





# A case study on the effect of smart meter sampling intervals and gap-filling approaches on water distribution network simulations

Jonas Kjeld Kirstein , Klavs Høgh , Martin Rygaard   
and Morten Borup 

## ABSTRACT

Water usage data collected from smart meters at the end user can improve the accuracy and applicability of water distribution network models. Collecting and storing large amounts of data across hundreds or more smart meters is costly, which makes it important to consider what constitutes a sufficient sampling interval. This paper explores the effect of varying sampling intervals in smart meter data on model performance in regard to flow, pressure and water age simulations. Furthermore, the effect of using linear interpolation, a demand pattern or a network-inflow-weighted approach to fill gaps when data are sampled coarsely, is investigated. The study was based on real data from 525 smart meters in a district metered area in Denmark. The results show that smart meter data can improve modelling results, and if the sampling intervals are coarser than 2 h, then a weighted gap-filling approach markedly outperforms linear interpolation and models with coarse bi-annual demand data.

**Key words** | gap filling, hydraulic model, sampling interval, smart meter, water distribution network

Jonas Kjeld Kirstein  (corresponding author)  
Martin Rygaard   
Morten Borup   
DTU Environment,  
Technical University of Denmark,  
Kgs. Lyngby,  
Denmark  
E-mail: [jkir@env.dtu.dk](mailto:jkir@env.dtu.dk); [jkj@niras.dk](mailto:jkj@niras.dk)

Jonas Kjeld Kirstein  
Klavs Høgh   
NIRAS A/S,  
Allerød,  
Denmark

## HIGHLIGHTS

- Study on the impact of varying smart meter data sampling intervals used as input to a demand-driven EPANET model.
- Finer resolution data improved flow, pressure and water age simulations.
- Smart meter data require gap-filling to estimate the demand at uniform intervals.
- Gap-filling based on demand multiplier patterns outperformed linear interpolation.
- Coarse smart meter readings may be sufficient for water loss assessment.

## INTRODUCTION

In recent years, utilities around the world have installed smart water meters at an increasing rate. This implementation entails a wide range of benefits for utilities and customers, including reduced operational costs, enhanced demand management, efficient pipe network infrastructure planning, accurate billing, improved leakage detection and higher customer satisfaction (Stewart *et al.* 2018; Monks *et al.* 2019). In model-based water distribution systems analysis, the quantity and quality of the available data are often

among the most restraining factors (Savic *et al.* 2009) and the expected widespread use of smart meter data has the potential to improve the applicability and accuracy of the models (Gurung *et al.* 2014, 2016).

The terms ‘sampling interval’, ‘frequency’ and ‘resolution’ describe the temporal distance between a device’s water meter readings (WMRs), which has a major influence on the usefulness of the data. Cominola *et al.* (2018) showed that fine sampling resolutions (ranging from minutes to

seconds) increased significantly the accuracy of end-use disaggregation. Furthermore, they showed that the magnitude and timing of the peak demand can vary by up to 62% and more than 15 h, respectively, when changing the sampling resolution from 10 s to 1 day. Also, Gurung *et al.* (2014) showed how enhanced demand patterns can be generated based on very fine sampling intervals (5 s) in smart meter data sets, thus reducing measured peak demand levels and updating the time of peak occurrence. This is crucial information, as demand patterns and associated peaks are used to design and optimise water distribution systems (Gurung *et al.* 2014; Cominola *et al.* 2018). In terms of water quality modelling, Blokker *et al.* (2008) showed that the spatial aggregation and hydraulic time step play a significant role. Time steps of 1 h can be sufficient for larger transportation networks and water quality models, including advection reactions, but at finer spatial aggregations, water quality models should also include dispersion reactions with time steps below 5 min. This should be done, for example, when simulating water age in parts of the networks where laminar flow occurs (Blokker *et al.* 2010). However, in Blokker *et al.* (2011), a hydraulic modelling time step of 15 min was determined to be sufficient for accurate residence time computations using a bottom-up demand allocation approach (i.e. unique demand multiplier patterns for each household). Creaco *et al.* (2017) showed that when using the conventional top-down approach (i.e. allocating strongly correlated demand multiplier patterns to nodes), pressure head simulations require larger ( $\geq 1$  h) time steps before being deemed reliable. Also, Creaco *et al.* (2017) showed that the bottom-up approach combined with extended period simulations is capable of generating accurate pressure and flow simulations at time steps larger than 2 min. However, at shorter time steps, unsteady flow modelling is required to reflect the behaviour of the water distribution network (WDN), but the computational overhead of such methods limits usability in real-time applications (Creaco *et al.* 2017).

Even though case studies have shown the benefits of finer sampling intervals, the increased data volumes, originating from the implementation of advanced metering infrastructure, also pose challenges for utilities. One example includes data management, as utilities may struggle with identifying the best type and frequency of data

needed for tasks related to the planning and operation of water distribution systems (Boyle *et al.* 2013). With a finer sampling interval, the required data storage per installed meter grows, and there is an important trade-off between sampling interval and battery life of smart meters. The emergence of low-power wide-area networks (Stewart *et al.* 2018) reduces the required energy for data transmission, but at the cost of coverage/range, payload length (bytes per message) and maximum number of messages sent per day per device (Mekki *et al.* 2019). Common to most large-scale roll-outs of smart meters is the fact that the smart meters use their own transmission channels with more focus on energy efficiency than transmission reliability, as opposed to traditional system sensors connected to the SCADA system. This makes smart meter data conceptually different from other types of data from water distribution systems. Where data from traditional sensors are typically transmitted and stored at regular configured intervals, smart meter readings can arrive at irregular intervals depending on the system setup, and data from even the same meter are likely to arrive at different intervals. This makes gap filling and resampling of the data an important task before the data can be used for driving a model – regardless of the simulation's time step.

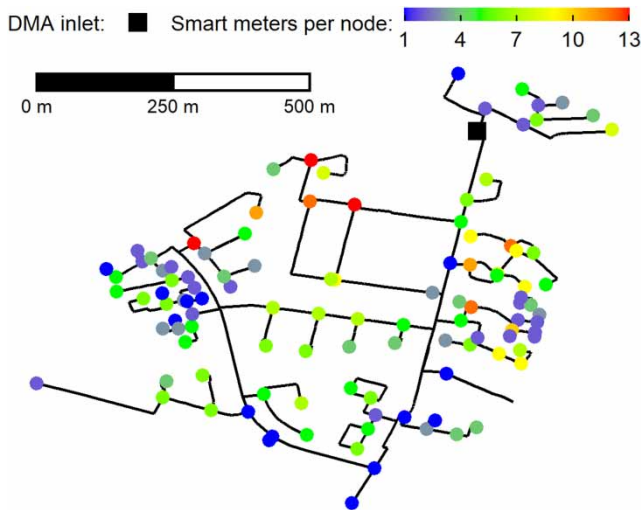
Our study aims to support utilities in selecting the sampling interval best suited to their needs, based on a district metered area (DMA) with full smart meter rollout. We analysed data intervals and resampled the collected fine-resolution data to coarser sampling intervals and investigated its impact on hydraulic modelling results. Compared with other stochastic bottom-up demand allocation approaches (e.g. Blokker *et al.* 2011; Creaco *et al.* 2017), this study used measured consumption data from household smart meters as input to a hydraulic model. We investigated how different approaches to fill data gaps in data sets with random or uniformly sampled intervals affect hydraulic simulations. This is conducted by comparing the filling of gaps between adjacent WMRs by linear interpolation with interpolation based on a demand multiplier pattern and a pattern resembling the DMA inflow. We compare simulation results of water consumption, pressure head and water age based on data-set scenarios with varying sampling intervals and gap-filling methods.

## CASE STUDY

The methodology was applied to a DMA in Brønderslev, Denmark, where all 525 consumers have smart meters installed. In the hydraulic model of the DMA, 111 of 128 nodes have smart meters assigned, averaging 4 smart meters per network node (Figure 1). According to the utility, leakages are insignificant in this DMA and can thus be ignored in the case study area.

Flow and pressure data at the DMA inlet were available at a uniform 5-min sampling interval. Furthermore, 2 times 2 weeks of smart meter data were available for analysis (Table 1). The raw smart meter data used in this study consisted of timestamped cumulative volume readings.

The smart meters are Diehl's water meter HYDRUS, based on ultrasonic technology (Diehl Stiftung & Co. KG 2019). The meters transmit the WMR in two signals, one short-range signal approximately every 20 s (the main purpose being to facilitate drive-by collection) and a long-range signal approximately every 5 min, which is collected



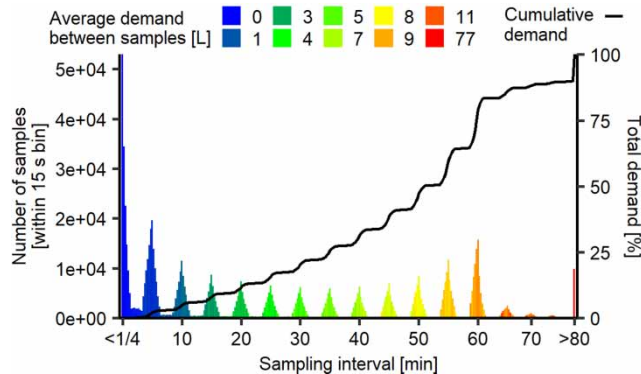
**Figure 1** | WDN model of the case study area and the number of smart meters assigned to the model nodes. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2020.083>.

**Table 1** | Information about the applied smart meter data in Brønderslev, Denmark

| Data set | Period             | Smart meters        | Total number of WMRs           | Average sampling interval (min) |
|----------|--------------------|---------------------|--------------------------------|---------------------------------|
| Training | 01–15 August 2018  | 525 (100% coverage) | 367,917 (1–2,403) <sup>a</sup> | 29 (8–20,160) <sup>a</sup>      |
| Testing  | 01–15 January 2019 |                     | 410,395 (1–2,379) <sup>a</sup> | 26 (8–20,160) <sup>a</sup>      |

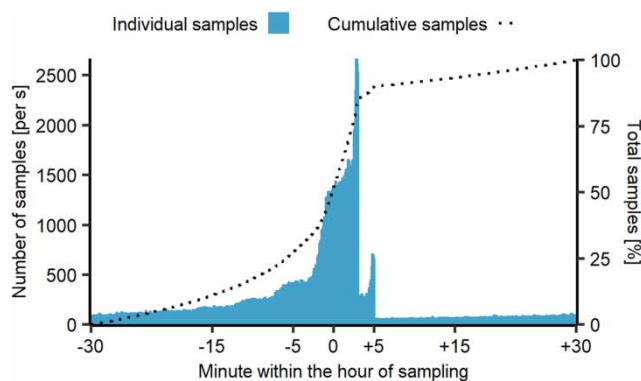
<sup>a</sup>(minimum–maximum per smart meter).

by data concentrators on antenna towers. The exact transmission interval varies between the different smart meters to avoid simultaneous interfering signals which increases the reliability of the data collection. In general, time stamps of WMRs may be inaccurate and lead to wrong modelling results, depending on the selected technology and system setup. For example, Kirstein *et al.* (2019) observed drifts and off-sets in the internal clock of flow meters installed in WDNs. To avoid such malfunctions, the smart meter technology provider in this study timestamps the WMRs centrally when the data reach the concentrators in the antenna towers (or the occasional drive-by vehicle), which ensures accurately timestamped WMRs. Multiple signals from the same smart meter may be received by the same antenna over the course of an hour, but in the current setup, only the reading closest to the nearest full hour is transferred to the utility's database while the others are discarded. However, more readings from each smart meter than one per hour may be present in the database if the smart meter data are collected by multiple concentrators, e.g. by two antenna towers or by both an antenna tower and drive-by collection. This leads to a smaller sampling interval than 1 h. Due to poor transmission signals, there may also be sampling intervals larger than 1 h. Thus, a mixture of signals will be collected, which implies that the overall sampling interval of smart meter data stored in the database can seem rather irregular or random. Figure 2 shows the frequency of the temporal distance, i.e. actual sampling interval, between adjacent WMRs and the share of total DMA demand that has been captured within the bins of varying temporal distances. The sampling interval between adjacent WMRs in the analysed periods ranges from 1 s to more than 80 h (maximum 125 h). Figure 2 shows, however, that close to 75% of the demand was sampled below a temporal distance of 60 min. A great proportion of adjacent samples have a temporal distance below 1 min, but less than 1% of the demand was sampled within this interval.



**Figure 2** | Distribution of temporal distances between adjacent WMRs (i.e. sampling interval) covering the training and testing period (Table 1). Smart meters with fewer than two samples within each period were omitted. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2020.083>.

The histogram in Figure 3 shows at which second of the hour samples were timestamped. The cumulative line represents the total share of samples received within this timeframe. The figure shows that most samples in the database are sampled around the full hour and that the remaining samples are timestamped irregularly. Samples close to the full hour may indicate a good connectivity of a smart meter to an antenna, whereas samples farther from the full hour may indicate smart meters with a reduced connectivity. For example, the 5-min transmission intervals visible in the form of recurring peaks in Figure 2 may indicate antennas with a reduced connectivity to smart meters. Another measure of the case study's data collection reliability can be found by sectioning the training and testing



**Figure 3** | Distribution of WMR timestamps according to their occurrence within the hour for the data from both the training and testing period (Table 1).

periods' data (Table 1) into uniform time intervals of 1 h, starting at 00:00. By doing so, only 294 out of the 525 smart meters have at least one WMR available for 95% of the time. This number, however, increases to 494 meters when the threshold is decreased to 75% of the time.

## METHODOLOGY

This paper assesses the effect of three different gap-filling methods and varying sampling intervals on WDN simulations. Thus, this methodology section includes information on:

- **Gap-filling methods:** Describes the investigated gap-filling methods which regularly align the data and fill the gaps between WMRs.
- **Sampling intervals:** Describes how different average sampling intervals are synthetically obtained.
- **Effect on WDN modelling:** Describes the assessment of how different gap-filling approaches and sampling intervals affect the modelling results.

### Gap-filling approaches

Three gap-filling approaches are tested in this article:

- (1) Linear interpolation, which to the best of our knowledge is currently used by smart meter technology providers, e.g. when providing data for hourly water loss assessments.
- (2) Application of a representative demand multiplier pattern of the consumers in the case study area, which weights the consumption depending on the hour of the day.
- (3) Another weighted approach where the demand multiplier pattern is calculated based on the actual DMA inflow to the case study DMA. In other words, the latter approach uses the actual variation in the inflow to the DMA as weights to fill gaps between adjacent WMRs.

For gap-filling method 2, the demand multiplier patterns for work and weekend days were generated based on all WMRs during the training period (Table 1). First, the data were resampled to the nearest uniform timestamp with an interval of 5 min, using the measured DMA inflow of the training period as a gap-filling approach. Next, the mean consumption of all consumers  $\bar{x}_t$  was calculated for each

5-min step  $t$  over 24 h using either all weekdays or all weekend days in the training period data set, resulting in two demand multiplier patterns. Each pattern was found by:

$$\text{Demand multiplier}(t) = \frac{\bar{x}_t \cdot n(\bar{x}_t)}{\sum_{t=1}^{\text{timestamps}} \bar{x}_t} \quad (1)$$

where  $n(\bar{x}_t)$  is the total number of timestamps per 24 h, equalling 288 time steps of 5 min each.

Figure 4 illustrates the principles behind the applied gap-filling approaches. In this theoretic example, four cumulative volume readings (black dots) of one smart meter with varying temporal distance (Figure 4(a)) were captured over the course of 1 day. The measured water demand is the difference between adjacent cumulative volume readings (y-axis). For example, the two first WMRs show that  $3 \text{ m}^3$  have been consumed over 7 h (Figure 4(a)), corresponding to a constant flow of  $0.43 \text{ m}^3/\text{h}$  (yellow line, Figure 4(c)). Owing to this coarse sampling interval, however, the actual time of water consumption is unknown, as water consumption in households occurs instantaneously and may only last from seconds to minutes. In Figure 4, two approaches are applied to estimate the time of consumption in between WMRs and to align the irregular data to regular timestamps: one where the data are aligned by linear interpolation (yellow line, Figure 4(a)), and another where gaps are filled with a demand multiplier pattern (blue line, Figure 4(a)) representing gap-filling approach 2 and 3. In this example, the demand multiplier pattern with hourly multipliers shown in Figure 4(b) was applied. Such patterns

are often available for utilities, as they are an integral part of hydraulic models for various types of consumers. Finally, when the volume at regular time steps has been estimated (here hourly), it is possible to compute the flow over the period of analysis. Depending on the selection of the gap-filling method, the estimated flow may vary considerably (Figure 4(c)). These differences increase with coarser sampling intervals and a growing number of smart meters bundled in the same nodes of the hydraulic model.

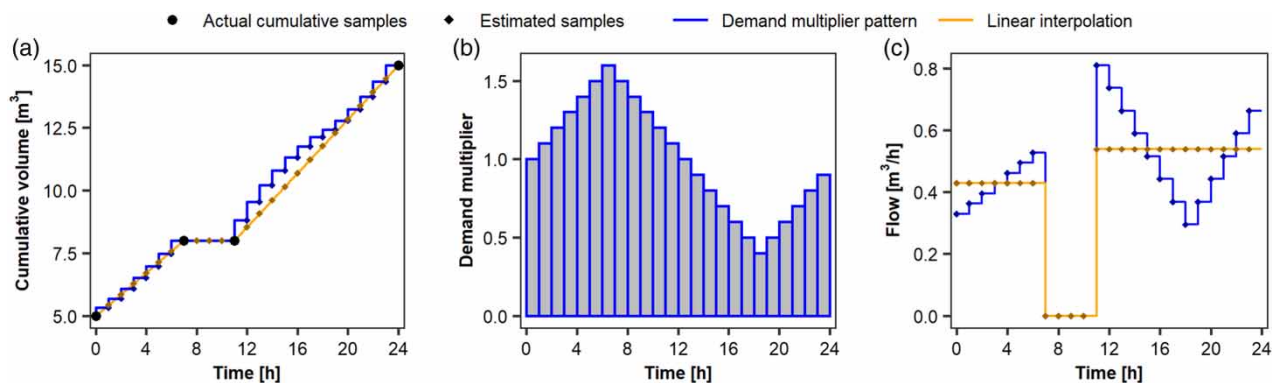
### Sampling intervals

The sampling interval describes the average temporal distance between WMRs. The sampling interval between a device's individual WMRs can be uniform (e.g. every full hour) or, as is the case with the presented real-world data, pseudorandom. Pseudorandom data may occur when data transfer depends on water consumption triggering the transmission of messages, when data connections are unstable, or due to specific system setups (as described earlier).

The sampling interval can be computed for the entire data set or each individual smart meter. The average sampling interval for the entire data set of  $N_{\text{devices}}$  smart meters is calculated as follows:

$$\text{average sampling interval} = \frac{T \cdot N_{\text{devices}}}{\sum_{i=1}^{N_{\text{devices}}} [n(i)]} \quad (2)$$

where  $T$  is the length of the analysed period (e.g. of the testing period in Table 1) and  $n(i)$  represents the number of



**Figure 4** | (a) Four cumulative volume readings from a single smart meter with irregular and unaligned timestamps. To estimate the water consumption at regular timestamps, gaps between adjacent samples have been filled using either linear interpolation or a demand multiplier pattern. (b) Demand multiplier pattern used for gap filling in (a). (c) Flow rates computed based on the estimated cumulative values. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2020.083>.

WMRs within  $T$  for each device  $i$ . A smart meter's individual average sampling interval can be computed for  $N_{\text{devices}}$  equal to 1. For example, the average sampling interval of the data shown in Figure 4(a) is 6 h. For the case study, the finest average sampling interval of a smart meter is around 8 min (Table 1), which is of much finer resolution than is usually the case.

The small sampling interval allows us to resample the data to represent smart meter data with coarser resolution to assess the effect of the sampling interval on the different gap-filling methods. Two types of sampling intervals are created: one where the sampling intervals are uniformly distributed, and one where the sampling intervals are pseudorandom but overall provides the required average sampling interval:

- To create uniform sampling intervals, the gaps between the WMRs available in the testing period (Table 1) for each smart meter were first filled by using the DMA inflow pattern as the demand multiplier pattern. Hereafter, coarser time series were created by resampling the gap-filled time series at uniform intervals.
- Scenarios with pseudorandom sampling intervals were created by deleting WMRs at random from the testing period's data set until the required sampling interval according to Equation (2) was achieved. The random deletion was performed by deleting one data point from a data set at a time, where all data points had the same probability of being deleted, irrespective of the temporal distance to other data points.

### Effect of gap-filling approaches and sampling intervals on WDN modelling

The impact of gap-filling methods, sampling methods and sampling intervals on total DMA consumption, pressure

**Table 2** | Various data set scenarios applied to WDN simulation

|                               |   |
|-------------------------------|---|
| Average sampling interval (h) | 0.5; 1; 2; 3; 4; 6; 8; 12; 24                               |
| Sampling methods              | Uniform; Random   |
| Gap-filling methods           | Linear interpolation; Demand multiplier pattern; DMA inflow |

head and water age simulations were assessed for a number of data scenarios (Table 2). The water age is an important parameter since this is one of the main drivers for all water quality simulations. The DMA consumption represents the summed water usage of all users in the DMA. All hydraulic simulations were run in EPANET (Rossman 2000). Measured pressure with a sampling interval of 5 min at the single inlet to the DMA was used as the boundary condition, and the model was accordingly run in 5-min time steps.

We ran the simulations at time steps of 5 min in order to (1) align the model results with the DMA inflow observations (having the same sampling interval) and (2) to reduce the loss of information in periods with small temporal distances between WMRs, as indicated in Figures 2 and 3. Therefore, the WMR data needed to be resampled at a 5-min interval, which implies estimating values in between the sometimes much coarser WMRs. All model simulations were based on data covering the testing period (Table 1).

Our analysis included two benchmark scenarios:

- (1) *Benchmark model*. The raw data set, containing the finest average sampling interval available (i.e. 26 min; Table 1), was aligned by using the DMA inflow during the same period as weight for filling gaps in between adjacent WMRs. This data set and the corresponding hydraulic model will represent the *benchmark*. It should be noted that this is only possible for the case study DMA because the majority of consumers are identical, i.e. single-family houses. This benchmark model has an average water demand of 5 m<sup>3</sup>/h and an average water age of 17.8 h at nodes with consumers assigned (Figure 1). The average pressure head lost from inlet to each node was 0.2 m.
- (2) *Bi-annual models*. These models represent today's best practice. First, the average demand of each consumer was found by taking the difference between the newest WMR from the training period and the oldest WMR from the testing period for each consumer, approximating a bi-annual audit data set. The average demand was then assigned to each node in the hydraulic model. Based on this, two different (top-down) models were constructed.

- (a) Smart meter-based demand multiplier pattern scaled by bi-annual consumption.

(b) DMA inflow pattern scaled by bi-annual consumption.

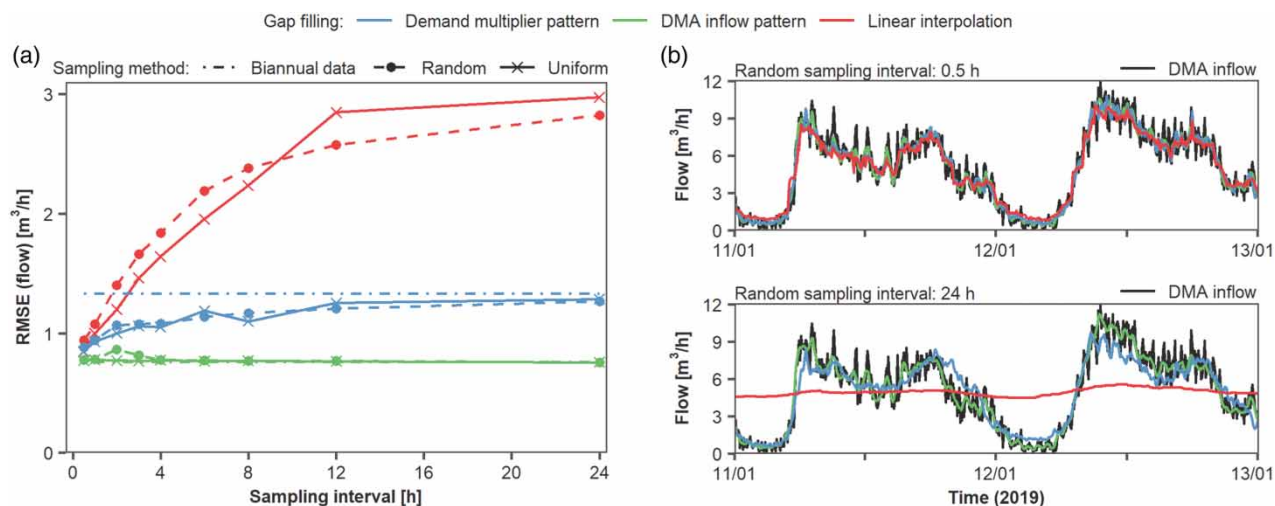
The root-mean-square error (RMSE) was used to assess the effect of the different scenarios on the WDN modelling results. The RMSE between the modelled and measured inflow to the DMA was computed for the data set scenarios (Table 2). Further, each scenario's simulated pressure and water age results were compared with the benchmark model's results. This was done by calculating the RMSE for all individual nodes having consumers assigned (Figure 1) and subsequently taking the mean of these nodal RMSEs for each data set scenario. The first 84 h were not included to avoid any impact from the initial water age conditions on the model results. All RMSE computations were based on a temporal resolution of 5 min.

## RESULTS

### Measured versus simulated consumption

Figure 5(b) shows the variation in the simulated consumption over 2 days for two selected sampling intervals and a good resemblance between the inflow and demand; thus, this confirms that leakage in the DMA is insignificant. The figure shows that even 24-h sampling intervals can provide

decent modelling results when a demand multiplier pattern or DMA multiplier pattern is used for gap filling. This will, of course, depend on the quality of the demand multiplier pattern. Figure 5(a) shows the RMSE based on the measured DMA inflow and the simulated consumption in the DMA for the testing period. DMA inflow data for gap filling gave the best performing data because inflow data in that case was partly compared with itself. The DMA inflow did thus not act as an independent reference in this specific case. The choice of the gap-filling method is more important than the sampling method for flow simulations, and finer sampling intervals of the WMR data gave better simulation results (Figure 5(a)). A demand multiplier pattern bi-annual model (dashed blue line) outperformed linear interpolation at sampling intervals coarser than 2 h (Figure 5(a)). However, at finer sampling intervals, there is only a slight difference between linear interpolation and the demand multiplier pattern-based approach. At sampling intervals equal to or greater than 3 h, a demand multiplier pattern approach should be favoured over linear interpolation. Furthermore, there is no clear trend in the difference between random and uniform sampling. It is noteworthy that Figure 5(a) does not indicate a lower limit under which a finer sampling interval will no longer improve the modelling results. If such a limit exists, it must be finer than 30 min and is not possible to explore with the current data.



**Figure 5** | (a) RMSE between measured inflow to the DMA and accumulated demand in WDN simulations at various sampling intervals and gap-filling methods during the testing period (Table 1). For the bi-annual data, the same demand multiplier pattern or DMA pattern was assigned to each node. (b) Simulated and measured DMA inflow. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2020.083>.

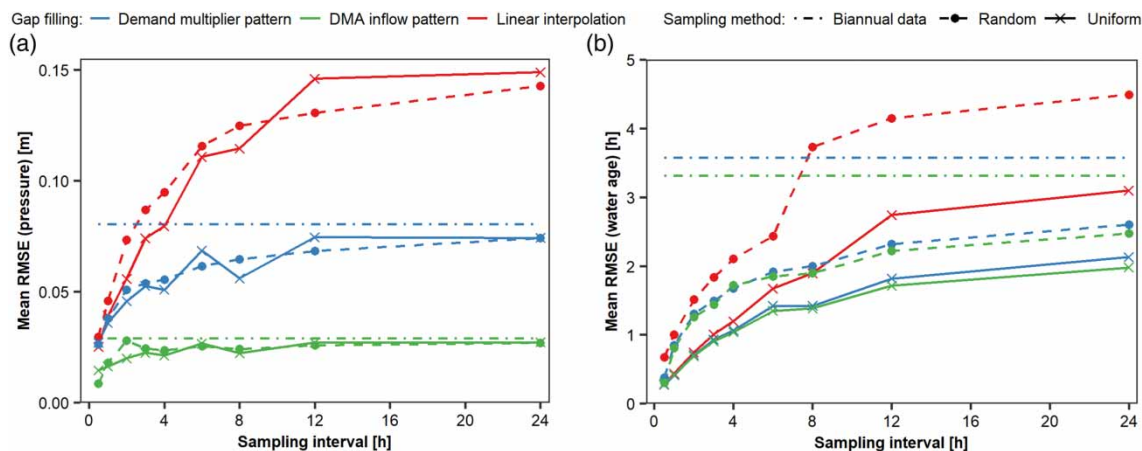
Figure 5 shows that the value of frequent smart meter readings in a DMA may be less than what can be gained from fewer demand readings combined with DMA inflow data or demand multiplier patterns. The added value from smart meters is less, where utilities have representative and up-to-date knowledge about consumer types and expected demands. Utilities may struggle to achieve such knowledge without the data available from smart meters. For example, water audits from utilities without automated readings available may be delayed and inconsistent and no longer represent an area's demand when available. Our case study area was made up of a single consumer type, making it possible to assign the same demand multiplier pattern to the consumers and to assume that the DMA inflow represents the consumers' variation in demand. In areas with varying consumer types, the overall DMA inflow may not be able to represent a consumer type's variation in demand and representative demand patterns may not be available without smart meters. Future studies should thus include and assess the impact of, among other things, varying consumer types, different DMA sizes and leakages.

### Simulation of pressure head and water age

The low head loss between the inlet and nodes with consumers assigned in the benchmark model (average of 0.2 m) is due to over-dimensioned pipes owing to fire safety regulations or outdated pipe-dimensioning standards typical of

Danish WDNs. Consequently, the mean RMSE of simulated pressure even for coarser sampling intervals was also low (Figure 6(a)). A maximum mean RMSE value of approximately 0.15 m was found for a sampling interval of 24 h. This sounds an insignificant error, but relative to the generally low head losses in the system the error is actually rather significant. Figure 6(a) shows that DMA pattern and demand multiplier pattern-based approaches outperformed linear interpolation in terms of simulating pressure head.

Depending on the sampling interval and gap-filling method, the mean RMSE of water age simulations varied between 16 min and 4.5 h (Figure 6(b)). As opposed to pressure and flow results, the choice of the gap-filling method had little impact on water age, especially for sampling intervals of 4 h or less. This can be explained by the fact that the age of the water is a result of many hours of consumption, so that age is less dependent on short-term variations. However, at sampling intervals coarser than 4 h, the demand multiplier pattern and DMA-based gap-filling approach resulted in better overall results than linear interpolation. Moreover, the water age simulations performed better with uniform sampling. This can be explained by the fact that the random sampling method can have smart meters without any or only few WMRs, whereas the uniform sampling method still required data from all meters at a given sampling interval. Thus, if certain areas in the network had no WMRs for long periods in the random sampling method, the error in water quality simulation at these locations



**Figure 6** | Mean RMSE of (a) pressure head and (b) water age simulations with varying sampling methods and sampling interval based on all nodes with smart meters installed. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2020.083>.



increased. At sampling intervals finer than 8 h, all sampling methods reduced the mean RMSE notably compared with the bi-annual models. As the water age simulation results are theoretical, future studies should include tracer measurements in the network to validate whether the models with smart meter data improve water age simulations.

The 'Case Study' section showed that the antenna tower concentrators reduced the smart meter data set to WMRs around the full hour. In terms of water quality simulations, this procedure removes valuable WMRs containing information about the actual time of consumption. In the future, more advanced reduction processes should be considered that keep samples with information about the actual time of consumption from being removed, ultimately reducing the need for water demand models to fill gaps between readings and improving the validity of water quality simulations.

Finally, it should be noted that the results presented in this case study are based on a single DMA. DMAs vary greatly in the number and types of customers, pipe age, size, level of leakage, etc. Thus, not all approaches presented in this study can be transferred directly to any other network and require further attention in future studies.

## CONCLUSIONS

By using data from 525 household smart meters to specify consumer demand in a water distribution network model, and by varying the sampling intervals of the smart meter data, we conclude the following:

- If there is a limit below which a finer sampling interval of smart meter data will not improve modelling results any further, then it is below 30 min.
- It is much better to use representative demand multiplier patterns than linear interpolation to fill gaps between observations. A sampling interval of 3 h with linear gap filling led to comparable or worse flow simulations than 24-h data with demand multiplier pattern-based gap filling.
- Water age simulation error between a benchmark model and models with coarser smart meter sampling intervals increased notably almost regardless of the applied gap-

filling method. Simulations of flow, on the other hand, were affected more by the choice of the gap-filling method.

Smart meter data can greatly improve modelling results, and if the average sampling interval is coarser than 2 h, then a weighted gap-filling approach should be used.

## ACKNOWLEDGEMENTS

We thank Brønderslev Forsyning A/S, in particular Henrik Horsholt Christensen, Per Grønvald and Thorkil Bartholdy Neergaard, for providing data and answering questions related to their water supply. This project was partly funded by the Danish Eco-Innovation Program (MST-141-01277/NST-404-00378).

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Blokker, E. J. M., Vreeburg, J. H. G., Buchberger, S. G. & van Dijk, J. C. 2008 [Importance of demand modelling in network water quality models: a review](#). *Drinking Water Eng. Sci.* **1**, 27–38. doi:10.5194/dwes-1-27-2008.
- Blokker, E. J. M., Vreeburg, J. H. G., Beverloo, H., Klein Arfman, M. & van Dijk, J. C. 2010 [A bottom-up approach of stochastic demand allocation in water quality modelling](#). *Drinking Water Eng. Sci.* **3**, 43–51. doi:10.5194/dwes-3-43-2010.
- Blokker, E. J. M., Beverloo, H., Vogelaar, A. J., Vreeburg, J. H. G. & van Dijk, J. C. 2011 [A bottom-up approach of stochastic demand allocation in a hydraulic network model: a sensitivity study of model parameters](#). *J. Hydroinf.* **13**, 714–728. doi:10.2166/hydro.2011.067.
- Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S. & Stewart, R. 2013 [Intelligent metering for urban water: a review](#). *Water* **5**, 1052–1081. doi:10.3390/w5031052.
- Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D. E. & Abdallah, A. M. 2018 [Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management](#). *Environ. Modell. Software* **102**, 199–212. doi:10.1016/j.envsoft.2017.11.022.

- Creaco, E., Pezzinga, G. & Savic, D. 2017 [On the choice of the demand and hydraulic modeling approach to WDN real-time simulation](#). *Water Resour. Res.* **53**, 6159–6177. doi:10.1002/2016WR020104.
- Diehl Stiftung & Co. KG 2019 *Hydrus*. Available from: <https://www.diehl.com/metering/en/portfolio/water-metering/water-metering-product/hydrus/63697/> (accessed 4 July 2020).
- Gurung, T. R., Stewart, R. A., Sharma, A. K. & Beal, C. D. 2014 [Smart meters for enhanced water supply network modelling and infrastructure planning](#). *Resour. Conserv. Recycl.* **90**, 34–50. doi:10.1016/j.resconrec.2014.06.005.
- Gurung, T. R., Stewart, R. A., Beal, C. D. & Sharma, A. K. 2016 [Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning](#). *J. Cleaner Prod.* **135**, 1023–1033. doi:10.1016/j.jclepro.2016.07.017.
- Kirstein, J. K., Høgh, K., Rygaard, M. & Borup, M. 2019 [A semi-automated approach to validation and error diagnostics of water network data](#). *Urban Water J.* **16** (1), 1–10. doi:10.1080/1573062X.2019.1611884.
- Mekki, K., Bajic, E., Chaxel, F. & Meyer, F. 2019 [A comparative study of LPWAN technologies for large-scale IoT deployment](#). *ICT Express* **5**, 1–7. doi:10.1016/j.ict.2017.12.005.
- Monks, I., Stewart, R. A., Sahin, O. & Keller, R. 2019 [Revealing unreported benefits of digital water metering: literature review and expert opinions](#). *Water* **11**, 838. doi:10.3390/w11040838.
- Rossman, L. A. 2000 *EPANET 2 Users Manual*. Cincinnati, OH.
- Savic, D. A., Kapelan, Z. S. & Jonkergouw, P. M. R. 2009 [Quo vadis water distribution model calibration?](#) *Urban Water J.* **6**, 3–22. doi:10.1080/15730620802613380.
- Stewart, R. A., Nguyen, K., Beal, C., Zhang, H., Sahin, O., Bertone, E., Vieira, A. S., Castelletti, A., Cominola, A., Giuliani, M., Giurco, D., Blumenstein, M., Turner, A., Liu, A., Kenway, S., Savić, D. A., Makropoulos, C. & Kossieris, P. 2018 [Integrated intelligent water-energy metering systems and informatics: visioning a digital multi-utility service provider](#). *Environ. Modell. Software* **105**, 94–117. doi:10.1016/j.envsoft.2018.03.006.

First received 8 May 2020; accepted in revised form 24 September 2020. Available online 20 October 2020