Building and Exploiting a Digital Twin for the Management of Drinking Water Distribution Networks

Conejos, P.*1, Martínez Alzamora, F.2, Hervás Carot, M.3, Alonso Campos, J.C.4

1 Assistant Prof. Dep. Hydraulic Engineering. Universitat Politècnica de València. 46022, Valencia (Spain)

2 Professor, Res. Inst. of Water and Env. Engineering (IIAMA). Universitat Politècnica de València. 46022

3,4 PhD Student. Dep. Hydraulic Engineering. Universitat Politècnica de València. 46022, Valencia (Spain)

1,3 Dep. of Network Control and Operation, EMIMET, C/ Pedrapiquers, 4, Valencia (Spain)

4 R&D Department, Global Omnium, C/ Pedrapiquers, 4, Valencia (Spain)

1 pconejos@dihma.upv.es (corresponding author), 2 fmartine@hma.upv.es
Building and Exploiting a Digital Twin for the Management of Drinking Water Distribution Networks

Digital Twins (DTs) are starting to be exploited in water distribution systems as a tool to improve their management and, in the future, they will be crucial for decision making. In this paper, the authors propose several requirements that a DT of a water distribution system should accomplish. However, the development of a DT is a challenge, and a continuous process of adjustments and learning is required. Due to the advantages of having a DT of the water distribution system, during the last years a strategy to build and maintain a DT of the water distribution network of Valencia (Spain) and its Metropolitan Area (1.6 million inhabitants) was developed. This is one of the first DTs of a water utility. The great benefits of their use in the daily operation of the system ensure that they will begin to be habitual in the most advanced smart cities.

Keywords: WDS, real time, detailed model, digital twin, Epanet

Introduction

There is a growing need for more sustainable and efficient nowadays cities, due to their increasing population and high resource consumption (Mohammadi and Taylor 2018). In that context, informed decision making should employ advanced tools that combine real-time sensor data gathering, statistical analysis and model-based capabilities to simulate “what-if” scenarios. As defined by various authors, a “Digital” or “Virtual” Twin of a defined system would fulfil those requirements.

The Digital Twin philosophy has been used since the 1960s by the National Aeronautics and Space Administration (NASA) because of the need to remotely operate and maintain the systems, although the Twin was then physical rather than virtual. As an example of its usefulness, the NASA could successfully perform the rescue mission of the Apollo XIII thanks to its twin on earth.

The concept of Digital Twin (DT) was definitively consolidated by M. Grieves in 2003 (Grieves 2018). It was first aimed at optimizing the life cycle of a product, in
relation to its design and manufacture process, and with the subsequent maintenance during its lifetime. According to Grieves, a Virtual Twin model must contain three main parts: the physical assets, a virtual model, and the connections of data and information that tie the virtual and real spaces together.

Since then to nowadays, numerous researchers have defined in more detail the theoretical foundations of DTs applied to the industry field (Tao et al. 2019). Most of the works highlight the following key technologies involved in building a DT: a) Modelling: physical and virtual models must describe the main features of the system. b) Connection: the physical and virtual system should be constantly connected. This theory includes data transmission, conversion, storing, protection, etc. c) Data mining: in order to get information from a dataset, it must be pre-processed (cleaned and filtered) and mined by data analysis techniques and Artificial Intelligence (AI) algorithms. d) Interaction and service: the DT should be able to suggest, operational settings, once validated by simulations, to optimize or to adapt the system processes to external changes.

DTs are not only useful in industry, but they can also be developed and exploited in a city management context, particularly in the drinking water supply system. Hydraulic models are more and more necessary and commonly exploited for decision-making by Water Utilities (WU). Water Distribution Systems (WDSs) have become complex hard-to-manage systems, since WU must deal with challenges like the population growth, the resource scarcity or the infrastructure wear and tear. Therefore, WDSs are also on the scope of the new technologies (Sarni et al. 2019), and the Digital Twin concept is starting to be used in order to solve many of the problems that concern the management of those systems (Shi, Xu, and Du 2019), like the asset management, the leak location, the optimization of the system operation, the energy efficiency, the water quality, the planning of the maintenance operations, the early response to emergencies, etc. In this
The DT will be an essential support system in the near future for the decision-making (Chacón Ramírez, Albarrán, and Cruz Salazar 2019).

There are few examples of precursors of a DT in WDSs. (Lin, Sedigh, and Miller 2009) proposed a preliminary Cyber-Physical System (CPS) combining the Epanet hydraulic simulator (Rossman 2000) with the mathematical Matlab software (Pratap 2002) to build an intelligent cyber WDS that can operate certain valves to optimally allocate demands with the goal of distributing the water as equitably as possible. The authors present it as a basic example of an intelligent decision support system.

Wang et al. (2015) focused their work in the definition of a CPS for the management of the water quality main parameters. According to Wang et al., the CPS needs: a) sensors and an effective communication network, b) computing technologies to build models, manage the data and perform data analysis, and c) predictive control technologies.

Kartakis, Abraham, and McCann (2015) built a small testbed able to simulate a real WDS, considering multiple failure and anomaly scenarios, like pipe bursts or communication issues. The testbed included real-time control algorithms fed from sensors data and aimed at energy optimization, automatic control and event-driven control.

The aforementioned studies share common attributes with DTs as defined in the industrial sector. One of the objectives of the present work is to propose a detailed description of the structure and the set of features that define a WDS Digital Twin, which must replicate the real system behaviour in a virtual model that serves as the basis for experimentation, that is, a virtual copy of reality that allows to perform any simulation. In fact, for WDSs, a complete DT would be a virtual replica of all the processes taking place from the raw water catchment until the users tap.
In this way, a DT would allow monitoring the whole process. It should analyse the performance of every phase by simulating the system behaviour under any scenario and request. It would also permit to improve the system performance and efficiency thanks to the ability to make informed and optimized decisions and, therefore, reduce risks.

However, this work is focused on the development and use of a DT just for the water distribution stage, where it is used in order to solve many of the problems concerning the distribution network management, from the design to the daily operation. The main capabilities of the proposed DT are the following:

- Optimal design of the network elements, such as pipes, valves, tanks or pumping stations, with the goal of minimizing the overall carbon footprint and subject to quality of service constraints, taking into account the demand forecast and the expected future extension of the network.
- Asset management, which combined with given design criteria will determine the optimal strategy for the renewal of the physical elements of the network and the strategic operations to be done when integrating the new assets in the system.
- Model based leak detection and pre-location, which will significantly reduce the leaks lifetime. Hence, the DT will help to preserve such a scarce resource, as well as to reduce operational costs.
- Determination of the optimal daily operation parameters with multiple objectives. These will include water velocity in pipes, service pressure and energy efficiency, among others.
- Early warning and informed response to emergencies, which will help the operators to make quick and effective decisions.
- Water quality prediction combined to online measurements of the main parameters will allow to detect quality weaknesses throughout the network and to find the best strategy for their improvement.

The next sections describe in more detail the requirements and capabilities that a Digital Twin must satisfy in order to accurately replicate the WDN behaviour, some practical issues about its implementation and a real use case of DT in the WDN of Valencia (Spain) and its Metropolitan Area.

Materials and methods

**WDS Digital Twin requirements**

As highlighted above, although there are many publications about the advantages of digital twins in general, few of them concern to WDSs. In fact, there is not a precise definition of the features that a digital twin of a WDS should have in order to differentiate it from other similar applications. From the authors point of view, a true digital twin of a WDS, according to the given definition and in order to be useful, must meet the following requirements:

- It must contain with absolute accuracy and fidelity all the elements that are part of the distribution network, and that can affect both the water flow regimes and the maintenance operations.
- It must accurately reproduce the behaviour of all the network elements, even in extreme operating conditions.
- The digital twin should also be able to exactly simulate all the controls programmed in the local automata, including the emergency response.
- Also, it must be able to reproduce the network behaviour during all the maintenance operations, including the tank cleaning, equipment tests, valve closure due to repairs, pipe filling and emptying conditions, etc.

- It must reproduce the demands on the house connections as accurately as possible. In this way, the consumptions should be registered with the highest available temporal and spatial resolution, being the last target to register all hourly-metered consumptions without delay.

- The digital twin should mimic the observed real behaviour of the network as accurate as possible, according to the data registered by all available sensors about levels, pressures, flows, water quality, etc. With this purpose, the model must be perfectly calibrated and must be able to reproduce all the control operations performed on the network, either manual or programmed.

- The last important aspect to highlight concerns to the reliability of the measurements collected by the sensors, which will be used to assess the degree of adjustment of the twin to reality. This reliability of measurements is also fundamental in real time decision making, and especially in alarm detection. A powerful anomaly detector must be developed in parallel, provided with tools to correct faulting of wrong values, based on the historic records and patterns, in order to feed properly the digital twin.

Nowadays it is possible to develop a Digital Twin because there is a lot of digital information that resides in different systems. Internet of Things (IoT) technologies have allowed the installation of sensors in isolated areas, where previously infeasible, thanks to the low energy consumption of the ultra-narrow band based IoT devices. Moreover, the ultra-narrow band is only one of the communication technologies available today. Among the set of Information and Communication Technologies (ICTs), data can be
retrieved by multiple means, including wired or wireless communications, internet based or radio based, etc. However, this information, in order to be useful, must be filtered, normalized and properly integrated, so a big data platform is necessary to develop a Digital Twin.

The main data sources managed by Digital Twins are the Geographical Information Systems (GIS), the Automated Meter Readings (AMR), the Computerized Maintenance Management Systems (CMMS), and the field data stored by Supervisory Control And Data Acquisition (SCADA) systems. In order to keep the Digital Twin always ready to use, it is important to be aware on the frequency of change of the data from the different sources. For instance, information from GIS should be updated every week at least, the AMR should desirable be hourly-based, CMMS must be imposed at the time they occur, while SCADA data should be updated every 5 minutes, or even every minute when operating in real time.

Also, in order to be able to simulate any scenario, the Digital Twin must work upon a hydraulic model that provides coherent and accurate results of the simulation. The hydraulic model must be accurately developed and calibrated, and as a part of the DT, it must be live and permanently updated, so the hydraulic model must be fed in a continuous way from all the information sources.

Finally, the DT not only has to be able to simulate hypothetical scenarios, but it also must include data analysis and optimization tools. Machine Learning algorithms, such as decision trees, support vector machines or artificial neural networks, can provide the capability to model complex sub-systems not included in the hydraulic mathematical model, particularly they can be used for demand forecasting or data filtering. Moreover, the DT can include optimization algorithms, i.e. linear and non-linear programming,
metaheuristics, etc., to generate optimal operation rules or optimal design based on different criteria (energy, water quality, quality of service, etc.).

Therefore, integrating the hydraulic model with all the information sources, and using artificial intelligence algorithms, a DT is obtained and all its potential can be exploited (Figure 1).

Figure 1. WDS Digital Twin Structure.

WDS Digital Twin implementation

Getting to these levels of detail is often a challenge, so the DT is not developed overnight. It consists of a continuous process of adjustments and learning, which enriches the hydraulic model with day-to-day knowledge, making use of the machine learning or deep learning techniques supported by large amount of field data stored in big-data platforms.

According to the previous requirements, the authors have designed a strategy to develop and maintain a Digital Twin of the water distribution network, integrating the hydraulic model with all the information sources (Bartolín et al. 2006) (Figure 2).
The application GO2HydNet has been developed to obtain from the different information sources an Epanet-based detailed model for a selected area and a certain time period, which can include all the pipes, operating elements and auxiliary elements that affect the water flows, such as tanks, pumps, air valves, drains, hydrants and house connections.

The application can rebuild the topological relations of all read elements by means of their coordinates. It can also import the pumps and valves as small linear elements with their own output and input nodes. Furthermore, the application contains tools to detect any error, both regarding the elements’ connectivity and their properties declaration. In order to solve those errors while keeping the process alive, GO2HydNet can be fed by auxiliary information, which contain instructions to add, modify or remove any element retrieved from the GIS database.

The application has also additional tools focused on the retrieved data management. In order to avoid redundancy of information, it can remove aligned vertices
and merge consecutive pipe segments if their diameter, material and age coincide. Moreover, in case the GIS contained branching without pipe splitting, the application can create T-shaped connections. Finally, GO2HydNet is able to detect isolated areas and to remove them if neglectable due to their size.

Finally, additional properties not defined in the GIS can be estimated and assigned to the model elements. For example: the pipes roughness is calculated by means of their material and age; the nodes elevation are obtained by interpolation of the Digital Terrain Model (DTM), freely available at the Spanish National Geographic Institute; the pipes status can be established depending on the status of the superposed manual valves; the flow and pressure meters can be linked to their corresponding pipe and node respectively; and the District Metered Areas (DMAs) can be identified by exploring the network and defining the DMA boundaries where flowmeters and closed manual valves exist.

Once completed the building of the physical model, GO2HydNet implements the processes to load the model by using the measured field data of a defined time period (Bou et al. 2006). First, the node demands must be assigned. In case the use consumptions are remotely available thanks to the smart metering, the values can be directly applied to the house connections (if considered in the model, otherwise they must be applied to the nearest nodes), defining a different pattern to each node. The model demands can also be simplified, establishing a single demand pattern for each DMA. In such case, the pattern is obtained from the DMA’s inlet flows.

Second, the information stored at the CMMS is retrieved to check if there were isolated zones due to maintenance operations during the queried period. In such case, the corresponding pipes statuses of the model are accordingly modified to reproduce the real topology of the network.
Finally, the application retrieves the field data from the SCADA system corresponding to the queried time period. These data permits, on the one hand, to adjust the model demands with the aim to reproduce the flow supplied by the tanks and reservoirs and, on the other hand, it allows to determine the control rules required to replicate the behaviour of all regulating elements.

This detailed model built by query represents the network behaviour for a certain time period (one or several days) and can be connected with the SCADA to make live simulations. Hence, it can be used as an assistant to make informed decisions in real time. However, sometimes, depending on the use, connecting with the SCADA a strategic model or macro model that contains only the main elements is more useful to have the big picture of the system behaviour. In that case the detailed model can be simplified automatically to obtain a strategic model, both of them connected.

Either with the strategic or the detailed model, the network behaviour can be simulated any time: in real time, at any past time or near future (next 24 hours) considering the real network conditions.

It is important to highlight that this is not an easy task, and several challenges were overcome:

1) It was necessary to develop algorithms to fit GIS information to the model requirements:

- The GIS and model topology are different, for example, valves are usually represented as a point in the GIS while in the model are lines.
- The GIS information is sometimes wrong, for example, it may have inconsistent properties, failures on the network topology, etc. Hence, it was necessary to develop automatic error detection and correction tools to feed the model.
• Finally, the GIS may not contain all the necessary information, e.g. pump curves, pipe roughness, etc., so a system to store all the necessary extra information was designed.

2) Algorithms to calculate users’ demands were developed. This is necessary because the consumption is registered with different types of meters, and therefore in different time frequency. For example, some of them are smart meters that register the consumption every hour, others every day or even every week, some others are not smart meters and the consumption is registered every certain months.

3) On the other hand, developing algorithms to filter and correct the raw information provided by the SCADA was necessary, because faulty data leads to faulty simulations.

4) And finally, the elements considered by Epanet are limited and sometimes it is necessary to use a combination of several elements to model the behavior of the real component (for example a valve that feeds a tank from the top …). In the future new elements must be added to the existing simulation software.

Results

According to the previous requirements, Global Omnium, GoAigua and the Universitat Politècnica de València (UPV) defined a strategy to develop and maintain a DT of the water distribution network of Valencia and its Metropolitan Area, which approximately supplies water to 1.6 million inhabitants (Conejos et al. 2019).

In a first phase, the application GO2HydNet was developed, which permits to obtain from the GIS system a detailed model of the selected area, which can include all the pipes, operating elements and auxiliary elements that affect the water flows. The
graph exploration algorithms were full developed from scratch in order to achieve the highest possible computational efficiency. Thereby, the process of building the detailed model of Valencia, which contains 118,000 house connections with their corresponding consumption pattern each one, and generates a total of 325,000 nodes, takes about 1 minute of computation time. The process included data retrieving, the spatial processing and the model description file writing (the Epanet input file). The time required for data pre-processing was not included in the analysis, neither that needed by Epanet to open the input file.

The model can simulate either one or several days, and it considers all the maintenance operations that affect the water flows occurred during the simulation period. Finally, the field data are imposed (up to 20,000 daily records) in order to reproduce the real behaviour of the network up to the minimum detail allowed by the data sources, using the Epanet engine. Figure 3 shows, on the left, the extension of the full hydraulic model, build by request in a couple of minutes, and a detailed view of the house connections on the city centre.

1 The test was carried out on a personal computer with an i7-3.2 GHz processor and 32 Mb RAM, employing one single processor.
Figure 3. Building the Valencia’s detailed model of 325,000 nodes, including house connections, by query. Red dots stand for demands in house connections.

However, this model is too big to be used in real time, so the next step is to automatically simplify that model in order to reduce it to a more operational strategic model of 10,000 pipes. Nowadays, the DT works with a base strategic model which is coupled with the field data registered by the SCADA system in order to adjust the demands, pressures, levels and status of the regulation elements at every time instant, in order to get an operating twin that is updated every 1 minute (Figure 3, right). With the adjusted strategic model, the operators can simulate the real-time decisions, analyse the system behaviour at any past time, or program the actions to be carried out in the near future, usually in the next 24 hours.

The strategic model is updated periodically by using the GIS information, the information provided by the smart meters and, if necessary, the CMMS operations (long-
term operations). With the proposed strategy, the simplified model keeps its connection with the detailed model, so every operation can be performed with either one or another model, in such a way that both together constitute the DT currently managed by Global Omnium (Figure 4).

![Figure 4. The DT of Valencia WDN connected in real time with field data. By means of double-text boxes real-time data are compared with simulated ones. From the synoptics of control stations, the proposed actions can be simulated before carrying them out.](image)

Regarding the visualization capabilities of the Valencia DT, various environments were developed. On the one hand, general synoptics were designed with the aim to provide a DT interface oriented to the network operators. The synoptics allow simultaneous observation and comparison of the raw measured variables, filtered variables and simulated ones. Moreover, the information is organized by different colour codes. Each remote station is represented by its own synoptic, where the user can perform any simulation, in a similar way than a real SCADA, and analyse its impact over the network. The synoptics also contain additional information, such as water balances, evolution charts or piezometric profiles.

Likewise, the simulated values can be displayed over a geographic background. That kind of visualization was thought to provide a DT interface for planification tasks.
It allows the visualization in real time of the hydraulic variables of the whole network. The main advantage of that kind of visualization is the possibility to approximate the details of the hydraulic state of the different DMAs and supplied municipalities, without the need to install multiple sensors along the network.

The DT of Valencia WDN supports the daily operation of the system and serves multiple uses. The main uses are detailed as follows:

For planning, the DT is used in the development of master plans, where it allows to simulate the network behaviour under long-term demand forecast, to design the replacement plan of the existing insufficient infrastructures, the future extension of the network and the future operating guidelines and operation rules. Also, the DT is exploited to simulate the network emergency response, which allows to determine the actions that must be taken in an emergency and the infrastructures needed for minimizing the risks and their consequences.

Figure 5. Example of the exploitation of the Global Omnium’s DT as a decision support system for operation and design. A) Simulation results of the normal conditions. B) Simulation of a failure scenario (pipe burst). C) Results of the best operation strategy to mitigate the effects of the burst. D) Simulation results with the new infrastructure needed to amend the effects of the burst.
For operation and maintenance, the DT permits to achieve an extended and deep knowledge of the network behaviour, by simulating in real time what is happening every moment at any point; predicting the system behaviour in the near future, which commonly considers 24 hours (Figure 6); and analysing what happened at any past time.

Figure 6. Forecasted tank level evolution for a near future by means of the Global Omnium’s DT.

The DT is also used in operation and maintenance for operators training. Moreover, the DT serves as a decision support system under emergency conditions, as it helps with the development and validation of emergency protocols and allows testing or simulating different actions before making decisions under unforeseen emergencies. For works planning in the network, with the DT, the best day and schedule can be chosen, and the necessary operation protocols defined. In addition, the DT allows to optimize the network operation by minimizing the energy consumption, preserving quality and hydraulic service levels, minimizing leaks, etc. and finally, it is very useful for anomalies
detection, by comparing the simulated with the observed values, and by identifying outliers.

**Conclusions**

Digital Twins have reached the WDS area to stay. The development and live maintenance of a DT of a water distribution network is the further objective pursuit by many water utilities, but it is not an easy task. Nevertheless, the data and tools available nowadays make it possible. This paper presents the efforts made by Global Omnium, GoAigua and the UPV to achieve this goal, becoming one of the first utilities to have a DT connected to the main corporative data sources. The great benefits ensure that it will begin to be habitual in the most advanced smart cities.

The application GO2HydNet is capable of building a detailed hydraulic model from the information stored at the corporative big data platform, which gathers all the information provided by GIS, AMR, CMMS and field sensors. The algorithms were developed with the aim to achieve the highest possible efficiency. Thanks to that, GO2HydNet is able to build a detailed model with 325,000 nodes in barely a minute.

The processes to load the hydraulic model from field data, as well as their preprocessing and filtering, were successfully designed and implemented. It allows the DT to reproduce any past scenario, as well as simulate in real time conditions and obtain the forecasted evolution of the near future.

However, there are still remaining tasks to develop. So far, the detailed model and the strategic simplified model are not yet connected. Further efforts towards an automatic process to connect both models are needed. Moreover, the water quality model has not been developed yet. Due to the growing concern of WUs about the control and improvement of water quality issues, it is essential to take advantage of the potential usefulness of the DT quality model.
Declaration of interest

The authors declare non conflicts of interest.

References


