CASE STUDY

water ENVIRONMENT RESEARCH

Automated data transfer for digital twin applications: Two case studies

Hanna Molin[1,2] 🟢 | Christoffer Wärff[1,3] | Erik Lindblom[1,2] |
Magnus Arnell[1,3] | Bengt Carlsson[4] | Per Mattsson[4] | Jonas Bäckman[5] |
Ulf Jeppsson[1]

[1]Division of Industrial Electrical Engineering and Automation (IEA), Department of Biomedical Engineering, Lund University, Lund, Sweden

[2]IVL Swedish Environmental Research Institute, Stockholm, Sweden

[3]RISE Research Institutes of Sweden AB, Gothenburg, Sweden

[4]Division of Systems and Control, Department of Information Technology, Uppsala University, Uppsala, Sweden

[5]Gemit Solutions AB, Stockholm, Sweden

Correspondence
Hanna Molin, IVL Swedish Environmental Research Institute, PO Box 21060, SE-10031 Stockholm, Sweden.
Email: hanna.molin@ivl.se

Abstract

Digital twins have been gaining an immense interest in various fields over the last decade. Bringing conventional process simulation models into (near) real time are thought to provide valuable insights for operators, decision makers, and stakeholders in many industries. The objective of this paper is to describe two methods for implementing digital twins at water resource recovery facilities and highlight and discuss their differences and preferable use situations, with focus on the automated data transfer from the real process. Case 1 uses a tailor-made infrastructure for automated data transfer between the facility and the digital twin. Case 2 uses edge computing for rapid automated data transfer. The data transfer lag from process to digital twin is low compared to the simulation frequency in both systems. The presented digital twin objectives can be achieved using either of the presented methods. The method of Case 1 is better suited for automatic recalibration of model parameters, although workarounds exist for the method in Case 2. The method of Case 2 is well suited for objectives such as soft sensors due to its integration with the SCADA system and low latency. The objective of the digital twin, and the required latency of the system, should guide the choice of method.

Practitioner Points

- Various methods can be used for automated data transfer between the physical system and a digital twin.
- Delays in the data transfer differ depending on implementation method.
- The digital twin objective determines the required simulation frequency.
- Implementation method should be chosen based on the required simulation frequency.

KEYWORDS

digital twins, edge computing, process modeling, real-time simulation, wastewater treatment, water resource recovery facility

This article is part of the special issue on Digital Twins in Water and Wastewater Treatment Systems.

INTRODUCTION

Digitalization is a global mega trend impacting all of society. Digital applications emerge and benefit at all levels of societies and businesses, that is, from component level to cross-sectoral systems (SWAN Forum, n.d.). However, making larger impact and realizing the larger vision of digitalization requires integrating solutions on systemic level (Arnell et al., 2023). One such example is digital twins. Digital twins have been gaining an immense interest in various fields over the last decade. Bringing conventional process simulation models into (near) real time will provide valuable insights for operators, decision makers, and stakeholders in many industries and parts of the society, from health sciences to manufacturing to city planning (Fuller et al., 2020). The great variation in application areas has, however, caused variations in the definition of a digital twin. In this work, we define a digital twin in line with Torfs et al. (2022) as

> ... a virtual copy of a physical object or system, with an automated live data connection to the physical entity. The digital twin should have measures to dynamically update and adjust the models based on relevant data to maintain an accurate description of the physical entity.

Inter-disciplinary comparisons are somewhat difficult due to the difference in the digital twin definition, but general application areas include monitoring, simulation, optimization and control, and verification of different activities in the real entity/process (Martinez et al., 2018). Advanced control and operational support are the most common applications in the wastewater context so far (Johnson, 2021; Sparks et al., 2023; Stentoft et al., 2020; Torfs et al., 2022), while other industries have come further, especially the manufacturing industry where several cases are documented in literature. Applications range from production planning and control, maintenance, and layout planning to product design and smart manufacturing (Kritzinger et al., 2018). The manufacturing industries are fundamentally different from the process industries in that the manufacturing industries in general are best described as discrete event systems, whereas the process industries operate continuously. This is also reflected in the advances in digital twin applications and implementation. The process industries see possible benefits in asset monitoring and maintenance, risk assessment, decision making, automation, and forecasts, as well as maximizing profit and production planning (Makarov et al., 2019; Wanasinghe et al., 2020). The wastewater sector is still lagging behind other process industries, but the activity is growing. The scientific literature is thus scarce with examples of digital twins in operation at water resource recovery facilities (WRRFs). Nonetheless, documented cases can be found in, for example, Johnson (2021) where a full-scale digital twin for water reuse and recovery was developed in Singapore to evaluate operational scenarios and for operational support and in Daneshgar, Polesel, et al. (2024) where a full-scale digital twin of Eindhoven WRRF for real-time prediction of plant operations is described. The focus in the wastewater sector has so far been on operational support and control with digital twins as an enabler for advanced control such as model predictive controllers (MPCs). Examples can be found in, for example, Stentoft et al. (2020) and Sparks et al. (2023).

An automated and continuous data connection requires soft digital infrastructure, such as standards, data models, policies, and so forth, and technical digital infrastructure such as standardized information and communication protocols, networks and services, storage, and sufficient bandwidth (Arnell et al., 2023; Barricelli et al., 2019; Schleich et al., 2017). There are several ways of automating the data transfer between the physical system and the virtual copy in a digital twin system. They can be divided into, at least, three main classes: (1) cloud computing on one end of the spectra; (2) a local set up on the other end; (3) and edge computing somewhere in between (Knebel et al., 2023; Rasheed et al., 2020). Cloud computing, either by using commercial cloud services or a cloud service developed in-house, is not presented in detail in this paper but is an alternative for realization of digital twins. Commercial cloud services make it possible to allocate data storage and computational power, with the drawback that data must be shared outside the organization. An in-house cloud service would diminish the need to share data outside the organization but can come with higher costs (Balasubramanian & Aramudhan, 2012; Knebel et al., 2023).

The data transfer itself can be constructed in several ways, depending on the type of industry and applications. O'Donovan et al. (2015) described data transfer for a manufacturing industry big data solution, where they used a process where data are collected in a distributed way from appliances at the plant. First, background applications running on local servers were scheduled to collect data and send to the cloud. In a next step, a temporary storage is used between the data collection and processing stage, to allow asynchronous operation and a more robust process. Another process is used to notify the relevant downstream process when new data is available for processing. Cakir et al. (2022) also describe

solutions for real time big data management in a manufacturing industry. Data collection is distributed, depending on the device used. Sensor data are, for example, collected through OPC-UA (Open Platform Communications Unified Architecture; Leitner & Mahnke, 2006). OPC-UA is a standard communication protocol used to securely exchange data between assets (Leitner & Mahnke, 2006). Data from energy analyzers are collected from the respective PLC and collected on an OPC server. From there, data are further distributed for processing. Daneshgar, Polesel, et al. (2024) presents an automated data pipeline for their digital twin. Data are collected from a database and automatically pre-processed using the Python package *wwdata* (De Mulder et al., 2018). The pre-processed data are published on a cloud server (Microsoft Azure) every 2 h from where it is accessible for simulations. The data transfer is automated using the Python package *wwtwin*, which pipelines the data transfer from raw data to simulation (Daneshgar, Borzooei, et al., 2024).

The objective of this paper is to describe two methods for implementing digital twins and highlight and discuss their differences and preferable use situations, with focus on the automated data transfer from the real process. We present two case studies on digital twins in Sweden—one that uses a custom-built local system for automated data transfer and one based on edge computing. In the next section, we describe the objective(s) of the digital twin and the method for automated data transfer for each case, respectively. The results and subsequent discussion focus on simulation frequency and run time, data transfer frequency, and a brief discussion on the time frame of the of the 'real-time' or 'live' concept, that is, how close to real-time is needed and what is achievable depending on infrastructure.

METHODOLOGY

In this section, we present the two case studies under development by the authors, focusing on the automated data transfer between the physical and virtual entities. Case 1 displays a tailor-made architecture for Öresundsverket WRRF in Sweden. Case 2 showcases an edge computing system using a commercial software platform developed for real-time simulation. For this paper, we consider "real-time" simulation to be with high frequency, for example, seconds, essentially as soon as data are recorded in the system a new simulation is run. Table 1 summarizes the differences between the two cases in terms of digital twin objectives, simulation frequency, and data transfer frequency, as well as data amount, types, and resolution.

Case 1—Öresundsverket WRRF

Plant description

The Öresundsverket WRRF is located in Helsingborg, Sweden. The plant receives a load of 164,000 population equivalent (mean over 2021–2022). It operates with an enhanced biological phosphorus removal (EBPR) activated sludge configuration with an activated primary clarifier setup for primary sludge hydrolysis and fermentation. The primary and secondary sludge are thickened in gravity thickeners and anaerobically digested. Ferric chloride is added to the primary sludge to bind orthophosphate and minimize problems with H_2S in the digesters. The plant suffers from periodic upsets in the EBPR process, which results in additional dosing of ferric chloride in the water line (mainly to the biological

TABLE 1 Summary of differences between the two cases.

	Definition	Case 1	Case 2
Digital twin objective	What is the digital twin used for?	Forecasting, advanced control, and process optimization	Decision support, fault detection, predictive maintenance
Simulation frequency	How often are simulations run?	Every hour	Every second
Data transfer frequency	How often are data transferred?	Every hour	Every second
Data resolution	What resolution does the data come in?	Minute	Second
Amount of data	How much data is transferred at a time?	146 tags 4 columns per tag 60 rows per transfer	1 data point per input and time step
Data types	What type of data is used in the digital twin?	Online data (currently)	Online data from sensors and equipment

reactors) to reach the effluent permits for total phosphorus of 0.5 mg P/L as annual average.

Models and objectives of the digital twin

The goal of the digital twin is to evaluate how and if it can be used to improve the process stability and minimize the use of chemicals and energy at the facility through optimization algorithms. The core of the digital twin is a plantwide mechanistic process model, developed in the commercial simulation software Sumo (Sumo 22.0, Dynamita, France). The model was initially manually calibrated and validated with in total 3 years of dynamic (hourly) data for the years 2020–2022. The digital twin is controlled through the Python (3.9.13) programming interface from the Sumo Digital Twin Toolkit. This means that data handling and corrections, writing dynamic influent files and process settings, running the model, and evaluating and visualizing results must be handled through Python scripts and requires programming skills to develop and make changes to. Rather than connecting the twin directly to outputs from sensors (although this is also possible with Sumo), the method uses a scheduling approach where simulations are scheduled and executed with a certain (regular or irregular) interval.

Optimization methods based on the Nelder–Mead algorithm have been developed and used with a simplified model to estimate dynamic influent data (Wärff et al., 2024). Further optimization methods are under development, with expectations to use the same algorithm, in the digital twin to improve plant operations. These include minimizing aeration energy while achieving target effluent ammonium concentrations by manipulating dissolved oxygen setpoints forward in time, minimizing total emissions of total phosphorus (TP) to the receiving water during high flows by manipulating the threshold for bypassing the biological reactors, and minimizing dosing of ferric chloride to achieve effluent TP complying with the permit. These types of optimizations are expected to run on an hourly basis.

Methods for automated data transfer

Data from online sensors are collected in the Supervisory Control and Data Acquisition (SCADA) system at the plant (Cactus Eye, Cactus Utilities, Sweden). Due to cyber security concerns, the operating utility (Nordvästra Skånes Vatten och Avlopp, NSVA) does not allow cloud connections to the SCADA system, and all non-essential software that is currently connected to the SCADA system are being disconnected. This sets the boundary limits for how data transfer and computations can be performed for digital twin purposes, as it is not possible to read data directly from the SCADA database. Instead, a data warehouse is used to store data from online sensors by transferring from the SCADA database, while laboratory data are directly saved to the data warehouse. The online data are transferred from the SCADA database to the data warehouse with a fixed sampling frequency, which currently is 1 h. This limitation in sampling frequency is due to the current data warehouse service subscription utilized by the utility, which limits it to 10 min. With another subscription, higher frequency would also be possible but is not deemed necessary for the purpose of this digital twin application. The plant is currently experiencing problems with delays in data transfer between the SCADA system and data warehouse, and in parallel, a new data lakehouse solution (Cinter, Sweden) is being developed that will directly stream new data without delay.

The data analytics software aCurve (Gemit Solutions, Sweden) is used to read specific data, relevant to the digital twin operation, from the data warehouse and write it to a secure FTP server (SFTP). For this operation, data for every data tag (e.g., sensor signals, laboratory measurement values, and process setpoints) are written to a unique file with comma separated values (i.e., a .csv file). Although there exist more efficient formats for storing and transferring large amounts of data (such as, e.g., Apache Parquet), .csv files were chosen since the data are easy to access and interpret. The system with one file per SCADA tag was also chosen to simplify accessibility to the files. Each .csv file is given a name corresponding to the name given in the Python script where the data is handled. All tag name handling (i.e., if a tag name is changed) is left out from the digital twin scripts and must be handled in the aCurve software. Due to the current experienced delay in data transfer, and that no delay between SCADA and database is expected shortly, the digital twin automation is developed with data collected with 24-h delay. In other words, data collected during 14.00–15.00 is transferred to the SFTP server at 15.00 the next day.

The content on the SFTP server can be synchronized to a folder on the PC running the digital twin model. This way, data can be accessed and read directly through Python scripts. Data validation, wastewater characterization, and influent generation can then be managed through Python scripts and transferred to the correct format through the Digital Twin Toolkit. Finally, Sumo can be run at fixed intervals with different objectives. In the current format, it is not possible to write results back to the SCADA system, but that can be done with minor

effort once the system is set up locally without Internet access by setting up an application programming interface (API) for communication between Python and the SCADA system. This remains to be implemented.

For this paper, an example is displayed where the digital twin is run to keep up to date with the latest data (i.e., no optimization runs or forecasting). This means that every hour when new data are available from online sensors on the past hour of operation, the digital twin is run with this set of data from the last update point 1 h earlier up until the endpoint of current available data. This simulation is the first step before any forecasting or optimization utilizing forecasted data, to provide initial conditions for such simulations. The full data pipeline for this is shown in Figure 1. The example was used to quantify lag in data transfer by (1) checking how long after a value was logged until it was available at the PC for processing and (2) how long the simulation time was for running the full plantwide model with the last hour of data (with a data time step of 1 h). Plant optimization and setpoint suggestions are not yet implemented, but when results are available, this will at first require manual input in the SCADA system from the user. At a later stage, setpoint implementation in the SCADA can be automated.

Case 2—henriksdals WRRF

Plant description

Henriksdals WRRF is located in Stockholm, Sweden, and operated by Stockholm Vatten och Avfall (SVOA). The plant serves approximately 780,000 people in the Stockholm metropolitan area. The process is an activated sludge process with pre- and post-denitrification. It is currently undergoing a substantial reconstruction to a membrane bioreactor (MBR) process where aerated membranes will replace the settlers. At the moment, one

out of eight treatment trains has been rebuilt. The primary and secondary sludge is thickened in belt thickeners and centrifuges, respectively, and then anaerobically digested.

SVOA are partners and co-funders of a research project with the objective to develop and evaluate applications of digital twins at WRRFs, including soft sensors, methods for fault detection, and advanced control to improve operation of various subprocesses. To evaluate the methods in real-time, a digital twin pilot has been launched. The pilot is done in collaboration with Siemens (Siemens, Sweden) but similar software from other companies also exists (ABB, 2023).

Models and objectives of the digital twin

The main objective of the digital twins is broad: to improve operations. Specific prioritized applications in the project include detection of process and sensor faults, robust control under faults, and methods to identify maintenance needs. The application areas so far deployed on the digital twin pilot are soft sensors and models for fault detection and isolation. One soft sensor was developed to monitor the dry solids content in thickened primary sludge using a recurrent neural network with 17 inputs (Molin et al., 2023). Once results have been verified online, the purpose is to use the soft sensor in a control loop for polymer dosing. The feedforward controller currently in place is thought to be inefficient since it fails to maintain the desired dry solids content and a more robust control strategy can potentially lower the polymer consumption and improve the overall process performance. The soft sensor has potential to be used to detect faults and to identify maintenance needs related to the process unit. The second soft sensor was developed to monitor the return activated sludge (RAS) flow rate since physical constraints at the treatment plant make it unfeasible to use conventional techniques for flow
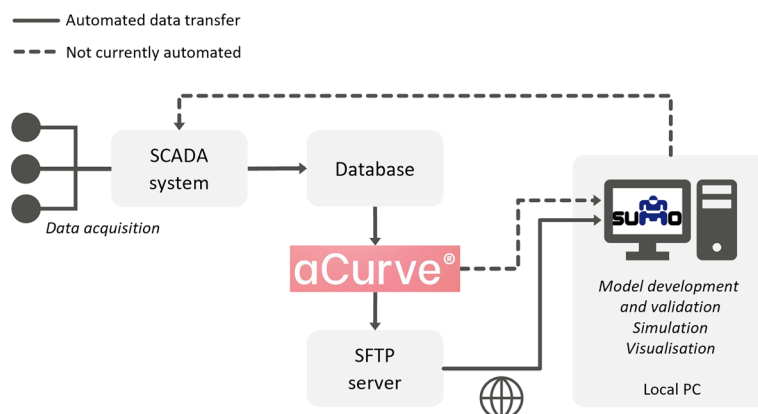

FIGURE 1 Data pipeline for the Öresundsverket WRRF digital twin in Helsingborg, Sweden. Dashed lines indicate that the data connection is not automated yet but can and will be in the future. Solid lines indicate that the data transfer is automated.

measurement at this location. Several methods were used for estimating the flow rate, including using a mass balance, a flow balance, and a pump model (Molin et al., 2022). The soft sensor was used to detect and isolate sensor faults using a statistical control chart in combination with a random forest classification model. The soft sensor can give more insight on the operational conditions and can be valuable to improve the control of the RAS pumps. The development and validation of all models were done offline in a conventional manner using historic data and the modeling software Matlab/Simulink R2021b and Python 3.10.

Although the processes involved are rather slow, the physical sensors that the soft sensors are to replace provide data with 1-s resolution. It is therefore desirable that the soft sensors provide data with the same resolution and as close to real time as possible.

Methods for automated data transfer

Siemens Industrial Edge offers various applications in edge computing. LiveTwin (version 2.1.16) is one of the applications, specifically developed for real-time simulations with high frequency data (Siemens, 2023). In edge computing, the simulations and computations are done close to the source which enables keeping information within the local network, decrease latency, and improve response times (Chen et al., 2020; Knebel et al., 2023). Cloud computing can achieve even faster response times than an edge computing system (e.g., Ren et al., 2023) but would require sharing data outside the organization or building an in-house cloud service, which is undesirable for SVOA. Furthermore, shorter response time is currently not required for the planned applications.

LiveTwin has full compatibility with Matlab/Simulink. Models developed in other software must be compiled as Functional Mock-up Units (FMUs) following the Functional Mock-up Interface (FMI) standard. The FMI standard is used for co-simulation of dynamic simulation models by compiling dynamic models as containers and

an interface using a combination of XML files, binaries, and C code, distributed as a ZIP file (Blochwitz et al., 2012; FMI, 2023). The Industrial Edge system is thus independent of the modeling software as long as the model can be compiled as an FMU. Python models can be compiled as FMUs using, for example, the PythonFMU package (Hatledal et al., 2020). However, most commercial wastewater modeling software do not currently follow this standard (FMI, 2024).

The compiled models are transferred to a management system manually, from where it can be deployed on an edge device (i.e., industrial PC). The simulations are initialized through the management system, where the user can change simulation frequency and parameter values, among other simulation instance settings. The default setting is to run simulations every 10 ms, but in practice, this will be limited by the frequency of the data source. It is possible to do simulations similar to those in Case 1, for example, scheduling an optimization algorithm once per hour, but that would require another Siemens application (Flow Creator, included in the Industrial Edge suite).

The edge device is in direct communication with the control system from where it reads data (Figure 2). The system can either read data from programmable logic controllers (PLCs) or via an OPC-UA server. The latter is used in the presented case. Data are accessible on the edge device as soon as it is presented on the OPC-UA server. The lag in data transfer, that is, how long time after a value is logged with a certain timestamp it is available on the edge device for simulation, was determined by comparing timestamps in the LiveTwin application and the SCADA system.

There are options on how to use the model output. Either the data can be visualized in the management system, or it can be written back automatically to the control system and used for monitoring or control. The latter option is currently not in use. Figure 2 shows a schematic overview of the architecture with dashed lines indicating that the data connection is not fully automated unless stated so by the user, and the dotted lines indicates that
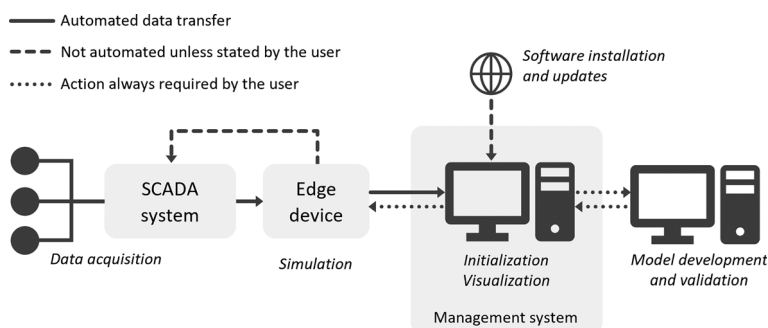

FIGURE 2 Schematic overview of the digital pilot at Henriksdals WRRF. Dashed lines indicate that the data connection is not fully automated unless stated so by the user. Dotted lines indicate that action is always required by the user. Solid lines indicate that the data transfer is always automated.

action is always required by the user. One example of this is the installation and updates of the Industrial Edge software or applications that require Internet access. On these occasions, standard security protocols are used. The transfer of process data to the edge device is always automated.

Combining simulation models is often an important feature of a digital twin. In the digital pilot at Henriksdals WRRF, this can be done in two ways: (1) by bundling models outside of LiveTwin or (2) by creating multiple FMUs and create data pathways between them within the Industrial Edge suite. The first option can potentially result in a complex model with issues when combining models from different software, while the second option makes it possible to run simulations on one or several of the FMUs simultaneously or consecutively giving more degrees of freedom to the user.

RESULTS AND DISCUSSION

In both case studies presented in this paper, the data latency and simulation time for one time step forward are low compared to the simulation frequency, causing no practical problems in running the simulations (Table 2). The architecture used in Case 1 is currently not able to feed the model with real-time data due to limitations in the data transfer rate from the database. If the database update frequency is increased to allow higher simulation frequency (e.g., 1 min), there would still be a constant lag of 2–3 min before the results could be utilized. The data lag for Case 1 will likely decrease once the digital twin is installed onsite and the data transfer to and from the SFTP-server are no longer necessary. However, the latency is still expected to be higher than in Case 2 as long as the utility does not allow direct connections to the SCADA system. For Case 2, data are accessible to the edge device and, thus for simulations, as soon as it is published on the OPC-UA server. This ensures rapid data transfer, in the range of milliseconds, which is a prerequisite for real-time simulations as is often needed for soft sensors. The sampling time is 1 s, and any delays shorter than that is not measurable, hence the reported latency of <1 s (Table 2). Since the data transfer is only for one time step at a time in the default case, the amount of data transferred at every instance is small compared to the data transferred in Case 1. The use of .csv files with

one file per tag in Case 1 is not optimal from a data transfer standpoint and could likely be optimized, but for this application, it does not have an impact.

Several of the methods described in the literature for manufacturing industries (e.g., Cakir et al., 2022; O'Donovan et al., 2015) use a distributed data collection and transfer system. This can be a viable option when data is not collected in a centralized location or to reduce complexity or latency by not requiring a connection to the full SCADA system or database when only a few data sources are used. For a full plant wide digital twin for a WRRF, a distributed system might, however, introduce more complexity than a centralized system due to the large number of inputs required for such digital twins. The so far published case studies about digital twins at WRRFs use centralized systems similar to Case 1 presented here (Daneshgar, Polesel, et al., 2024; Johnson, 2021).

It is important to note that the simulated models in Cases 1 and 2 are not identical but differ substantially in complexity. The values on *simulation run time* presented in Table 2 are therefore not directly comparable between the two methods and should be analyzed in view of the different objectives of the digital twins. In Case 1, as well as in Johnson (2021) and Daneshgar, Polesel, et al. (2024), a scheduling approach is used for simulations, meaning that simulations are scheduled with a fixed time interval (e.g., every hour). This can be useful both to update the model with the latest data and make forecasts and execute optimization scripts. Simulation near real time (i.e., every second) is not required for forecasting and optimization as the optimization horizon normally would be hours into the future. For comparison, Daneshgar, Polesel, et al. (2024) uses a 2-h interval to ensure adequate data pre-processing and thus data quality. The applications in Case 2 have so far been soft sensors which require real-time data to fully replace the physical sensors. If the goal is to continuously update the model and use the output for, for example, fault detection, the edge computing system in Case 2 could be more convenient as it assures rapid, continuous data transfer. It should be noted that none of the cases presented here are in practice limited to one or the other (scheduled or continuous simulation) but have been developed with different primary areas of use in mind and are thus better suited for different applications.

TABLE 2 Comparison of latency and simulation runtime between the two cases.

	Definition	Case 1	Case 2
Latency	Are there lags in the data transfer?	2 min	<1 s
Simulation run time	How long does a simulation take?	12 s	0.2–0.6 ms

Data transfer between a digital twin and the real process can be one-directional or bi-directional, with the recurring discussion point of whether bi-directional communication with the real system is required for the definition of a digital twin (e.g., Fuller et al., 2020). Torfs et al. (2022) argues that the virtual-to-real entity connection can be automated or manually performed. In practice, the need of a bi-directional connection is determined by the digital twin objective. A bi-directional automated connection is possible to establish in both presented case studies (Case 1 and Case 2) but is yet to be implemented. Until it has been implemented and launched, the simulation outputs (e.g., controller setpoints, suggested actions, or sensor values) are accessible to the operators for decision making in both cases (i.e., using humans in the loop).

Besides the data transfer from the physical plant to the digital twin, and from the digital twin to the physical plant, data transfer or information transfer between models can be a crucial part of the digital twin system as it may consist of several models. Simulations in parallel or series (i.e., co-simulation or model exchange) is a requested future functionality in both presented case studies and can be done in both systems. The difference between the cases does not lie in functionality but in the way it is handled. The overarching Python script used to initialize simulations in Case 1 allows simulation of multiple models in parallel or series. In Case 2 in can be done either by creating data flows between models within the Industrial Edge suite, or by combining models before deployment on the edge device.

CONCLUSION

The developments that have followed the ongoing digitalization have increased the options for implementation of digital twins. There are tools that can be used to locally setup an architecture tailored to the specific plant, as described in Case 1, and there are commercial edge or cloud computing services available, as exemplified in Case 2. The two case studies presented are different in terms of objectives of the digital twins, and thus sets different requirements on the automated data transfer. The system requirements depend on the objective of the digital twin and thus the required simulation frequency. For digital twins where the signals are used directly, such as for continuous model simulation or fault detection, the approach of Case 2 is likely necessary. For digital twins where plant optimization and forecasting are used, the scheduling approach of Case 1 is likely more efficient. Both approaches could be used for essentially the same applications, though. A combination of systems or a layered approach could be considered to fully utilize the strengths in both systems.

AUTHOR CONTRIBUTIONS

Hanna Molin: Writing - original draft; methodology; formal analysis; visualization; conceptualization; investigation. **Christoffer Wärff:** Writing - original draft; methodology; formal analysis; visualization; conceptualization; investigation. **Erik Lindblom:** Writing - review and editing; supervision. **Magnus Arnell:** Writing - review and editing; supervision. **Bengt Carlsson:** Writing - review and editing; supervision. **Per Mattsson:** Supervision; writing - review and editing. **Jonas Bäckman:** Software; resources. **Ulf Jeppsson:** Supervision; writing - review and editing.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Hanna Molin https://orcid.org/0009-0004-8299-8114

REFERENCES

ABB. (2023). Digital twin applications. Retrieved November 24, 2023, from https://new.abb.com/control-systems/features/digital-twin-applications

Arnell, M., Miltell, M., & Olsson, G. (2023). Making waves: A vision for digital water utilities. *Water Research X*, *19*, 100170. https://doi.org/10.1016/j.wroa.2023.100170

Balasubramanian, R., & Aramudhan, M. (2012). Security issues: Public vs private vs hybrid cloud computing. *International Journal of Computer Applications*, *55*(13), 35–41. https://doi.org/10.5120/8818-2808

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

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 *9th International MODELICA Conference, 3-5 September, Munich, Germany* (pp. 173–184). The Modelica Association. https://doi.org/10.3384/ecp12076173

Cakir, A., Akin, Ö., Deniz, H. F., & Yilmaz, A. (2022). Enabling real time bid data solutions for manufacturing at scale. *Journal of Big Data*, 9, 118. https://doi.org/10.1186/s40537-022-00672-6

Chen, Y., Deng, S., Ma, H., & Yin, J. (2020). Deploying data-intensive applications with multiple services components on edge. *Mobile Network Applications*, 25, 426–441. https://doi.org/10.1007/s11036-019-01245-3

Daneshgar, S., Borzooei, S., Peeters, R., Weijers, S., Nopens, I., & Torfs, E. (2024). wwtwin: An open-source modular Python package for digital twin applications of WRRFs. 9th IWA Water Resource Recovery Modelling Seminar (WRRmod), 6–10 April, Notre Dame, USA.

Daneshgar, S., Polesel, F., Borzooei, S., Sørensen, H. R., Peeters, R., Weijers, S., Nopens, I., & Torfs, E. (2024). A full-scale operational digital twin for a water resource recovery facility—A case study of Eindhoven Water Resource Recovery Facility. *Water Environment Research*, 96(3), e11016. https://doi.org/10.1002/wer.11016

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

FMI. (2023). About. Retrieved November 24, 2023 from https://fmi-standard.org/about/

FMI. (2024). Tools. Retrieved May 17, 2024 from https://fmi-standard.org/tools/

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

Hatledal, L. I., Zhang, H., & Collonval, F. (2020). Enabling Python driven co-simulation models with PythonFMU. In M. Steglich, M. Mueller, G. Neumann, & M. Walther (Eds.), *Proceedings of 34th International ECMS Conference on Modelling and Simulation*. European Council for Modeling and Simulation. https://doi.org/10.7148/2020-0235

Johnson, B. (2021). Water reuse and recovery facility connected digital twin case study: Singapore PUB's Changi WRP process, control, and hydraulics digital twin. WEFTEC 2021, 16–20 October, Chicago, USA.

Knebel, F. P., Revisan, R., do Nascimento, G. S., Abel, M., & Wickboldt, J. A. (2023). A study on cloud and edge computing for the implementation of digital twins in the oil & gas industries. *Computers & Industrial Engineering*, 182, 109363. https://doi.org/10.1016/j.cie.2023.109363

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

Leitner, S. H., & Mahnke, W. (2006). OPC UA–service-oriented architecture for industrial applications. *Softwaretechnik-Trends*, 26(4). Bonn: Geselllschaft für Informatik e.V. PISSN: 0720-8928. Berichte aus den Fachgruppen und Abeitskreisen.

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

Martinez, G. S., Sierla, S., Karhela, T., & Vyatkin, V. (2018). Automatic generation of a simulation-based digital twin of an industrial process plant. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, 21–31 October, Washington DC, United States* (pp. 3084–3089). IEEE. https://doi.org/10.1109/IECON.2018.8591464

Molin, H., Bröndum E., Nilsson, S., Lindblom, E., Saagi, R., Carlsson, B., & Jeppsson, U. (2023). Using machine learning to predict the total solids concentration in thickened primary sludge thickener at Henriksdal WRRF. 11th IWA Symposium on Modelling and Integrated Assessment Watermatex 2023, 23–27 September, Quebec City, Canada.

Molin, H., Lindblom, E., Saagi, R., Carlsson, B., & Jeppsson, U. (2022). Process monitoring and fault detection using a soft sensor for the return activated sludge flow rate at Henriksdal WRRF. World Water Congress 2022, 11–15 September, Copenhagen, Denmark.

O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data*, 2, 25. https://doi.org/10.1186/s40537-015-0034-z

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

Ren, J., Xu, C., Wang, J., Zhang, J., Mao, X., & Shen, W. (2023). An edge-fog-cloud computing-based digital twin model for prognostics health management of process manufacturing systems. *Computer Modeling in Engineering & Sciences*, 135(1), 599–618. https://doi.org/10.32604/cmes.2022.022415

Schleich, B., Anwer, N., Mathieu, L., & Wartzack, S. (2017). Shaping the digital twin for design and production engineering. *CIRP Annals – Manufacturing Technology*, 66, 141–144. https://doi.org/10.1016/j.cirp.2017.04.040

Siemens. (2023). LiveTwin. Retrieved November 24, 2023, from https://www.dex.siemens.com/edge/manufacturing-process-industries/livetwin?cclcl=en_US

Sparks, J., Vanrolleghem, P. A., & Bott, C. (2023). Full scale digital twin with integrated hybrid model predictive controller for ammonia-based aeration control. 11th IWA Symposium on Modelling and Integrated Assessment Watermatex 2023, 23–27 September, Quebec City, Canada.

Stentoft, P. A., Vezzaro, L., Mikkelsen, P. S., Grum, M., Munk-Nielsen, T., Tychsen, P., Madsen, H., & Halvgaard, R. (2020). Integrated model predictive control of water resource recovery facilities and sewer systems in a smart grid: Example of full-scale implementation in Kolding. *Water Science and Technology*, 81, 1766–1777. https://doi.org/10.2166/WST.2020.266

SWAN Forum. (n.d.). What is a smart water network? The Smart Water Network (SWAN) Forum. Retrieved 2021-05-04 from https://swan-forum.com/smart-water-network/ https://www.swan-forum.com/swan-tools/a-layered-view/

Torfs, E., Nicolai, N., Daneshgar, S., Copp, J. B., Haimi, H., Ikumi, D., Johnson, B., Plosz, B. B., Snowling, S., Townley, L. R., Valverde-Pérez, B., Vanrolleghem, P. A., Vezzaro, L., & Nopens, I. (2022). The transition of WRRF models to digital twin applications. *Water Science Technology*, 85, 2840–2853. https://doi.org/10.2166/wst.2022.107

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*, *8*, 104175–104197. https://doi.org/10.1109/ACCESS.2020.2998723

Wärff, C., Carlsson, B., Arnell, M., Saagi, R., Micolucci, F. & Jeppsson, U. (2024). Comparing methods for reconstructing high frequency WRRF influent data from downstream measurements. 9th IWA Water Resource Recovery Seminar (WRRmod), 6–10 April, Notre Dame, Indiana, USA.