



The transition of WRRF models to digital twin applications

Torfs, Elena; Nicolaï, Niels; Daneshgar, Saba; Copp, John B.; Haimi, Henri; Ikumi, David; Johnson, Bruce; Plosz, Benedek B.; Snowling, Spencer; Townley, Lloyd R.

Total number of authors:
14

Published in:
Water Science and Technology

Link to article, DOI:
[10.2166/wst.2022.107](https://doi.org/10.2166/wst.2022.107)

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):

Torfs, E., Nicolaï, N., Daneshgar, S., Copp, J. B., Haimi, H., Ikumi, D., Johnson, B., Plosz, B. B., Snowling, S., Townley, L. R., Valverde-Pérez, B., Vanrolleghem, P. A., Vezzaro, L., & Nopens, I. (2022). The transition of WRRF models to digital twin applications. *Water Science and Technology*, 85(10), 2840-2853. <https://doi.org/10.2166/wst.2022.107>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The transition of WRRF models to digital twin applications

Elena Torfs^{IWA id a,b,*}, Niels Nicolai^{IWA id c,d}, Saba Daneshgar^{IWA id a,b}, John B. Copp^{IWA id e}, Henri Haimi^{IWA id f}, David Ikumi^{IWA id g}, Bruce Johnson^{IWA id h}, Benedek B. Plosz^{IWA id i}, Spencer Snowling^{IWA id j}, Lloyd R. Townley^{IWA id k,l}, Borja Valverde-Pérez^{IWA id m}, Peter A. Vanrolleghem^{IWA id c,d}, Luca Vezzaro^{IWA id m} and Ingmar Nopens^{IWA id a,b}

^a Biomath, Ghent University, Coupure links 653, 9000 Gent, Belgium

^b Centre for Advanced Process Technology for Urban Resource recovery (CAPTURE), Frieda Saeystraat 1, 9000 Gent, Belgium

^c modelEAU, Université Laval, 1065 avenue de la Médecine, Québec G1 V 0A6, QC, Canada

^d CentrEau, Québec Water Research Centre, 1065 avenue de la Médecine, Québec G1 V 0A6, QC, Canada

^e Primodal, Inc., 122 Leland Street, Hamilton, Ontario L8S 3A4, Canada

^f FCG Finnish Consulting Group Ltd, Osmontie 34, P.O. Box 950, FI-00601 Helsinki, Finland

^g Water Research Group, Department of Civil Engineering, University of Cape Town, Rondebosch, 7700 Cape, South Africa

^h Jacobs, 9191 Jamaica St, Englewood, CO 80112, USA

ⁱ Department of Chemical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, UK

^j Hatch, Ltd, 2800 Speakman Dr, Mississauga, ON, Canada, L5 K 2R7

^k Nanjing University Yixing Environmental Research Institute, 128 Hengtong Road, Yixing Jiangsu, 214200, China

^l Nanjing Smart Technology Development Co. Ltd, Yanlord Landmark Building B Suite 706, Jiangxinzhou Jianye, Nanjing Jiangsu 210019, China

^m Department of Environmental Engineering, Technical University of Denmark, Bygningstorvet, Building 115, 2800, Kongens Lyngby, Denmark

*Corresponding author. E-mail: elena.torfs@ugent.be

id ET, 0000-0002-5629-6950; NN, 0000-0003-2794-2930; SD, 0000-0002-6407-1145; JBC, 0000-0002-5080-9997; DI, 0000-0001-9158-2525; BJ, 0000-0002-0012-6477; BBP, 0000-0002-7081-7038; SS, 0000-0001-8037-8983; LRT, 0000-0003-2923-628X; BV-P, 0000-0001-7255-1395; PAV, 0000-0003-1695-1313; LV, 0000-0001-6344-7131; IN, 0000-0001-6670-3700

ABSTRACT

Digital Twins (DTs) are on the rise as innovative, powerful technologies to harness the power of digitalisation in the WRRF sector. The lack of consensus and understanding when it comes to the definition, perceived benefits and technological needs of DTs is hampering their widespread development and application. Transitioning from traditional WRRF modelling practice into DT applications raises a number of important questions: When is a model's predictive power acceptable for a DT? Which modelling frameworks are most suited for DT applications? Which data structures are needed to efficiently feed data to a DT? How do we keep the DT up to date and relevant? Who will be the main users of DTs and how to get them involved? How do DTs push the water sector to evolve? This paper provides an overview of the state-of-the-art, challenges, good practices, development needs and transformative capacity of DTs for WRRF applications.

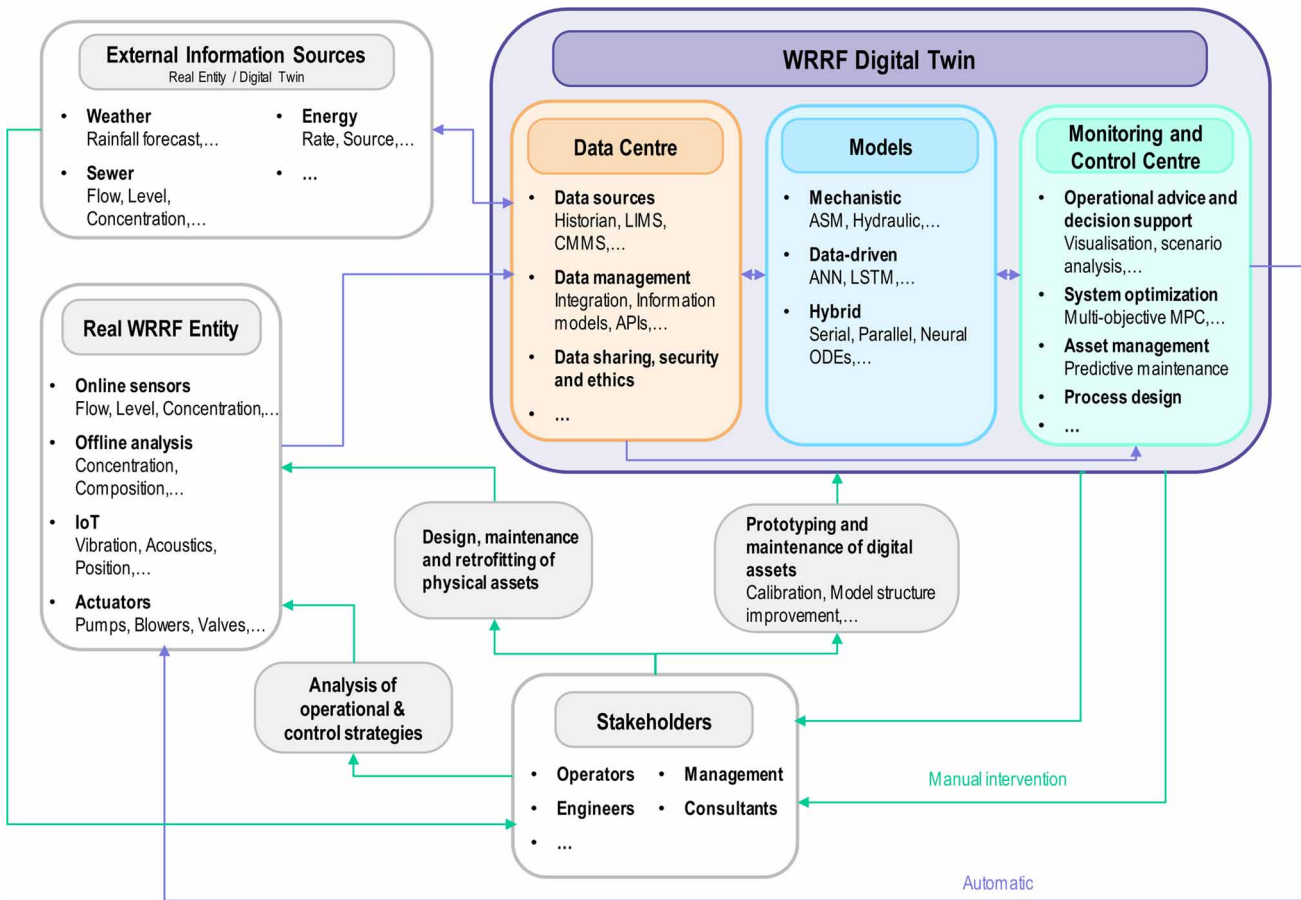
Key words: data management, hybrid models, interoperability, model predictive control, online optimisation, proactive management

HIGHLIGHTS

- A Digital Twin distinguishes itself from a simulation model by a continuous, automated data connection.
- Current DT projects in WRRFs focus on operational support and control.
- Combining mechanistic models and data-driven techniques into hybrid models can accelerate the adoption of DTs.
- Data and information models are key for the implementation and upscaling of DTs.
- WRRF staff should be included during the development stages of DTs.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

GRAPHICAL ABSTRACT



1. INTRODUCTION

The water industry is transforming and digitalisation is a key part of this transition (IWA & Xylem Inc 2019). However, digitalisation and its derivatives such as Digital Twins (DTs) seem to be buzzwords adopted from other industries and used without sufficient consensus and in-depth understanding of the challenges and development needs ahead. Moreover, many digitalisation initiatives seem to be skewed towards drinking water distribution applications (e.g. IWA's digital water program, the SWAN working group on Digital Twin applications etc.) with Water Resource Recovery Facility (WRRF) applications gaining less attention. This delayed uptake in the WRRF sector can be explained by the complex biological reactions that are governing typical treatment systems leading to specific challenges with data collection and model development.

The transition from conventional process models to DTs should be motivated by the plant's objective. As such, offline process models used for design and decision-making and DTs used for automated live process monitoring, control, and/or optimisation can be complementary. To achieve the transition to a DT, dramatic changes at both cultural and technical levels need to be put into practice. For example, the predictive power of conventional mechanistic modelling practices for WRRFs (e.g. ASM, ADM, and BSM) should be assessed and, if needed, existing model structures should be combined with data-driven techniques. The way water and wastewater utilities collect and store data can be optimised to have more relevant and consistent data. Finally, trust between the operational staff, leadership, and the DT will be needed. It is important to note that by looking more holistically, a WRRF DT can be integrated in a higher level decision support system. This multi-disciplinary approach with inter-connection between various subsystems from different technical domains brings a whole new set of challenges and makes the need for some level of agreement on DT definitions and elements even more critical.

In order to realise the perceived benefits from the rapidly expanding concept of DTs, consensus is needed to provide clear definitions of what DT is. Currently, consensus is missing in the WRRF modelling community concerning the state-

of-the-art, challenges, good practices, development needs and transformative capacity of DTs for WRRF applications. To overcome this barrier, a dedicated workshop was organised during the (virtual) 7th IWA/WEF Water Resource Recovery Modelling Seminar 2021. This paper summarises the results of this workshop and presents: (1) a definition of DTs in the WRRF sector based on key features that distinguish a DT from conventional simulation models, (2) guidelines on the transition of WRRF modelling efforts to successful application of DTs, (3) a comprehensive overview of technical and non-technical challenges and development needs for DTs, and (4) an outlook for the future of DTs as a transformative digital decision-making tool.

2. DIGITAL TWIN DEFINITION

The origin of using the DT terminology goes back to the early 2000s but it did not appear in official publications until the early 2010s where it was first introduced in the manufacturing industry (Glaessgen & Stargel 2012; Grieves 2014) as a new tool for the management of a product's lifecycle. Since then, the concept of DTs has spread throughout many different sectors including the water industry.

Despite the growing popularity of DT terminology in the domain of wastewater treatment, the distinction between DTs and conventional process modelling is not consistently defined, leading to the erroneous use of the term in many studies where models are applied for process design and decision-making. This inconsistency may lead to misunderstanding, contribute to the idea that DTs are merely hype and consequently may even slow down the adoption of DTs (Wright & Davidson 2020). Hence, there is an increasing need to define what differentiates DTs from conventional simulation models so that a common understanding on the topic can be maintained within our domain.

Many definitions of DTs can be found in the literature across different domains (Curl *et al.* 2019; Karmous-Edwards *et al.* 2019; Fuller *et al.* 2020; Wright & Davidson 2020). General consensus can be found in the idea that what distinguishes a DT from a conventional simulation model is its continuous **connection** with the physical twin (Jones *et al.* 2020). This connection should include the following features:

1. The idea of a twin inherently assumes that **there exists a physical counterpart**. Hence, a DT should be connected to a real entity (e.g. equipment, process or product) even though its development process can start as soon as the real entity is in its conceptualisation stage. This is especially relevant in discussions on the potential of DTs for design (greenfield facilities, retrofitting existing reactors, extending existing plants with new treatment lines).
2. A DT should have an **automated, live data connection to the real entity**. This connection is preferably bi-directional such that insight generated by the virtual twin is fed back to the real entity. Whereas the data connection from the real entity to the DT should be automated, the virtual-to-real entity connection may be either automated (direct input from the DT into SCADA) or manually performed using humans in the loop.
3. Twins should evolve together. Hence, the DT should include the means to **dynamically update** or adjust the models based on relevant data to maintain an accurate description of the real entity as it evolves over time. A DT thus becomes a continuously updated knowledge repository that is automatically kept relevant and thus provides a much higher level of usefulness as compared to a conventional simulation model executed independently.

In the context of wastewater treatment and resource recovery, the definition and application of DTs has been discussed in a recent white paper by the IWA specialist group on modelling and integrated assessment (IWA 2021a) as well as during a dedicated workshop at the WRRmod2021 conference. Although there is general agreement on the features listed above, there are still unresolved questions particularly with respect to the aspect of real-time connectivity. A real-time data connection is an ill-defined concept because the time frame is arbitrary. Are data from 1 hour or 1 day ago still considered real-time? Therefore, the authors propose that the time horizon for model updates and simulation should be defined by the DT objectives (e.g. instrument failure detection, real-time control) and the relevant system dynamics (aeration vs. sludge age). Hence, the automated data-feed and dynamic model updating are what defines a DT whereas the time horizon over which this is performed is flexible.

An overview of the components of a DT and its connections to the real entity and relevant stakeholders is provided in Figure 1.

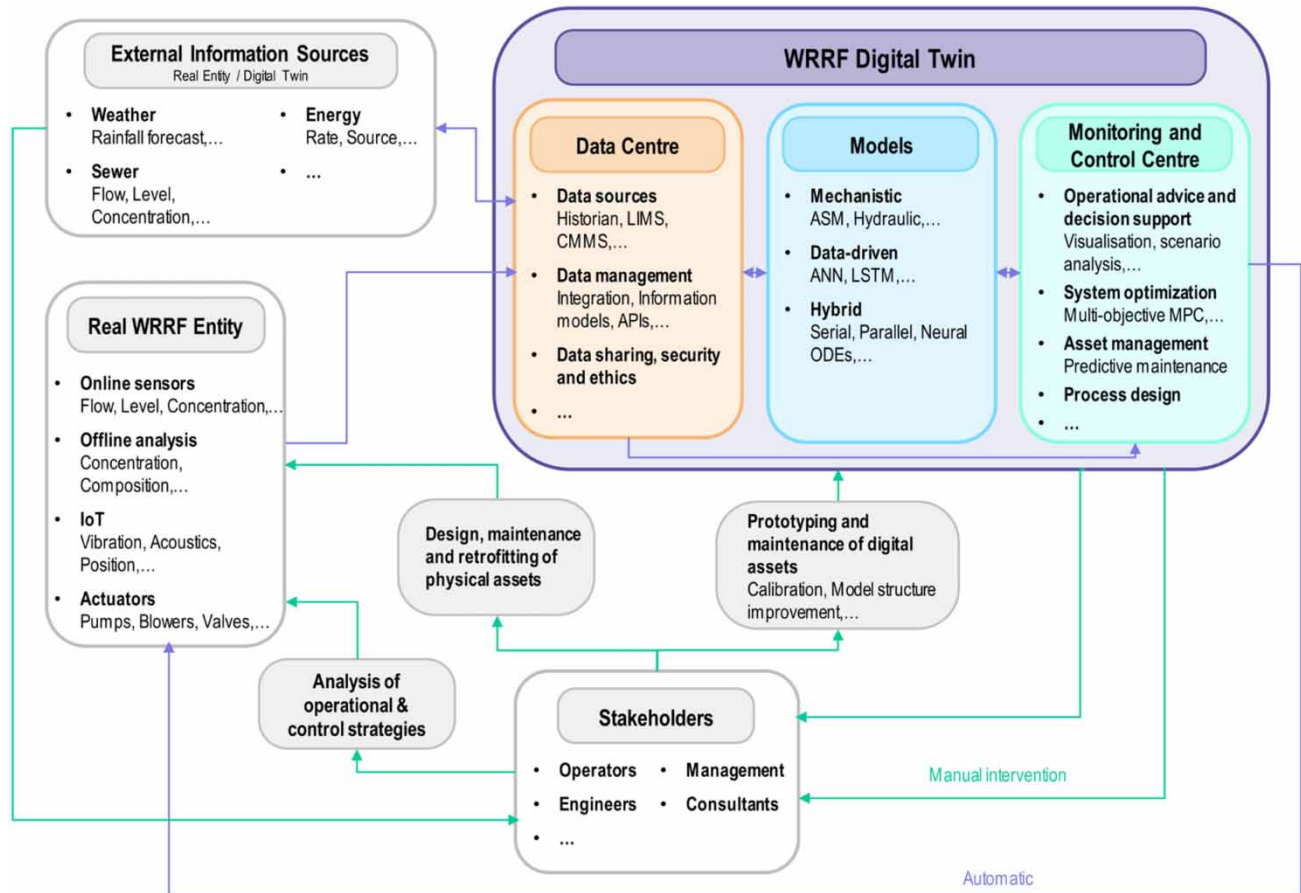


Figure 1 | Schematic representation of a Digital Twin and its interaction with the physical system, control systems and stakeholders.

3. DIGITAL TWIN APPLICATIONS IN WRRFS

In general, the application of DTs supports operations and management staff in making data both time-relevant and actionable in a way that traditional data analysis and modelling cannot. Based on data from the physical asset or system, a DT unlocks value principally by supporting improved decision-making, which creates the opportunity for positive feedback into the physical twin (Bolton *et al.* 2018).

As such, in WRRFs, DTs can support the transition towards proactive management, whereby different processes and assets can be operated and maintained, to mitigate disturbances and other issues before they have undue adverse impacts on performance (Karmous-Edwards *et al.* 2019). As a result, there is significant potential for economic savings (e.g. energy optimisation) and more effective protection of the environment (e.g. better nutrient removal and recovery, reduction in GHG emissions). With WRRFs contributing approximately 1–3% of total global energy consumption (Mamais *et al.* 2015) and about 2–3% of total global GHG emissions (Maktabifard *et al.* 2019), significant gains can be foreseen from real-time optimisation and advanced control supported by DTs. Moreover, DTs can accelerate the transition to a circular water economy by tailoring effluent quality towards water reuse purposes.

DT applications include (but are not limited to): (i) data-driven decision support for the selection of different operational strategies and operator training (Johnson *et al.* 2021); (ii) online (multi-objective) system optimisation (e.g. model-predictive control) for energy or resource savings or compliance management (e.g. to minimise carbon footprint) (Stentoft *et al.* 2020, 2021); (iii) failure analysis (Jain *et al.* 2020); (iv) asset management and predictive maintenance (Bartos & Kerkez 2021); (v) investment planning (Ruohomäki *et al.* 2018); and (vi) decision support for policy making (Poch *et al.* 2020). Although the potential for application of DTs is far-reaching, only a limited number of successful implementations are documented in the literature for WRRF systems and they mainly focus on operational decision support and online system optimisation through advanced control. In the list above, for example, only Johnson *et al.* (2021) and Stentoft *et al.* (2020, 2021) describe

applications in the WRRF sector, whereas other examples are taken from related domains such as water network management, smart cities, and the manufacturing industry. Some examples of application of DTs in the WRRF sector for operational decision making and online optimisation are described in more detail below.

3.1. Digital twins for operational advice

A DT was developed for the Singapore PUB Changi Water Reclamation Plant (WRP) (Johnson *et al.* 2021). This DT includes the whole plant process, hydraulics, and controls and automatically accepts over 1200 data streams from both SCADA and Laboratory Information Management systems. It is implemented on a dedicated server system located at the Changi WRP accepting data from a replicated historian system. It has been implemented as a hybrid model that uses both mechanistic and data-driven models, and includes some level of automatic calibration through the use of both the SCADA and laboratory data connections. The Changi WRP DT includes current status evaluations of measured versus DT results, automated scenario analysis (i.e. what happens if secondary settling tanks or bioreactors are out of service?), and a 5-day hourly prediction (a wastewater 'weather' forecast) of the plant performance. Operations staff are free to use the information provided by the DT to help them transition to a more proactive operational mode. It has been implemented in the advisory-only mode at this point to provide operations staff time to both improve accuracy and develop the needed trust prior to proving its control authority. The DT can also be used for operator training as it allows the operators to 'operate' the facility virtually in a safe environment under conditions that are rarely observed. These conditions include equipment failures, extreme high or low flows, emergency shutdowns and system restarts.

3.2. Digital twins for model predictive control

Model predictive control (MPC) allows for online optimisation of WRRFs. Stentoft *et al.* (2020) presented an MPC for integrated control of a pumping station in the catchment of Kolding WRRF (Denmark). The aim of the controller was to minimise energy usage while respecting discharge limits. The DT consisted of a catchment model (including both the sewer line and dry weather inlet flow forecast) as well as an effluent quality model, both being empirical models calibrated on an hourly basis. The effluent quality model uses measurements of ammonia and nitrate concentrations at 2 minute intervals to approximate the effluent TN. First, the data are resampled to hourly values and then, model parameters are estimated using maximum likelihood estimation and a Kalman filter. The catchment model was, however, calibrated offline and predicts the flow rate based on the time of the day under dry weather conditions. The DT is implemented as an extra module of the control system. The example falls within the scope of a DT, as the models are dynamically updated with data from the physical twin, while the DT can also communicate with the physical counterpart to improve its performance. The MPC optimises the flow rate of the pumping station every hour with a prediction horizon of 24 hours. The controller was successfully validated during two operational periods over a total of 7 days. Currently, the developer is working on an extension of the MPC to optimise WRRFs based on competing operational objectives, such as cost, effluent quality, and carbon footprint (Stentoft *et al.* 2021). To that end, new data-driven models need to be developed and validated, as it is important that the DT adapts to the new optimisation needs.

3.3. Digital twins for design

The potential application of DTs for design of WRRFs is a topic of much discussion as automated data coupling is often lacking when an entity is in its design phase. However, in the manufacturing industry several applications of DTs for product design can be found (Wright & Davidson 2020): so-called prototyping DTs are fed with data from existing objects or products and used to update the design of new versions, which may or may not reach the commercial phase, but generate new data to update the DT. Within the WRRF sector, the main interest is currently on DTs for operational purposes, as current practice for model-based design does not include any automated data coupling (Pedersen *et al.* 2021). This, however, does not mean that a model developed in a WRRF's design phase cannot transition into an operational DT as the reactor or facility goes into operation. A recent case study that investigates this transition between design and operation is the work performed by Alex *et al.* (2020), which illustrates how a detailed process model of a plant including its control system and actuators is used for virtual commissioning of the automation system. This process model developed during the planning and design phase of the automation concept can continue its lifecycle as an operational DT once the automation system is connected to the real entity. A similar example is the use of CFD models for design and DT development. Whereas CFD models are not yet directly used in DT applications for WRRFs, a detailed CFD model developed in the design phase can serve as an important information source for the development of an operational DT through the derivation of compartmental models (Le Moullec

et al. 2011). Finally, DTs of existing systems can be powerful tools for evaluating potential process design changes because, by definition, the DT represents an accurate and relevant description of the real entity's current state.

4. MODELLING TOOLS AND PREDICTIVE POWER

A DT has a mathematical model at its core. In principle, a DT can use any type of model that is a sufficiently accurate representation of its physical counterpart (mechanistic or first principles models; fully data-driven models; or a combination of both, so-called hybrid models). In practice, the choice of model structure will be influenced by a number of factors that are specifically related to the application and specific objectives of the DT as this will determine the timescale over which a DT needs to be evaluated and updated. Given the dominant dynamics of most governing processes in WRRFs, this will likely be in the range of minutes, hours or days for operational purposes. However, for applications such as predictive maintenance or investment planning, the timescale can be much larger (Wright & Davidson 2020).

In an ideal world where computational power is not an issue and all processes in WRRFs are perfectly understood, DTs would make use of so-called mechanistic or first principles models. Mechanistic models describe the system in a fixed, structured way derived from underlying physical, chemical, and biological mechanisms, often represented by balance equations of some quantities which change in time and space. The parameters of a mechanistic model have a physical meaning in the system. In general, mechanistic models have a relatively high computational power requirement making them less suitable for high-frequency DT applications where model updates are needed in real-time (for example for MPC applications) (Pantelides & Renfro 2013; Pedersen *et al.* 2021). For DT applications with a longer actionable time-horizon such as maintenance or investment planning, mechanistic models are highly relevant provided that their predictive power is sufficiently high.

At the other end of the spectrum, data-driven or so-called black box models describe the system purely based on information extracted out of the process data. The difference between pure black-box models (for example neural networks) and empirical relations derived from data (such as Monod kinetics or settling velocity functions) lies in their unstructured nature – the number and nature of the parameters are flexible and not fixed in advance by knowledge (von Stosch *et al.* 2014). These models do not incorporate any information on the underlying process dynamics and are therefore only reliable within the region of input parameter space from which the data used to construct the model was taken. Extrapolation of data-driven models beyond the operating space for which they were developed is a dangerous approach. Therefore, compared to mechanistic models, data-driven models need larger datasets to accurately predict a range of operational conditions. Moreover, their lack of interpretability is another issue for their use in DT applications. However, data-driven models are much less computationally expensive, making them very interesting for real-time DT applications such as data-driven soft sensor development for real-time monitoring (Haimi *et al.* 2013) and model predictive control for online optimisation (Stentoft *et al.* 2018). Historically, WRRF modellers are accustomed to developing and using mechanistic models whereas data-driven models are not yet widely used for these systems. This is related to the lack of expertise in data-driven techniques within the WRRF modelling community and the absence of sufficient trust in black-box modelling performance. It is also linked to the inherent nature of WRRF processes which are non-stationary and have significant temporal dependency which requires frequent retraining of the data-driven model. A comprehensive overview of barriers to the application of Big Data analytics in the domain of WRRFs as well as further examples of applications of data-driven models for real-time monitoring, fault detection, and control in WRRFs can be found in Newhart *et al.* (2019).

Two possible solutions exist to overcome the disadvantages of both mechanistic and data-driven models. A first very promising approach is the combination of both modelling paradigms into hybrid models (Lee *et al.* 2005; Quaghebeur *et al.* 2022). This creates a modelling paradigm that includes the best of both worlds: a mechanistic backbone incorporating relevant process knowledge and thus providing interpretability and extrapolation capabilities, as well as a data-driven part that augments the overall model's predictive power by including information on lesser-known subprocesses at reduced computational cost. An overview of different hybrid model configurations and applications for advanced control and online optimisation can be found in von Stosch *et al.* (2014).

A second solution to overcome the high computational demand of mechanistic models is the development of surrogate models. Surrogate models are simplified approximations of a more complex model of the system that statistically relate the input and output of the system. They are considerably useful when the underlying relationship between input and output of the system is unknown or in the present context computationally expensive to evaluate. Here, a highly accurate mechanistic model is used to generate a set of (simulated) data within the relevant operating space which can subsequently

be used to train a data-driven model (Chinesta *et al.* 2020). Depending on the definition, the creation of surrogate models can also be seen as hybrid modelling practice.

Regardless of the selected model to be used in a DT of a WRRF, a challenging aspect to include in a DT is the means to **dynamically update** or adjust the models based on relevant data to maintain an accurate description of the real entity as it evolves over time. Therefore, the models need to be validated frequently and if necessary recalibrated or retrained. While this can be done in a manual way, automatic validation/(re)calibration of the models is one of the key aspects of DT. This is a crucial step in developing a DT so that it can deal with the known uncertainties in WRRF modelling without the need to interrupt the operation of the DT. There are some important aspects to consider for the implementation of such a validation/(re)calibration procedure. For example, choosing the frequency for the validation/(re)calibration could be dependent on the source and frequency of the available data (e.g. laboratory data versus online sensor data), and the nature of the process and parameters to be calibrated (e.g. settling vs biological processes). Typical parameter estimation methods for model (re)calibration are optimisation methods such as recursive weighted least square methods and Bayesian estimation using a Markov chain Monte Carlo approach. However, these techniques can be time-consuming and computationally expensive. A data-driven approach, like artificial neural networks, could be an appropriate choice for model (re)calibration in real-time (Samad & Mathur 1992; de Almeida Martins *et al.* 2021). However, the performance of these models is highly dependent on the quality of the training datasets, and they may need to be retrained from time to time should completely new scenarios appear in the data patterns. Existing model calibration protocols that are recognised as part of the Good Modelling Practices (GMP) do not account for the new challenge of automatic online calibration (Rieger *et al.* 2012). This does not necessarily mean that the traditional protocols are obsolete, but rather emphasises the need for changes or adaptations based on the specific requirements of a DT compared to traditional modelling practices.

Alternatively, state estimators such as extended Kalman filters and other data assimilation methods can be included in the DT to match the model predictions to the most recent observations. In this way the model can continuously account for measurement inaccuracies and slowly changing process dynamics (Pantelides & Renfro 2013; Afshari *et al.* 2017). However, the number of states to be corrected by state estimation methods cannot exceed the number of measurements. Hence, a subset of uncertain/varying states should be selected on which online model correction through state estimators can be applied (Patwardhan *et al.* 2007).

In summary, state estimators allow for continuous model correction to account for small inaccuracies and slowly changing dynamics whereas major changes in the operating or influent space (for example a shift in microbial dynamics) require recalibration of the DT model.

5. DATA MANAGEMENT IN WRRFS

As stated in Section 2, the distinction between a DT and a conventional simulation model is primarily determined by its automated and continued connection to the real entity. This connection is achieved by (bi-directional) data transfer. Indeed, if a DT is to consistently represent the structure, state, and/or behavior of its physical counterpart, then it needs to be provided with operational data in various forms. Different categories of data usage can be identified for a DT, e.g. to calibrate and train mechanistic and data-driven models, to supply models with input data for simulation, to evaluate the state of the dynamic subsystems, to feed soft sensors, and/or to provide the necessary inputs for automated closed-loop control strategies. Ideally, both the objectives of the DT as well as the means towards its objectives, i.e. the models and control strategies, will determine which data are required. In practice, however, there exist technological and monetary constraints that limit data collection and management at WRRFs, which means data access will ultimately dictate how the DT can optimally exploit the available data. Even with limited data, a DT can add value by providing context and making the data actionable.

Fortunately, the amount and reliability of data that is being collected by WRRFs is on the rise (Olsson *et al.* 2014). New, more reliable sensor technologies (e.g. liquid phase N₂O (Fenu *et al.* 2020)) combined with faster computations are making the storage of frequent and high-resolution water quantity and quality data more affordable. For a DT focused on the operation and maintenance stage of the WRRF lifecycle, this steady increase of real-time process signals offers opportunities to determine and deliver a proactive course of action for WRRFs (GWRC 2021).

While WRRF operators and engineers are focused on the dynamic signals that relate to plant performance, at the same time, other organisational activities that support WRRFs (e.g. equipment and infrastructure maintenance, finance, resource planning, compliance, research, and development, etc.) are also generating significant amounts of digitised data. Although

the water sector is lagging behind other sectors in adopting these digital technologies (TWI2050 2019), the information gathered from digital workflows provides the context for interpretation of WRRF measurements and states. For example, dynamic time series data stored in a process historian can be routinely annotated with the rich but static and unstructured contextual data often found in operational logbooks, and computerised maintenance management systems. In this way, a necessary base for fault detection is made available. In contrast to process measurements, those metadata that relate to process equipment, infrastructure, location, maintenance history, quality, purpose, range etc., typically do not need to be updated at high frequency (IWA 2021b). However, they still need to be curated actively if reliable conclusions are to be extracted from a DT. This means that high levels of workflow organisation and automation are needed to ensure that new information generated by one party becomes rapidly available to all WRRF stakeholders. Data curation also requires that external contractors' non-editable contextual document handover (e.g. PDF files of P&IDs, screenshots of CAD models, protected automation logic, etc.) be replaced with data handover (e.g. P&IDs, CAD models and automation programs in file formats native to the design software) that seamlessly integrate into a DT (Brendelberger *et al.* 2019; Wiedau *et al.* 2019).

To understand the challenges related to data and their usage in the context of a DT for the water and wastewater industries, abstract representations of data flows can be useful. Therrien *et al.* (2020) describe the concept of a data pipeline that illustrates how raw data are transformed into intelligent action. A crucial weak spot in this flow is the need for qualitative data upon which operators and engineers can make reliable decisions, and mathematical models can be fine-tuned. The struggle for high quality data is particularly applicable to WRRFs, as these data are naturally dependent on various water quality sensors that are notorious for low veracity when exposed to the harsh wastewater medium combined with inadequate maintenance (Therrien *et al.* 2020). The COVID-19 pandemic has highlighted this problem. As on-site staff were reduced and selective maintenance programs were introduced, the quality of online sensor data collected in about one-third of treatment plants around the world decreased considerably (Rahman *et al.* 2021).

The relation between DTs and data quality can be compared to the causality dilemma of the chicken or the egg. While the quality of DT predictions is inherently influenced by the quality of the data input, the DT will also improve the quality of collected data by providing the means for automatic and live data checks and contextualization. Thorough data analysis and reconciliation to detect, isolate, identify and correct measurement faults (Olsson & Newell 1999) should therefore be included early in the conception of the DT. Given the short-term operational objectives, data reconciliation methods should be sufficiently automated to allow for quick decisions and rapid action (De Mulder *et al.* 2018). While mechanistic models such as mass balances can assist in data reconciliation (Spindler & Vanrolleghem 2012; Le *et al.* 2018), the bulk of this process will likely be performed by generically applicable data-driven methods (Corominas *et al.* 2018; Newhart *et al.* 2019). Using data-driven approaches means that periodic re-training will be needed to compensate for the time-varying nature of most environmental time series. This is preferably achieved using automatically adapting fault detection techniques (Haimi *et al.* 2013). This will lead to a need for sufficient training data so that domain experts are not overloaded with routine data labelling tasks required for supervisory model training. Instead, clever methods should be devised that exploit uncertainty in the detection to iteratively improve future model prediction performance while limiting the need for human input (Russo *et al.* 2020).

Collecting data without a proper management strategy can result in underutilised data sets that lack the necessary structure for users and/or automated tools to efficiently interact with the data (IWA 2021b). An important piece of any DT's application is therefore consistent with the vision put forward by the Findable Accessible Interoperable Reusable (FAIR) guiding principles for scientific data management and stewardship (Wilkinson *et al.* 2016). That is, data should be FAIR within the boundaries and permissions needed for the DT to operate. A key aspect of the FAIR principles is the adoption of open interfaces and protocols for authorised data access, such that vendor lock-in for both software and hardware is abolished. This philosophy for open data has recently also been adopted by some WRRF simulation software providers, which nowadays provide built-in support for open communication protocols used in industrial automation (e.g. OPC UA), connectivity to commercial cloud computing infrastructure (e.g. MS Azure), as well as application programming interfaces (API) for open-source scripting languages (e.g. Python) to name a few. The latter is especially important in the context of hybrid modelling. APIs that allow for easy coupling of data-driven modelling techniques and tools to existing simulators could facilitate the transition of the WRRF modelling community from the purely mechanistic paradigm to applications with hybrid models.

A secondary benefit that stems from FAIR data is the opportunity for data monetisation. Clean data streams can be offered via a controlled marketplace, which developers can utilise to develop new applications and business models (FIWARE 2021). Subscription-based decentralised networks, based on blockchain technology, are already available for secure data sharing

while guaranteeing fair profits to its owners (cf. Streamr, <http://streamr.network>). The opportunity for new revenue streams for WRRF utilities in a data economy may also provide additional incentive for the adoption of FAIR data principles. Moreover, opening data to the research community or the general public can help in advancing data-driven research breakthroughs by introducing solutions from entirely different fields, while also creating public awareness (Bonabeau 2009; Quay *et al.* 2021).

Underlying all the mathematical modelling and data management lies the IT software architecture and hardware infrastructure needed to support the DT. Here, choices will have to be made based on maintenance needs, user interaction, and cybersecurity of the DT. This includes choosing between on-premises software or cloud solutions, open-source or closed-source software, and off-the-shelf or custom software applications. With respect to data management, the DT will probably need to query various operational databases of a WRRF. This can include process historians, laboratory information systems (LIMS), and computerised maintenance management systems (CMMS), but also databases that are related to business activities higher up in the organisational structure like those used for finance and human resources. It is also not beyond the realm of possibility that the DT may need to access data sources external to the organisation like those providing multi-day weather forecasts, energy prices or smart city data brokers. At the same time, new types of data will have to be stored and version controlled by the DT. These include simulation results, calibration/training data sets, model performance metrics, historic parameter values, input data for the modelling of future scenarios, user interaction, and clickstream data, etc. It should thus be clear that sophisticated data architectures and computational workflows, composed of structured and unstructured data import, storage, and processing, will be needed to serve the DT. According to Fuller *et al.* (2020), the cost of installing and running such infrastructure is likely to be the biggest challenge for any DT project. Moreover, with the advent of increased internet connectivity and the introduction of IoT tools, questions need to be addressed around data sharing, cybersecurity, and ethics (GWRC 2021).

The successful adoption of open data models, FAIR data principles and powerful data architectures is of course influenced by the culture of the organisation but also of the regulatory frameworks surrounding water management. Whereas data standardisation, openness and interoperability are high on the development agenda in Europe and North America, the same cannot be said for many other regions world-wide. In such countries or regions, the first step towards DT applications should naturally be a paradigm shift from data acquisition based solely on regulatory compliance to being driven by the need for holistic and sustainable multi-criteria water management practices.

6. DIGITAL TWINS BEYOND THE FENCE OF WRRFS

DTs may be developed for a range of purposes and operate at different scales. A DT could be as simple as a component or model-based controller for a specific task. However, a DT can also be scaled up, to be plant-wide, then to the integrated urban watershed, to city-wide, and beyond. The scale on which a DT is developed and deployed is determined by the objectives. WRRF DTs could benefit from integrating data sources from outside their fence such as weather forecasts (Heinonen *et al.* 2013), energy tariffing (Aymerich *et al.* 2015), data on the ecological status of receiving waters (Muschalla 2008) etc., for decision support. Such novel sources of information provide new opportunities to optimise and manage WRRFs from a holistic perspective rather than a domain-specific one. DTs of WRRFs can not only improve treatment plant performance but can also be used to serve societal benefits. For example, the successful implementation of reuse strategies for wastewater effluent will depend on the specific composition of waste streams but also on the needs and opportunities in the surrounding urban/industrial/natural landscape, e.g. reuse of wastewater effluent for irrigation in agriculture (Neto *et al.* 2021), use in industry, aquifer replenishing for indirect potable reuse (Bullard *et al.* 2019) etc. Looking even further outside of the scope of the water sector, DTs can be used to predict the impact of a new or retrofitted wastewater treatment plant on energy systems, e.g. potential power demand for pumping and aeration, supply from onsite renewable power generation, waste heat recovery potential from treated wastewater etc.

A high-level DT would allow interdependencies across sectors to be understood in a way that organisation level or sector-based DTs could not satisfy. However, for the water sector, we are not yet there: bringing modelling tools together over different scales to address high-level technological or societal challenges is still rarely achieved and few DTs at present are connected or share data across organisations, sectors or geographies. Lack of interoperability is a key constraint. Achieving a high level DT requires effective communication between different subsystems to allow automatic reasoning and optimisation in support of decision making. Hence, it is not a matter of developing a single all-encompassing model of a region

but rather a matter of building the foundation for an ecosystem of interconnected DTs where interoperability and transparency are key through common and automated data and context information models (Bolton *et al.* 2018).

Exchange of data from potentially different domains, each characterised by a specific jargon, leads to the need for standardised data structuring and communication conventions. Several initiatives are emerging aimed to develop semantic information models or ontologies consisting of controlled vocabularies, relationships, constraints, and rules to provide organised, stable, and shareable data for specific domain contexts which will significantly increase the accessibility and transferability of data. Examples can be found in the domains of geospatial sciences (OGC 2021), public health environmental surveillance (PHES-ODM 2021), chemical process industry (DEXPI 2021), production systems engineering (AutomationML 2021), building information management (ISO 2018). Within the specific context of smart cities, the Open and Agile Smart Cities initiative (<https://oascities.org/>) has defined minimal interoperability mechanisms (MIMS) including common data models (DTDLE 2021; FIWARE 2021) to ensure cross-sectorial communication of data and models. The goal of each of these initiatives is to capture and map complex relationships that exist in the physical world and translate them into the digital data world using a set of concepts and categories defined by the domain ontology.

Some first examples of high-level DTs based on common data and information models can be found in literature of other domains: the application of a decision support system for real-time air quality control in the city of Singapore by coupling weather predictions with combustion processes and city topology models (Farazi *et al.* 2020) or even the construction of a national DT for energy management in the UK (Akroyd *et al.* 2021).

7. THE HUMAN FACTOR IN A DIGITAL WORLD

Effective use of a DT will require a significant cultural change across an organisation. WRRFs that are mostly run from a reactive stance will shift to proactive understanding based on a stream of (near) real-time information. This demands that the entire workforce understands the value of the data they generate or maintain with respect to other activities that exist along an organisation's data pipeline.

In most ongoing DT projects in the WRRF sector, the primary user of the DT will typically be an operations staff member who is accustomed to having to make day-to-day decisions, based on sometimes incomplete and/or delayed information (e.g. laboratory tests). A DT opens a whole new spectrum of potentially relevant information. Hence, it is important that the DT is designed taking into account the needs of frontline staff, i.e. to present the right information in the right format for operational decision-making. If it can make their life easier in some significant way, without adding more work, WRRF operators are likely to adopt this technology enthusiastically. However, it is important to understand that frontline staff have varying degrees of digital literacy, i.e. familiarity and comfort with digital tools and workflows. Moreover, users of any DT will need to gain trust in the suggestions provided by the DT, especially in the case when these suggestions are directly used to control the plant's operation.

It is up to each organisation's leaders to facilitate this novel synergy between human and DT. Involving frontline staff in the development of the DT content and design and giving staff an ownership stake in the success of the technology is critical. This includes education and training, such that it is very clear what the DT can and cannot do. It is not uncommon for a layman to think that a DT can do everything (Fuller *et al.* 2020), but when something does not match these expectations, the entire DT is branded a failure, a situation that is hard for an organisation to recover from. Therefore, it is also advised that any new DT goes through a proving stage to give all staff confidence in its suggestions and/or actions.

Eerikäinen *et al.* (2020) interviewed employees of WRRFs and other related stakeholders about their expectations of new data tools. It was found that they were expecting a next generation of tools in the near future – tools that combine the competencies of both automation/software providers and wastewater process experts with a thorough understanding of process behaviour. It is clear that a DT makes use of relatively novel technologies stemming from a wide range of technical information and engineering domains. Having in-house access to all this expertise is highly unlikely for most water and wastewater utilities. New hiring policies and lifelong learning initiatives can help utilities overcome some of the knowledge and skill gaps but implementation of the DT will most likely be performed to some degree in partnership with consultancy firms and academic partners. Moreover, some of the data interpretation within the DT may also be outsourced. Subscription platforms for the external, certified, evaluation of specific measurements already exist to date (e.g. vibration and acoustic signal analysis for condition monitoring of critical assets) (cf. Zensor, <https://www.zensor.be/>). By using IoT devices, data are sent directly to cloud applications of service providers for specialised analysis and subsequent presentation to non-experts.

Eventually, the DT could tap into this external knowledge base to provide even more relevant and accurate predictions. While depending on external parties may provide answers, it also comes at additional costs. A workaround that is regarded as a key enabler for DTs are so-called low-code and no-code application development platforms (Michael & Wortmann 2021). By providing modular preconfigured building blocks with standardised communication interfaces in a user friendly, often graphical environment, the complex computational workflows at the basis of the DT can become manageable by the non-experts themselves which reinforces their degree of ownership of and trust in the DT technology.

Lastly, it is important to recognise the potential of young water professionals in the digital transformation of the water sector, including the introduction of DTs. With estimates on water workforce retirement in the next 10 years ranging from 30 to 50% in the USA, a new wave of millennials is expected to reinforce the sector (Dickerson & Butler 2018). Raised in an online world, these digital natives have (un)consciously created a distinctive way to look for information and solve challenges faced on the modern work floor and beyond (IWA 2021b). Of course, the continued beneficial use of a DT depends on accurate documentation, a maintenance program and succession plans should a company's champions leave. However, the fact that the DT itself ideally acts as a high fidelity knowledge repository composed of models and data on the WRRF past, present, and future can contribute significantly in the transfer of knowledge from experts to inexperienced newcomers. Hence the DT could also help mitigate knowledge loss as a result of the ageing workforce (Kadiyala & Macintosh 2018).

8. CONCLUSIONS

In many ways, the advent of DTs has the potential to entirely change the way utilities operate and design their facilities. The existence of a live, automated data connection between the DT and the real entity ensures that data is actionable and knowledge is up to date. As such, utilities can transition from reactive to proactive and holistic management. However, accomplishing such aspirational goals requires a combination of technological advances and a clear buy-in (through clear social and financial benefits) from relevant stakeholders involved.

Despite growing interest, successful full-scale applications in the WRRF sector are still rare. The current focus of DT development in the WRRF sector lies mostly with online operational support and real-time advanced control. The potential of other applications that are well documented in other sectors, such as asset management, retrofitting, etc. should however not be overlooked.

The predictive power of the model used in a DT as well as its ability to adapt and respond to new scenarios and disturbances is one of the key aspects required to build trust among operational staff and utility managers. This concerns not only the application of DTs but is also important for keeping them alive and relevant in the long-term. Hence, developing a model that is highly predictive with a robust validation/(re)calibration protocol on top is critical. Hybrid models can be important tools in boosting predictive power by balancing mechanistic models with data-driven techniques. Their application and development for WRRF processes can encourage and amplify the successful implementation of DTs. Moreover, the data connection of the DT to its physical counterpart is crucial to ensure a continuous feed of high-quality data to the modelling core of the DT. A proper data architecture and data management strategy as well as automated data analysis and reconciliation are essential for any DT project. DT projects should hereby follow the philosophy put forward by the Findable Accessible Interoperable Reusable (FAIR) guiding principles for data management.

The development of DTs opens up an enormous potential for optimisation outside the fence of a WRRF. Open data models and standards allow connection of WRRF DTs to other smart city data sources. Thus, to foster and stimulate truly holistic decision making in the water sector, any DT should be considered as part of a modular inter-connected structure using common and automated data and context information models supporting minimum interoperability mechanisms (MIMS) that allow automated reasoning over their combined structure.

Finally, the expectations of the DT end users should match the capabilities and limitations of the models. No technology is useful unless people want to use it and trust it. This means different things to different stakeholders. The organisational structure around the DT should account for these differences. This emphasises the importance of including WRRF staff during the development of the DT.

ACKNOWLEDGMENTS

This position paper gives an overview of the discussion that took place at the similarly titled workshop of the (virtual) 7th IWA/WEF Water Resource Recovery Modelling Seminar WRRmod2021. It is the result of a collective effort by industry

practitioners and academics who are gratefully acknowledged for all their input. The work of Niels Nicolai and Peter Vanrolleghem was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2021-04347 towards digital twin based control of water resource recovery facilities – Methods supporting the use of adaptive hybrid digital twins.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Afshari, H. H., Gadsden, S. A. & Habibi, S. 2017 Gaussian filters for parameter and state estimation: a general review of theory and recent trends. *Signal Processing* **135**, 218–238. doi:10.1016/j.sigpro.2017.01.001.
- Akroyd, J., Mosbach, S., Bhave, A. & Kraft, M. 2021 Universal digital twin – A dynamic knowledge graph. *Data-Centric Engineering* **2** (4). doi:10.1017/dce.2021.10.
- Alex, J., Hübner, C. & Förster, L. 2020 Planning, testing and commissioning of automation solutions for wastewater treatment plants using simulation. *IFAC-Papers OnLine* **53**, 16665–16670. doi:10.1016/j.ifacol.2020.12.1084.
- AutomationML 2021 Available from: <https://www.automationml.org/> (Accessed December 29, 2021).
- Aymerich, I., Rieger, L., Sobhani, R., Rosso, D. & Corominas, L. 2015 The difference between energy consumption and energy cost: modelling energy tariff structures for water resource recovery facilities. *Water Research* **81**, 113–123. doi:10.1016/j.watres.2015.04.033.
- Bartos, M. & Kerkez, B. 2021 Pipedream: an interactive digital twin model for natural and urban drainage networks. *Environmental Modelling and Software* **144**, 105–120. doi:10.1016/j.envsoft.2021.105120.
- Bolton, A., Butler, L., Dabson, J., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., Luck, A., Pawsey, N., Saville, S., Schooling, J., Sharp, M., Smith, T., Tennison, J., Whyte, J., Wilson, A. & Makri, C. 2018 *The Gemini Principles: Guiding Values for the National Digital Twin and Information Management Framework*. Centre for Digital Built Britain and Digital Framework Task Group. doi:10.17863/CAM.32260.
- Bonabeau, E. 2009 Decisions 2.0: the power of collective intelligence. *MIT Sloan Management Review* **50** (2), 45–52.
- Brendelberger, M., Christmann, U., Dubovy, M., Gundlach, C. S., Hartmann, W., Hoelzke, U., Landa, D. V., Mertens, M., Modersohn, A., Nothdurft, L., Rahm, J., Schüller, A., Temmen, H. & Zgorzelski, P. 2019 *Digital Twin in Process Industry, NAMUR – Interessengemeinschaft Automatisierungstechnik der Prozessindustrie E.V.* Technical Report. Leverkusen, Germany, p. 4.
- Bullard, M. G., Widdowson, M., Salazar-Benites, G., Heisig-Mitchell, J., Nelson, A. & Bott, C. 2019 Managed aquifer recharge: transport and attenuation in a coastal plain aquifer. In: *World Environmental and Water Resources Congress 2019: Groundwater, Sustainability, Hydro-Climates/Climate Change, and Environmental Engineering* (Bullard, M. G. ed.). American Society of Civil Engineers, Reston, VA, pp. 108–120.
- Chinesta, F., Cueto, E., Abisset-Chavanne, E., Duval, J. L. & El Khaldi, F. 2020 Virtual, digital and hybrid Twins: a new paradigm in data-based engineering and engineered data. *Archives of Computational Methods in Engineering* **27**, 105–134. doi:10.1007/s11851-018-9301-4.
- Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. & Poch, M. 2018 Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques. *Environmental Modelling & Software* **106**, 89–103. doi:10.1016/j.envsoft.2017.11.023.
- Curl, J. M., Nading, T., Hegger, K., Barhoumi, A. & Smoczynski, M. 2019 Digital twins: the next generation of water treatment technology. *Journal American Water Works Association* **111**, 44–50. doi:10.1002/awwa.1413.
- de Almeida Martins, J. P., Nilsson, M., Lampinen, B., Palombo, M., While, P. T., Westin, C.-F. & Szczepankiewicz, F. 2021 Neural networks for parameter estimation in microstructural MRI: application to a diffusion-relaxation model of white matter. *NeuroImage* **244**. doi:10.1016/j.neuroimage.2021.118601.
- 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 & Software* **107**, 186–198. doi:10.1016/j.envsoft.2018.05.015.
- DEXPI – Data Exchange in the Process Industry 2021 Available from: <https://dexpi.org/> (accessed December 29, 2021).
- Dickerson, S. T. & Butler, A. 2018 Resolve workforce challenges to ensure future success at water and wastewater utilities. *Opflow* **44**, 8–9. doi:10.1002/opfl.1063.
- DTDL – Digital Twin Definition Language 2021 Available from: <https://github.com/Azure/opendigitaltwins-smartcities> (accessed December 29, 2021)

- 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* **82** (12), 2681–2690. doi:10.2166/wst.2020.311.
- Farazi, F., Salamanca, M., Mosbach, S., Akroyd, J., Eibeck, A., Aditya, L. K., Chadzynski, A., Pan, K., Zhou, X., Zhang, S., Lim, M. Q. & Kraft, M. 2020 Knowledge graph approach to combustion chemistry and interoperability. *ACS Omega* **5** (29), 18342–18348.
- Fenu, A., Wambeq, T., de Gussem, K. & Weemaes, M. 2020 Nitrous oxide gas emissions estimated by liquid-phase measurements: robustness and financial opportunity in single and multi-point monitoring campaigns. *Environmental Science and Pollution Research* **27**, 890–898. doi:10.1007/s11356-019-07047-0.
- FIWARE – The Open Source Platform for Our Smart Digital Future 2021 Available from: <https://www.fiware.org/> (accessed December 29, 2021).
- Fuller, A., Fan, Z., Day, C. & Barlow, C. 2020 Digital twin: enabling technologies, challenges and open research. *IEEE Access* **8**, 108952–108971. doi:10.1109/ACCESS.2020.2998358.
- Glaessgen, E. & Stargel, D. 2012 The digital twin paradigm for future NASA and U.S. Air force vehicles. In *Proceedings of 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference – Special Session on the Digital Twin*.
- Grieves, M. 2014 *Digital Twin: Manufacturing Excellence Through Virtual Factory Replication*. NASA, Washington, DC, USA, White Paper 1.
- GWRC 2021 *The Digital Water Utility of the Future: Whitepaper on the Enablers, Applications and Risks of Digitalisation for the Global Water Research Coalition*. Technical Report. Unley, Australia, p. 37.
- Haimi, H., Mulas, M., Corona, F. & Vahala, R. 2013 Data-derived soft-sensors for biological wastewater treatment plants: an overview. *Environmental Modelling & Software* **47**, 88–107. doi:10.1016/j.envsoft.2013.05.009.
- Heinonen, M., Jokelainen, M., Fred, T., Koistinen, J. & Hohti, H. 2013 Improved wet weather wastewater influent modelling at viikinkmäki WWTP by on-line weather radar information. *Water Science and Technology* **68** (3), 499–505. doi:10.2166/wst.2013.213.
- ISO – International Organisation for Standardization 2018 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling – Part 1: Concepts and principles – ISO 19650-1:2018. (accessed December 29, 2021).
- IWA 2021a *Digital Water: Operational Digital Twins in the Urban Water Sector*. White Paper. London, UK, p. 17.
- IWA 2021b *Digital Water: The Value of Meta-Data for Water Resource Recovery Facilities*. White Paper. London, UK, p. 15.
- IWA & Xylem Inc 2019 *Digital Water Report*. Technical Report. London, UK, p. 44.
- Jain, P., Poon, J., Singh, J. P., Spanos, C., Sanders, S. R. & Panda, S. K. 2020 A digital twin approach for fault diagnosis in distributed photovoltaic systems. *IEEE Transactions on Power Electronics* **35** (1), 940–956. doi: 10.1109/TPEL.2019.2911594.
- Johnson, B. R., Kadiyala, R., Owens, G., Mak, Y. P., Grace, P., Newbery, C., Sing, S., Saxena, A. & Green, J. 2021 Water reuse and recovery facility connected digital twin case study: Singapore PUB's Changi WRP process, control, and hydraulics digital twin. In: *Proceedings of WEFTEC 2021*, Chicago, USA.
- Jones, D., Snider, C., Nassehi, A., Yon, J. & Hicks, B. 2020 Characterising the digital twin: a systematic literature review. *CIRP Journal of Manufacturing Science and Technology* **29**, 36–52. doi:10.1016/j.cirpj.2020.02.002.
- Kadiyala, R. & Macintosh, C. 2018 *Leveraging Other Industries – Big Data Management (Phase I)*. Available from: <https://www.waterrf.org/>
- Karmous-Edwards, G., Conejos, P., Mahinthakumar, K., Braman, S., Vicat-Blanc, P. & Barba, J. 2019 *Smart Water Report 2019: Foundations For Building A Digital Twin For Water Utilities*. Available from: www.swan-forum.com.
- Le, Q. H., Verheijen, P. J. T., van Loosdrecht, M. C. M. & Volcke, E. I. P. 2018 Experimental design for evaluating WWTP data by linear mass balances. *Water Research* **142**, 415–425. doi:10.1016/j.watres.2018.05.026.
- Lee, D. S., Vanrolleghem, P. A. & Park, J. M. 2005 Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of Biotechnology* **115**, 317–328. doi:10.1016/j.jbiotec.2004.09.001.
- Le Moullec, Y., Potier, O., Gentric, C. & Leclerc, J. P. 2011 Activated sludge pilot plant: comparison between experimental and predicted concentration profiles using three different modelling approaches. *Water Research* **45**, 3085–3097. doi:10.1016/j.watres.2011.03.019.
- Maktabifard, M., Zaborowska, E. & Makinia, J. 2019 Evaluating the effect of different operational strategies on the carbon footprint of wastewater treatment plants – case studies from northern Poland. *Water Science and Technology* **79** (11), 2211–2220. doi:10.2166/wst.2019.224.
- Mamais, D., Noutsopoulos, C., Dimopoulou, A., Stasinakis, A. & Lekkas, T. 2015 Wastewater treatment process impact on energy savings and greenhouse gas emissions. *Water Science and Technology* **71** (2), 303–308. doi:10.2166/wst.2014.521.
- Michael, J. & Wortmann, A. 2021 Towards development platforms for digital twins: A model-driven low-code approach. In: *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. APMS 2021. IFIP Advances in Information and Communication Technology*, Vol. 630 (Dolgui, A., Bernard, A., Lemoine, D., von Cieminski, G. & Romero, D. eds.). Springer, Cham. Available from: https://doi-org.acces.bibl.ulaval.ca/10.1007/978-3-030-85874-2_35
- Muschalla, D. 2008 Optimization of integrated urban wastewater systems using multi-objective evolution strategies. *Urban Water Journal* **5** (1), 59–67. doi: 10.1080/15730620701726309.
- Neto, O. B. L., Haddon, A., Aichouche, F., Harmand, J., Mulas, M. & Corona, F. 2021 Predictive control of activated sludge plants to supply nitrogen for optimal crop growth. *IFAC-PapersOnLine* **54** (3), 200–205. doi:10.1016/j.ifacol.2021.08.242.
- Newhart, K. B., Holloway, R. W., Hering, A. S. & Cath, T. Y. 2019 Data-driven performance analyses of wastewater treatment plants: a review. *Water Research* **157**, 498–513. doi: 10.1016/j.watres.2019.03.030.
- OGC – Open Geospatial Consortium 2021 Available from: <http://www.opengeospatial.org/> (accessed December 29, 2021).
- Olsson, G. & Newell, B. 1999 *Wastewater Treatment Systems*. IWA publishing, London, UK.

- Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K. V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodríguez-Roda, I., Steyer, J.-P., Takács, I., Vanrolleghem, P. A., Vargas, A., Yuan, Z. & Åmand, L. 2014 Instrumentation, control and automation in wastewater – from London 1973 to Narbonne 2013. *Water Science and Technology* **69** (7), 1373–1385. doi:10.2166/wst.2014.057.
- Pantelides, C. & Renfro, J. 2013 The online use of first-principles models in process operations: review, current status and future needs. *Computers & Chemical Engineering* **51**, 136–148. doi:10.1016/j.compchemeng.2012.07.008.
- Patwardhan, S. C., Prakash, J. & Shah, S. L. 2007 Soft sensing and state estimation: review and recent trends. *IFAC Proceedings Volumes* **40** (19), 65–72. doi: 10.3182/20071002-MX-4-3906.00012.
- Pedersen, A. N., Borup, M., Brink-Kjær, A., Christiansen, L. E. & Mikkelsen, P. S. 2021 Living and prototyping digital twins for urban water systems: towards multi-purpose value creation using models and sensors. *Water* **13**, 592. doi:10.3390/W13050592.
- PHES-ODM – The Public Health Environmental Surveillance Open Data Model 2021 Available from: <https://github.com/Big-Life-Lab/PHES-ODM> (accessed December 29, 2021)
- Poch, M., Garrido-Baserba, M., Corominas, L., Perelló-Moragues, A., Monclús, H., Cermerón-Romero, M., Melitas, N., Jiang, S. C. & Rosso, D. 2020 When the fourth water and digital revolution encountered COVID-19. *Science of the Total Environment* **744**, 140980. doi:10.1016/j.scitotenv.2020.140980.
- Quaghebeur, W., Torfs, E., Nopens, I. & De Baets, B. 2022 Hybrid differential equations: integrating mechanistic and data-driven techniques for modelling of water systems. *Water Research* **213**, 118166. doi.org/10.1016/j.watres.2022.118166.
- Quay, A., Fiske, P. S. & Mauter, M. S. 2021 Recommendations for advancing FAIR and open data standards in the water treatment community. *ACS EST Engg.* doi:10.1021/acsestengg.1c00245.
- Rahman, A., Belia, E., Kirim, G., Hasan, M., Borzooei, S., Santoro, D. & Johnson, B. 2021 Digital solutions for continued operation of WRRFs during pandemics and other interruptions. *Water Environment Research* **93** (11), 2527–2536. doi:10.1002/wer.1615.
- Rieger, L., Gillot, S., Langergraber, G., Ohtsuki, T., Shaw, A., Takács, I. & Winkler, S. 2012 *Guidelines for Using Activated Sludge Models*. Scientific and Technical report. IWA Publishing, London, UK.
- Ruohomäki, T., Airaksinen, E., Huuska, P., Kesäniemi, O., Martikka, M. & Suomisto, J. 2018 Smart city platform enabling digital twin. In: *Proceedings International Conference on Intelligent Systems (IS)*. IEEE, pp. 155–161. doi: 10.1109/IS.2018.8710517.
- Russo, S., Lürig, M., Hao, W. J., Matthews, B. & Villez, K. 2020 Active learning for anomaly detection in environmental data. *Environmental Modelling & Software* **134**, 104869. doi:10.1016/j.envsoft.2020.104869.
- Samad, T. & Mathur, A. 1992 Parameter estimation for process control with neural networks. *International Journal of Approximate Reasoning* **7** (3–4), 149–164. doi:10.1016/0888-613X(92)90008-N.
- Spindler, A. & Vanrolleghem, P. A. 2012 Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Science and Technology* **65** (12), 2148–2153. doi:10.2166/wst.2012.125.
- Stentoft, P. A., Munk-Nielsen, T., Vezzaro, L., Madsen, H., Mikkelsen, P. S. & Møller, J. K. 2018 Towards model predictive control: online predictions of ammonium and nitrate removal by using a stochastic ASM. *Water Science and Technology* **79** (1), 51–62. doi:10.2166/wst.2018.527.
- 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. doi:10.2166/WST.2020.266.
- Stentoft, P. A., Munk-Nielsen, T., Møller, J. K., Madsen, H., Valverde-Pérez, B., Mikkelsen, P. S. & Vezzaro, L. 2021 Prioritize effluent quality, operational costs or global warming? – using predictive control of wastewater aeration for flexible management of objectives in WRRFs. *Water Research* **196**, 116960. doi:10.1016/j.watres.2021.116960.
- Therrien, J. D., Nicolai, N. & Vanrolleghem, P. A. 2020 A critical review of the data pipeline: how wastewater system operation flows from data to intelligence. *Water Science and Technology* **82**, 2613–2634. doi:10.2166/WST.2020.393/778299/WST2020393.
- TWI2050 - The World in 2050 2019 *The Digital Revolution and Sustainable Development: Opportunities and Challenges. Report Prepared by the World in 2050 Initiative*. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. Available from: www.twi2050.org.
- von Stosch, M., Oliveira, R., Peres, J. & Feyo de Azevedo, S. 2014 Hybrid semi-parametric modeling in process systems engineering: past, present and future. *Computers and Chemical Engineering* **60**, 86–101. doi:10.1016/j.compchemeng.2013.08.008.
- Wiedau, M., von Wedel, L., Temmen, H., Welke, R. & Papakonstantinou, N. 2019 ENPRO data integration: extending DEXPI towards the asset lifecycle. *Chemie Ingenieur Technik* **91**, 240–255. doi:10.1002/cite.201800112.
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Niklas Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L. O., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, F., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. 2016 The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3**, 160018. doi:10.1038/sdata.2016.18.
- Wright, L. & Davidson, S. 2020 How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences* **7**, 1–13. doi:10.1186/S40323-020-00147-4.