TEMPORAL SCALE EFFECT ANALYSIS FOR WATER SUPPLY SYSTEMS MONITORING BASED ON A MICROCOMPONENT STOCHASTIC DEMAND MODEL

Sarai Díaz[1] and Javier González[2]

[1]Corresponding author: Dr. Eng, Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain). E-mail: Sarai.Diaz@uclm.es

[2]Dr. Eng, Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain). // HIDRALAB INGENIERÍA Y DESARROLLOS, S.L., Spin-Off UCLM, Hydraulics Laboratory Univ. of Castilla-La Mancha, Av. Pedriza, Camino Moledores s/n, 13071 Ciudad Real (Spain). E-mail: Javier.Gonzalez@uclm.es

The final published version is available at
https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0001352

ABSTRACT

Water demands are the main random factor that conditions flow variability within drinking water supply systems. The importance of using high-resolution demands in distribution mains is already well-known, but there is little knowledge of how the temporal scale (i.e. sampling frequency) affects the ability of a metering or monitoring system to explain network performance. The aim of this paper is to analyse the variability (i.e. information) that is lost because of not using a more frequent sampling rate to characterize water demands. For such purpose, a novel analytical approach based on a conceptualization of the microcomponent-based SIMDEUM model (SIMulation of water Demand, an End-Use Model) is presented. This methodology provides the statistical properties of water demands over different sampling frequencies. It is here applied to Benthuizen case study to

further explore the effect of temporal and spatial scaling laws under realistic conditions. Results are of major importance for monitoring design, as they highlight the need for properly combining measurements with different levels of resolution. Moreover, they enable to assess the impact of the sampling selection on the potential characterization level of monitored demands within urban water modelling applications.

INTRODUCTION

Drinking water supply systems have traditionally been modelled following a deterministic approach, based on assumed average values for input data, such as water demands or pipe roughness. Water demand has been identified as a major source of uncertainty among these, as its variability affects the reliability of the spatial and temporal distribution of the hydraulic variables resulting from the model (Magini et al. 2008). Conventional hydraulic models commonly average water demands spatially and temporally. Spatial averaging is usually undertaken by aggregating several water users into a single demand node, whereas time averaging consists on smoothing the instantaneous variations in demands (Buchberger and Wu 1995). From a temporal point of view, pseudo-steady models are typically assumed, i.e. demand multiplier patterns are assigned to the average demand of each node (Blokker et al. 2011a). Such spatial and temporal approximations may be sufficient for the arteries that transport water to District Metered Areas (DMAs), but the stochastic nature of demands becomes especially important when modelling distribution mains that deliver water to final users. In these last downstream pipelines, there is high spatial and temporal variability of demands, with low auto and cross correlation among individual homes (Filion et al. 2008). The pursuit of more realistic hydraulic models has motivated the development of stochastic demand models that enable to simulate the spatial and temporal complexity of water demands (Vertommen et al. 2012).

Buchberger and Wu (1995) presented the first stochastic model for residential water demands. This approach assumes that Poisson Rectangular Pulses (PRP) can be used to simulate the intensity, duration and frequency of water consumption at a household. The model conceives the household as a whole, so that PRP parameters and probability functions can be adjusted based on flow measurements at monitored homes (Buchberger and Wells 1996). This method established a basis for the analysis, over which several other pulse models have been presented (see Creaco et al. 2017 for a literature review). Years later, an alternative to available household-based methods came forward: the so-called SIMDEUM model (Blokker et al. 2009; Blokker et al. 2010; Blokker et al. 2011b). SIMDEUM is a microcomponent-based model that builds the overall water demand at a household by aggregating demand pulses for each inhabitant (i.e. end-user) at a fixture level (e.g. tap, shower, washing machine) (Creaco et al. 2017). Rather than relying on flow measurements like the first type of models, the end-use approach is fed with survey-based parameters. This implies dealing with a greater number of input parameters, which are easier to obtain (i.e. surveys instead of experimental campaigns). The original SIMDEUM model relies on Monte Carlo simulations. Each simulation provides one high-resolution water demand pattern. Spatial resolution can be adjusted by aggregating pulse demands as required, and a small time scale (1 second) is used (Blokker et al. 2010). PRP-like and SIMDEUM models have proven to give similar results for different spatial and temporal scales (Blokker et al. 2009; Creaco et al. 2017). However, as Monte Carlo simulations may lead to important computational times at large urban areas or biased results if the number of simulations is not sufficient (Blokker et al. 2011a), Díaz and González (2020) have recently presented an analytical approach to SIMDEUM model that provides statistical characterization (i.e. mean and variance) of instantaneous demands, avoiding the need for Monte Carlo simulations. Such a tool has proven to be useful in order to assess network spatial scale effects under heterogeneous uses conditions, but its potential to evaluate the temporal scale effect has not been explored yet.

As well as in other research fields (like hydrology, e.g. Rodriguez-Iturbe 1986), the relevance of temporal scale effects has already been discussed in water supply systems. Tessendorff (1972) suggests adopting different time intervals for peak flow estimation: 15 s for customer's installation lines, 2 min for service lines, 15 min for supply lines and 30 min for water mains. This is due to the fact that temporal resolution affects the variability of water demands: considering longer time intervals implies losing information about consumption signals, resulting in lower variance values (Buchberger and Nadimpalli 2004). For this reason, small time intervals (i.e. high temporal resolution) are required in the terminal branches of a water supply system in order to simulate their full variability, whereas longer time intervals can be considered as aggregating upstream because lower relative variability is expected. Scaling laws have been presented in the literature before as an analytical way of estimating realistic values for water consumption moments (mean, variance and covariance) under varying spatial and temporal resolutions. Magini et al. (2008) and Vertommen et al. (2012) presented simple scaling laws that provided demand moments according to the number of aggregated users (i.e. spatial scaling). Vertommen et al. (2015) explicitly incorporate the spatial and temporal correlation into the scaling laws when considering two groups with different characteristics. In what regards the temporal effect in the statistical distribution of water consumption, Kossieris and Makropoulos (2018) analysed the statistical characteristics of stochastic residential demands on a 15-60 minutes temporal scale (standard time resolutions in many urban water modelling applications) by systematically analysing demand records. Shortly afterwards, Kossieris et al. (2019) presented a strategy based on the Nataf's joint distribution to statistically model water demands in the range 1 min - 1h. Despite these efforts, there are still many issues to discuss about spatial and temporal behaviour of urban water demand, in particular related to the sampling rate of metering devices and its implications on registered (i.e. apparent) and unregistered (i.e. missed) information. Nowadays, monitoring systems in water supply networks combine different sampling frequencies, depending on the technology used for measuring in each case. On many occasions there is no formal knowledge of the sampling effect in the ability of the generated records to explain network behaviour and performance.

The aim of this paper is to analyse the sampling rate (i.e. temporal scale) effect in the capacity of a metering and/or monitoring system to detect network performance. In particular, the analysis here presented focuses on water demand monitoring, as demands highly condition flow variability. Note that demands are the engine that puts water flow in motion along the network, so they determine other network variables, like the pressure regime or water quality. The main question that is to be answered in this work is: how much information is lost because of not using a more frequent sampling period at a particular metering device or demand monitoring system? When a sampling period is set, only the average consumption over the period is recorded, without any further information about the specific demand sequence. This paper intends to characterize the variability of what may occur during the sampling interval, i.e. the importance of the non-recorded behaviour. For this purpose, the analytical approach to SIMDEUM model presented by Díaz and González (2020) is adapted, so that it can assess the statistical properties of water demand over different time scales (i.e. sampling rates). Note that the methodology presented in Díaz and González (2020) analytically provides mean and variance values of instantaneous demands according to the same input parameters than SIMDEUM model, but without the need for Monte Carlo simulations. This is possible by assuming that no correlation among end-uses and end-users exists (i.e. they are independent among each other), and this assumption will be kept in the methodology here presented.

This novel approach enables to analyse different temporal scales (i.e. sampling frequencies) and to explore the effect of the spatial scale law under realistic conditions. Conclusions are relevant for monitoring design, as they help to decide the most suitable sampling time for metering devices, which determines the monitoring potential. Moreover, the presented approach goes one step forward on the path towards the systematic consideration of stochastic demand information (rather than average demand values) in real systems, narrowing the gap between the traditional super-fine scope of stochastic models (1 s - 1 min) and traditional times for water systems analysis (10 min - 1 hour). This is especially important for monitoring applications which, like state estimation, intend to identify the most likely hydraulic state of a system based on all the available data (Kumar et al. 2008; Díaz et al. 2018). Available data in water distribution systems typically include water demand pseudomeasurements (i.e. estimations from historical data) but also readings from metering devices (Díaz et al. 2016b), which can in turn be associated with volumetric measurements every hour (e.g. volumetric smart meters at house connections) or high-frequency flow measurements (e.g. flow meters at water mains). The present paper contributes to finding a solid scaling law that can be relied upon in order to make available sources of information compatible among each other. As mentioned before, the analytical approach here presented is inspired on the microcomponent-based

SIMDEUM model. This makes it a suitable tool for such an objective, because it is grounded in a realistic model but it can be treated analytically, thus minimizing computational time.

The rest of the paper is organised as follows. First, the methodology to analytically compute statistical properties of water demands over different temporal scales is presented. This includes a brief explanation of the basics behind the analytical approach for computing mean instantaneous demands and instantaneous demand variances presented by Díaz and González (2020), and the transition required to assess internal variability over a given period. Then, the method is applied to Benthuizen case study (Blokker et al. 2011a) and validated with equivalent Monte Carlo simulations. Once validated, microcomponent-based analytical results are discussed to assess temporal and spatial scale effects. Finally, conclusions are concisely drawn.

METHODOLOGY

Analytical approach for instantaneous statistical properties of water demands

The analytical approach presented by Díaz and González (2020) is based on the original SIMDEUM model for residential water demand (Blokker et al. 2010) and it uses the same input information in order to assess the same end-uses (bathroom tap, outside tap, water closet -WC-, bathtub, shower, dishwasher, washing machine and kitchen tap). Its key assumption is that it considers the activation/opening of each end-use, each inhabitant and each household independent among each other. This implies the absence of covariance terms all along and guarantees that mean and variance values can be progressively added up to consider spatial aggregation on demand variability (Buchberger and Wu 1995; Magini et al. 2008). External factors that cause correlation among population groups can be considered in the model parametrization, by varying parameters in distribution functions and events' probabilities. This limits the random process to the individual behaviour of each end-user, who operates individually under the assumed external factors.

Mean instantaneous demand

Considering that water demands are random variables, the mean instantaneous demand at a specific time t and level of spatial aggregation x ($\mu_{t,x}$) can be computed by adding mean values of water consumption for all inhabitants j ($\mu_{hab_{j_t}}$) and kitchen tap end-uses k ($\mu_{ktap_{k_t}}$) within a household i ($\mu_{hou_{i_t}}$):

$$\mu_{t,x} = \sum_{i=1}^{n_{hou}} \mu_{hou_{i_t}} = \sum_{i=1}^{n_{hou}} \left(\sum_{j=1}^{n_{hab_i}} \mu_{hab_{j_t}} + \sum_{k=1}^{4} \mu_{ktap_{k_t}} \right). \tag{1}$$

Note that this equation treats the kitchen tap separately because, as assumed in Blokker et al. (2010), this end-use is typically associated with common activities for all household inhabitants (there are four k values to account for four activities: consumption, doing dishes, washing hands and others). The mean value for each inhabitant ($\mu_{hab_{j_t}}$) must consider the individually-activated end-uses u (rest of taps, WC, etc.), each of which is associated with a mean demand μ_{u_t}:

$$\mu_{hab_{j_t}} = \sum_{u=1}^{n_{use}} \mu_{u_t}. \tag{2}$$

It can be derived that μ_{u_t} can be computed as:

$$\mu_{u_t} = \mu_{N_u} \cdot P_{ou}(t) \cdot \mu_{i_u}, \tag{3}$$

where μ_{N_u} represents the mean frequency of use for that particular end-use (i.e. mean number of openings per day), $P_{ou}(t)$ is the unitary probability of one opening of the end-use u being on/open at time t and μ_{i_u} is the mean intensity of the end-use when it is open. $P_{ou}(t)$ can be computed according to the typical duration Cumulative Distribution Function (CDF) for each end-use and each inhabitant's CDF along a day (i.e. daily pattern). In this work, each inhabitant is considered to behave according to one out of the five different types of users (people who work from home, people who do not work, senior people, teenagers and children) identified by Blokker et al. (2010) at The Netherlands. The reader may refer to Díaz and González (2020) for details. Note that computing the mean instantaneous demand value with Eqs. (1)-(3) implies that the population that uses water at a specific location within the network (i.e. level of spatial aggregation x) and at a specific time t behaves according to these five daily patterns, which are conditioned by identical external factors.

Instantaneous demand variance

Under independence hypotheses, demand variance at a specific time and level of spatial aggregation ($\sigma^2_{t,x}$) can be computed by adding demand variances for all inhabitants:

$$\sigma^2_{t,x} = \sum_{i=1}^{n_{hou}} \sigma^2_{hou_{i_t}} = \sum_{i=1}^{n_{hou}} \left(\sum_{j=1}^{n_{hab_i}} \sigma^2_{hab_{j_t}} + \sum_{k=1}^{4} \sigma^2_{ktap_{k_t}} \right). \tag{4}$$

Eq. (4) is analogous to Eq. (1). Similarly to Eq. (2), it can be stated that:

$$\sigma^2_{hab_{j_t}} = \sum_{u=1}^{n_{use}} \sigma^2_{u_t}. \tag{5}$$

The variance of each end-use ($\sigma^2_{u_t}$) can in turn be computed as:

$$\sigma^2_{u_t} = \mu_{N_u} \cdot P_{ou}(t) \cdot \left(\sigma^2_{i_u} + (\mu_{i_u} - \mu_{u_t})^2 \right) + (1 - \mu_{N_u} \cdot P_{ou}(t)) \cdot \mu^2_{u_t}, \tag{6}$$

where $\sigma^2_{i_u}$ is the variance of the intensity of the end-use u when it is open. Note that Eq. (6) presents two terms: the left-hand side refers to the possibility of the end-use being on and the right-hand side refers to the possibility of the end-use being off. In Eq. (3) the off-term disappeared because the intensity of the end-use when it is closed is assumed to be zero (i.e. water-tight closure). In Eq. (6) it remains because even though the variance when the end-use is closed is also taken as zero, the second moment of a variable with respect to a position displaced from the origin must consider the distance between such points (Haan 1977). This explains why μ_{u_t} is involved in Eq. (6).

From instantaneous variability to internal variability over a time period

Previous findings refer to instantaneous properties of water consumption. In this paper, a methodology for computing the statistical properties of microcomponent-based stochastic water demand over a specific time interval is presented. Fig. 1 helps to better understand the concept of internal variability and its connection to readings from metering devices and instantaneous properties. The first graph within this figure shows the water demand series that may take place one particular day (i.e. realization 1) at time t over a specific time period Δt and a spatial aggregation level x. In that particular scenario, a flow meter with a Δt sampling rate would register the accumulated value of water flow over the time interval, which can be understood as an average value $m_{1_{t,\Delta t,x}}$. Due to the stochastic nature of water demands, the same reading at a different day is very likely to be different (realization 2 with $m_{2_{t,\Delta t,x}}, \ldots,$ n with $m_{n_{t,\Delta t,x}}$), so the mean ($\mu_{m_{t,\Delta t,x}}$) and the variance ($\sigma^2_{m_{t,\Delta t,x}}$) of the recorded readings can be computed. These properties can be understood as "apparent" or registered statistical properties, i.e. properties that can be computed based on available records. However, the internal variability within the sampling rate at each realization is not recorded by the metering device whatsoever. There is an internal oscillation over the time period for each realization r, which can be expressed as $z_{r_{t,\Delta t,x}}; \forall r = 1, \ldots, n$. This "missed" variance ($\sigma^2_{z_{t,\Delta t,x}}$) would help to assess the effect of the sampling rate selection on water demand characterization.

Three statistical properties are computed for water demand at each time t for different Δt at a particular level of spatial aggregation x along this paper:

- The apparent average, which is the mean of the average water demands over the Δt time period ($\mu_{m_{t,\Delta t,x}}$). Therefore, it is comparable to the average of registered readings from metering devices with the same Δt sampling rate. It can be computed as:

$$\mu_{m_{t,\Delta t,x}} = \frac{m_{1_{t,\Delta t,x}} + m_{2_{t,\Delta t,x}} + \cdots + m_{n_{t,\Delta t,x}}}{n} \tag{7}$$

  As such mean is considering all possible solutions for water demand over the period, it must be equal to the mean instantaneous demand (Eqs. 1-3) if the mean instantaneous demand keeps constant along the time interval, i.e. $\mu_{m_{t,\Delta t,x}} = \mu_{t,x}$.

- The missed variance, which is the variance of the internal deviations within the time period ($\sigma^2_{z_{t,\Delta t,x}}$). This value cannot be obtained from readings from metering devices, but it could theoretically be computed as the mean of the internal variability of water demands within

Δt:

$$\sigma^2_{z_{t,\Delta t,x}} = \frac{s^2_{1_{t,\Delta t,x}} + s^2_{2_{t,\Delta t,x}} + \cdots + s^2_{n_{t,\Delta t,x}}}{n}, \tag{8}$$

where $s^2_{r_{t,\Delta t,x}}$ can be computed over the interval for each realization r by considering the distribution function of the noise over $\tau \in [0, \Delta t]$, i.e. $f(z_{r_{t,\Delta t,x}}(\tau))$:

$$s^2_{r_{t,\Delta t,x}} = \int_{\Delta t} \left(z_{r_{t,\Delta t,x}}(\tau) - \mathrm{E}[z_{r_{t,\Delta t,x}}(\tau)]\right)^2 \cdot f(z_{r_{t,\Delta t,x}}(\tau)) \cdot d\tau; \forall r = 1, \ldots, n \tag{9}$$

As $z_{r_{t,\Delta t,x}}(\tau)$ deviations are defined as noise:

$$\mathrm{E}[z_{r_{t,\Delta t,x}}(\tau)] = \int_{\Delta t} z_{r_{t,\Delta t,x}}(\tau) \cdot f(z_{r_{t,\Delta t,x}}(\tau)) \cdot d\tau = 0; \forall r = 1, \ldots, n \tag{10}$$

As the time interval decreases, the variability during the interval will tend to zero. As Δt increases, missed variance will become closer to the instantaneous demand variance, i.e. it will consider a broader time period and therefore it will become closer to the instantaneous value. This will be demonstrated and discussed later over results.

- The apparent variance, which is the variance of the average water demands over the Δt period ($\sigma^2_{m_{t,\Delta t,x}}$). Therefore, it is comparable to the variance of the readings from a metering device with a Δt sampling frequency. It can be computed based on the instantaneous demand variance (Eqs. 4-6) and the missed variance (Eq. 8). Note that any water demand record at a time t and spatial level of aggregation x on a r day (i.e. realization r) can be computed as:

$$Q_{r_{t,x}} = m_{r_{t,\Delta t,x}} + z_{r_{t,\Delta t,x}} \tag{11}$$

Therefore, its variance should be computed as:

$$\sigma^2_{Q_{r_{t,x}}} = \sigma^2_{m_{r_{t,\Delta t,x}}} + \sigma^2_{z_{r_{t,\Delta t,x}}} + 2 \cdot \mathrm{Cov}(m_{r_{t,\Delta t,x}}, z_{r_{t,\Delta t,x}}); \tag{12}$$

As internal deviations $z_{r_{t,\Delta t,x}}$ are by definition associated with an expected value equal to zero

(Eq. 10) and $m_{r_{t,\Delta t,x}}$ is constant for a specific recording period, Eq. (12) can be simplified as:

$$\sigma^2_{Q_{r_{t,x}}} = \sigma^2_{m_{r_{t,\Delta t,x}}} + \sigma^2_{z_{r_{t,\Delta t,x}}}; \tag{13}$$

By averaging this expression over all possible realizations within an homogeneous time-period:

$$\sigma^2_{t,x} = \sigma^2_{m_{t,\Delta t,x}} + \sigma^2_{z_{t,\Delta t,x}} \tag{14}$$

And thus the apparent variance can be computed as the difference between the instantaneous demand variance ($\sigma^2_{t,x}$) and the missed variance ($\sigma^2_{z_{t,\Delta t,x}}$):

$$\sigma^2_{m_{t,\Delta t,x}} = \sigma^2_{t,x} - \sigma^2_{z_{t,\Delta t,x}} \tag{15}$$

Analytical approach for water demands statistical properties over a time interval

In order to simplify the formulation that is to be derived to statistically characterize water demand variability over a time period, steady conditions are here assumed around time t. This is valid when small temporal scales are considered, which is reasonable given that sampling rates (i.e. temporal scales for metering devices) in water supply systems are traditionally below the hour. If this analysis had to be extrapolated to greater time intervals, seasonality would have to be considered and non-homogeneous behaviours should be taken into account.

Apparent average over a time interval

Eqs. (1)-(2) can be adapted in order to provide the apparent average of water demands ($\mu_{m_{t,\Delta t,x}}$) for a particular time t, time interval Δt and spatial aggregation level x:

$$\mu_{m_{t,\Delta t,x}} = \sum_{i=1}^{n_{hou}} \left(\sum_{j=1}^{n_{hab_i}} \sum_{u=1}^{n_{use}} \mu_{m_{u_{t,\Delta t}}} + \sum_{k=1}^{4} \mu_{m_{ktap_{k_{t,\Delta t}}}} \right). \tag{16}$$

It must be highlighted that $\mu_{m_{t,\Delta t,x}}$ is here computed by aggregating individual $\mu_{m_{u_{t,\Delta t}}}$ for each end-use u, including the bathroom tap, outside tap, WC, bathtub, shower, dishwasher and washing machine. Like before, the kitchen tap ($\mu_{m_{ktap_{k_{t,\Delta t}}}}$) is considered separately in Eq. (16), but it can be calculated like any other tap, and hence it can be considered as an additional ordinary use $\mu_{m_{u_{t,\Delta t}}}$.

Computing the apparent average of water demands for an end-use over a time interval requires considering all the possible pulse scenarios that may take place over Δt. This implies having to evaluate the probabilities of having a different number of pulses over the selected time interval. The procedure adopted to compute such probabilities is:

1. Assuming that each pulse arrives according to a PRP process (Buchberger and Wu 1995), compute the probability of p pulses taking place for a particular end-use u over one day (PRP_u):

$$PRP_{u_{p+1}} = \frac{\mu_{N_u}^{p} \cdot e^{-\mu_{N_u}}}{p!}; \text{ for } p = 0, 1, \ldots, p_{max} \tag{17}$$

   In this work, $p_{max} = 6 \cdot \text{ceil}(\mu_{N_u})$ because this guarantees that $\sum_{p=0}^{p_{max}} PRP_{u_{p+1}}$ is equal to 1 with a 10^{-4} tolerance. Note that $\text{ceil}(\mu_{N_u})$ is the ceiling function, i.e. it rounds up the number of openings per day to the nearest integer number. Table 1 shows all input parameters for different end-uses in the Netherlands to highlight the variability of their frequency of use.

2. Evaluate the probability of pulses falling within Δt. This requires making multiple combinations that consider that all, some or none of the openings per-day are taking place within the interval. The probability of one single pulse of end-use u falling in the interval ($PS_{u_{t,\Delta t}}$, where S stands for single) can be computed as:

$$PS_{u_{t,\Delta t}} = f_j(t) \cdot (\Delta t + \mu_{d_u}), \tag{18}$$

   where $f_j(t)$ represents the slope of the daily pattern CDF for inhabitant j at time t and μ_{d_u} is the average duration for the particular end-use. As in Blokker et al. (2010), five different types of inhabitant (people who work from home, people who do not work, senior people, teenagers and children) are here assumed. Eq (18) uses $\Delta t + \mu_{d_u}$ in order to consider all the pulses that partly fall in the interval (i.e. they have an initial time of up to μ_{d_u} before Δt starts). This guarantees that when the time interval tends to zero, the probability corresponds to the instantaneous value. Note that taps are typically taken to follow a lognormal CDF for duration, and this complicates probability calculations in Díaz and González (2020). The reason why $PS_{u_{t,\Delta t}}$ keeps simple even for taps is that $f_j(t)$ values are assumed to be constant along the day and equal to the slope of the CDF at that time (i.e. steady state assumption). Once $PS_{u_{t,\Delta t}}$ is obtained, it is necessary to make the convenient combinations. These combinations are organized within a matrix $P_{u_{t,\Delta t}}$ of dimensions $(p_{max}+1) \times (p_{max}+1)$. Rows p represent the possible number of openings per day (from 0 to p_{max}) and columns c the number of openings that may fall within the time interval (from 0 to p). Such a lower triangular matrix can be built as:

$$P_{u_{t,\Delta t_{p+1,c+1}}} = \frac{p!}{c!\,(p-c)!} \cdot PS^{c}_{u_{t,\Delta t}} \cdot \left(1 - PS_{u_{t,\Delta t}}\right)^{p-c};\ \text{for } p = 0, 1, \ldots, p_{max}; c = 0, 1, \ldots, p \tag{19}$$

Therefore, the sum of each row within $P_{u_{t,\Delta t}}$ is equal to 1 according to the previously defined tolerance. Fig. 2 shows the matrix construction process for a particular end-use u.

3. Compute the joint probability of the pulse for that end-use ($PPRP_{u_{t,\Delta t}}$) as:

$$PPRP_{u_{t,\Delta t_{p+1,c+1}}} = P_{u_{t,\Delta t_{p+1,c+1}}} \cdot PRP_{u_{p+1}};\ \text{for } p = 0, 1, \ldots, p_{max}; c = 0, 1, \ldots, p_{max} \tag{20}$$

   This implies that $\sum_{p=0}^{p_{max}} \sum_{c=0}^{p_{max}} PPRP_{u_{t,\Delta t}} = 1$.

The sum of the terms in each column of $PPRP_{u_{t,\Delta t}}$ represents the probability of finding c pulses for that end-use in the interval, with $c = 0, 1, \ldots, p_{max}$. Therefore, the apparent average for end-use u can be computed as:

$$\mu_{m_{u_{t,\Delta t}}} = \sum_{c=0}^{p_{max}} \left[\left(\sum_{p=0}^{p_{max}} PPRP_{u_{t,\Delta t_{p+1,c+1}}} \right) \cdot \mu_{u_{t,\Delta t_c}} \right] \tag{21}$$

Note that the probability is here multiplied by the mean intensity over the interval considering c pulses within the interval ($\mu_{u_{t,\Delta t_c}}$). Mean intensity can be computed by multiplying the c number of pulses within the interval by the intensity of the end-use when it is on (μ_{i_u}) by the mean duration of the pulse (μ_{d_u}), over the possible initial times so that the pulse falls in the interval ($\Delta t + \mu_{d_u}$):

$$\mu_{u_{t,\Delta t_c}} = c \cdot \frac{\mu_{i_u} \cdot \mu_{d_u}}{\Delta t + \mu_{d_u}};\ \text{for } c = 0, 1, \ldots, p_{max} \tag{22}$$

Missed variance over a time interval

As end-uses are considered independent all along this paper, Eqs. (4)-(5) can be converted to provide the missed variance when considering a time interval Δt ($\sigma^2_{z_{t,\Delta t,x}}$) by aggregating the missed variance for each end-use u ($\sigma^2_{z_{u_{t,\Delta t}}}$):

$$\sigma^2_{z_{t,\Delta t,x}} = \sum_{i=1}^{n_{hou}} \left(\sum_{j=1}^{n_{hab_i}} \sum_{u=1}^{n_{use}} \sigma^2_{z_{u_{t,\Delta t}}} + \sum_{k=1}^{4} \sigma^2_{z_{ktap_{k_{t,\Delta t}}}} \right). \tag{23}$$

Analogously to Eq. (21), the missed variance for each end-use u has to be computed considering the probabilities of a different number of pulses falling within Δt and the variance related to such pulses:

$$\sigma^2_{z_{u_{t,\Delta t}}} = \sum_{c=0}^{p_{max}} \left[\left(\sum_{p=0}^{p_{max}} PPRP_{u_{t,\Delta t_{p+1,c+1}}} \right) \cdot \sigma^2_{u_{t,\Delta t_c}} \right], \tag{24}$$

where $\sigma^2_{u_{t,\Delta t_c}}$ is the variance of the intensity within the interval considering c pulses within Δt. This variance can be computed in a simplified way by multiplying the variance of one single pulse falling within the interval ($\sigma^2_{u_{t,\Delta t_1}}$) by the number of pulses that actually fall in the interval (c):

$$\sigma^2_{u_{t,\Delta t_c}} = c \cdot \sigma^2_{u_{t,\Delta t_1}};\ \text{for } c = 0, 1, \ldots, p_{max} \tag{25}$$

This implies that the method will not be able to properly take into account pulse overlap when computing variability. This simplification will provide an underestimation of missed variance within the interval, but the application of the method to a case study will prove that this is negligible due to the low probability of overlap taking place.

In order to simplify the equations for the missed variance of one single pulse falling within the interval ($\sigma^2_{u_{t,\Delta t_1}}$), the formulation is here derived for uses that imply a fixed-intensity discharge of water over a fixed duration, and then extended to the rest (random intensity and random duration). According to Table 1, fixed intensity and duration end-uses are WCs, bathtubs, dishwashers and washing machines, whereas taps and showers are more random uses. Please note that dishwashers and washing machines discharge water over several cycles within the full duration of the end-use. As $f_j(t)$ are considered constant over the day for each inhabitant j, they can be simplified as single discharge end-uses with a mean number of openings equal to $\mu_{N_u} \cdot n_{cycles}$.

If only uses associated with a fixed duration and intensity are considered, it can be stated that $\mathrm{E}[d_u] = \mu_{d_u}$ and $\mathrm{E}[i_u] = \mu_{i_u}$. Computation of $\sigma^2_{u_{t,\Delta t_1}}$ for these end-uses must cover two scenarios:

1. The expected value of the pulse duration is equal or lower than the time interval ($\mu_{d_u} \leq \Delta t$). At the same time, two possible situations must be assessed:

    - The pulse falls fully within the time interval. The upper-left part in Fig. 3 shows that the probability of the pulse falling fully within the time interval (P_{1full}) can easily be computed, and so can the associated average (m_{1full}) and variability (s^2_{1full}) values for that particular realization.
    - The pulse falls partly within the time interval, with only a (s) within Δt. The bottom-left part in Fig. 3 gathers the probability (P_{1part}), average (m_{1part}) and variability (s^2_{1part}) of the pulse falling partly within the time interval. Note that the a duration has been averaged over its possible values ($a \in [0, \mu_{d_u}]$).

    A value for $\sigma^2_{u_{t,\Delta t_{1-1}}}$ for this first scenario can be obtained by computing the weighted average of the two scenarios within 1.

2. The expected value of the pulse duration is greater than the time interval ($\mu_{d_u} > \Delta t$). Similarly, two possible situations have to be considered:

    - The pulse falls fully within the time interval. The upper-right part in Fig. 3 shows that the probability of the pulse falling fully within the time interval (P_{2full}) can be computed, and so can the associated average (m_{2full}) and variability (s^2_{2full}) values for that particular case. Note that as the pulse duration is greater than the time interval, the average corresponds to the intensity of the end-use and variability is null over the interval.

- The pulse falls partly within the time interval, with only a (s) within Δt. The bottom-right part in Fig. 3 gathers the probability (P_{2part}), average (m_{2part}) and variability (s^2_{2part}) of the pulse falling partly within the time interval. Note that the a duration has been averaged over its possible values ($a \in [0, \mu_{d_u}]$).

The weighted average of the variability provides $\sigma^2_{u_{t,\Delta t_{1-2}}}$ under the second scenario.

According to Fig. 3, the variance of one pulse falling within Δt ($\sigma^2_{u_{t,\Delta t_1}}$) can be computed as:

$$\sigma^2_{u_{t,\Delta t_1}} = \begin{cases} \frac{\Delta t - \mu_{d_u}}{\Delta t + \mu_{d_u}} \cdot \frac{\mu^2_{i_u}}{\Delta t} \cdot \left(\mu_{d_u} - \frac{\mu^2_{d_u}}{\Delta t}\right) + \frac{2 \cdot \mu_{d_u}}{\Delta t + \mu_{d_u}} \cdot \frac{\mu^2_{i_u}}{\Delta t} \cdot \left(\frac{\mu_{d_u}}{2} - \frac{\mu^2_{d_u}}{3 \cdot \Delta t}\right) & \text{if } \mu_{d_u} \leq \Delta t \\ \frac{\mu^2_{i_u} \cdot \Delta t}{3 \cdot (\Delta t + \mu_{d_u})} & \text{if } \mu_{d_u} > \Delta t \end{cases} \tag{26}$$

Eq. (26) can be reorganized in order to group the summands as if it was a polynomial on μ_{d_u}, $\mu^2_{d_u}$ and $\mu^3_{d_u}$. Actually, this equation can also be written in a more general way by substituting $\mu_{d_u} = \mathrm{E}[\mu_{d_u}]$, $\mu^2_{d_u} = \mathrm{E}[\mu^2_{d_u}]$, $\mu^3_{d_u} = \mathrm{E}[\mu^3_{d_u}]$ and $\mu^2_{i_u} = \mathrm{E}[i^2_u]$:

$$\sigma^2_{u_{t,\Delta t_1}} = \begin{cases} \frac{\mathrm{E}[i^2_u]}{\Delta t + \mathrm{E}[d_u]} \cdot \mathrm{E}[d_u] - \frac{\mathrm{E}[i^2_u]}{\Delta t \cdot (\Delta t + \mathrm{E}[d_u])} \cdot \mathrm{E}[d^2_u] + \frac{\mathrm{E}[i^2_u]}{3\Delta t^2 \cdot (\Delta t + \mathrm{E}[d_u])} \cdot \mathrm{E}[d^3_u] & \text{if } \mathrm{E}[d_u] \leq \Delta t \\ \frac{\mathrm{E}[i^2_u] \cdot \Delta t}{3 \cdot (\Delta t + \mathrm{E}[d_u])} & \text{if } \mathrm{E}[d_u] > \Delta t, \end{cases} \tag{27}$$

Note that not only fixed duration and intensity uses but any end-use is now represented in Eq. (27). It is only required to introduce the convenient value for $\mathrm{E}[i^2_u]$, $\mathrm{E}[d_u]$, $\mathrm{E}[d^2_u]$ and $\mathrm{E}[d^3_u]$ according to the assumed intensity and duration distributions in Table 1:

$$\mathrm{E}[i^2_u] = \begin{cases} \mu^2_{i_u} & \text{if intensity is fixed} \\ \mu^2_{i_u} + \sigma^2_{i_u} & \text{if intensity follows a uniform CDF} \end{cases} \tag{28}$$

$$\mathrm{E}[d_u] = \begin{cases} \mu_{d_u} & \text{if duration is fixed} \\ e^{\mu_{dn_u} + \frac{\sigma^2_{dn_u}}{2}} = \mu_{d_u} & \text{if duration follows a lognormal CDF} \end{cases} \tag{29}$$

$$\mathrm{E}[d_u^2] = \begin{cases} \mu^2_{d_u} & \text{if duration is fixed} \\ e^{2\cdot\mu_{dn_u} + 2\cdot\sigma^2_{dn_u}} = \mu^2_{d_u} + \sigma^2_{d_u} & \text{if duration follows a lognormal CDF} \end{cases} \tag{30}$$

$$\mathrm{E}[d_u^3] = \begin{cases} \mu^3_{d_u} & \text{if duration is fixed} \\ e^{3\cdot\mu_{dn_u} + \frac{9}{2}\cdot\sigma^2_{dn_u}} & \text{if duration follows a lognormal CDF,} \end{cases} \tag{31}$$

where μ_{dn_u} and $\sigma^2_{dn_u}$ are the corresponding mean and variance values of the associated normal distribution, which can be computed from lognormal input parameters μ_{d_u} and $\sigma^2_{d_u}$ in Table 1.

The missed variance over Δt can therefore be computed with Eqs. (23)-(25) and (27)-(31).

Apparent variance over a time interval

The apparent variance ($\sigma^2_{m_{t,\Delta t,x}}$) can be computed adapting general Eq. (15) to the reality of the end-uses u within the spatial aggregation level x:

$$\sigma^2_{m_{t,\Delta t,x}} = \sum_{i=1}^{n_{hou}} \left[\sum_{j=1}^{n_{hab_i}} \sum_{u=1}^{n_{use}} \left(\sigma^2_{u_t} - \sigma^2_{z_{u_{t,\Delta t}}} \right) + \sum_{k=1}^{4} \left(\sigma^2_{ktap_{k_t}} - \sigma^2_{z_{ktap_{k_{t,\Delta t}}}} \right) \right]. \tag{32}$$

Note that the instantaneous demand variance for each end-use ($\sigma^2_{u_t}$ or $\sigma^2_{ktap_{k_t}}$) can be computed according to Eq. (6), and calculation of the missed variance over Δt ($\sigma^2_{z_{u_{t,\Delta t}}}$ or $\sigma^2_{z_{ktap_{k_{t,\Delta t}}}}$) has just been explained.

CASE STUDY AND TEMPORAL SCALE EFFECT ANALYSIS

The analytical approach proposed in this paper to compute the statistical properties of water demands over a time interval is here applied to Benthuizen case study, which has been presented in the literature before by Blokker et al. (2011a). This case study is a test area of approximately 140 homes and 300 inhabitants (130 occupied households assumed in this work) located at Benthuizen, a village in the Netherlands. This is convenient given that SIMDEUM was originally developed at the Dutch country (Blokker and Vreeburg 2005), and water use survey-based parameters are well-characterized there. Table 1 gathers overall neighbourhood input parameters for SIMDEUM model, which are here used to run the analytical approach (Blokker et al. 2010). Note that only $f_j(t)$ values (i.e. slope of the daily pattern CDF for each type of end-user) from Blokker et al. (2010) need to be additionally incorporated to start the model, as explained in Díaz and González (2020).

SIMDEUM is nowadays considered a well-fitted model to reality. SIMDEUM model and input parameters for Benthuizen were already validated for this case study in Blokker et al. (2011a). Furthermore, the analytical approach for computing the instantaneous mean and variance values of water demands, taken as starting point in this paper, has already been successfully applied to Benthuizen case study (Díaz and González 2020). The analytical approach proposed in this paper to compute statistical properties of water demands over a time interval is here run based on the same parameters, so it can be considered a good representation of a water supply system reality. The proposed methodology will be here validated by comparing analytical results and equivalent Monte Carlo simulations. Then, results analysis will focus on exploring temporal scale (i.e. sampling rate) effects so that similar rules can be used to assess other networks under different specific conditions.

The formulation proposed in this paper provides the statistical properties of water demands at a time t, considering a specific Δt for a particular level of spatial aggregation x. All along the results section, the spatial aggregation level (x) corresponds to the full Benthuizen test area, so no distinction among end-uses is here made. However, it is important to highlight that in order for these results to be extrapolated to other neighbourhoods and their subjacent water supply systems, the distribution of end-uses must be similar. As highlighted by Díaz and González (2020), heterogeneous end-uses coexist in each water supply system, so results are extendable provided that the distribution of end-uses and inhabitants is similar or maybe changes proportionally. Time t is here varied to assess the temporal scale effect at different times. In order to facilitate the interpretation of results, three times are selected along the discussion: 03:30 (at night, minimum flow values), 08:30 (in the morning, maximum flow values), 20:30 (in the evening, intermediate flow values). In what regards Δt, different values are taken from:

$$\Delta t = \frac{\Delta t_{max}}{2 \cdot n}, \text{ with } n = 0.5, 1, 2, \ldots, \frac{\Delta t_{max}}{2}. \tag{33}$$

Actually, only those associated with integer Δt values are selected (considering seconds as temporal units), because this is a valuable asset for the analytical model validation with Monte Carlo simulations. As $\Delta t_{max} = 3600$ s is considered, 37 Δt values are dealt with in this work, well distributed between $\Delta t = 3600$ s and $\Delta t = 1$ s.

Analytical model validation

In order to validate the methodology presented in this paper, analytically computed interval properties for the full test area are compared to those from Monte Carlo simulations. Monte Carlo simulations are here conceived to simulate real pulses, so that the hypotheses and formulation of the analytical approach are tested. 1000 Monte Carlo simulations are considered in this work for each end-use at each time t, so 1000 water demand scenarios are simulated for $\Delta t = \Delta t_{max}$. As Δt reduces, water demand simulated scenarios are rearranged so that the number of simulations is equal to $2000 \cdot n$.

Fig. 4 shows the statistical properties vs the time interval size according to the analytical approach here presented and Monte Carlo simulations at three different times for the full Benthuizen neighbourhood. The first row of graphs within the figure confirms the correct implementation of the analytical approach, as the apparent average of water demands over different Δt values is coincident with Monte Carlo simulation results. As expected, both methodologies provide a value that coincides with the mean instantaneous demand, which varies with the time of day: lowest values at 03:30, maximum values at 08:30 and intermediate values at 20:30. Graphs in the second row show that there is an almost perfect match for the analytical and numerical approach in terms of the missed variance within Δt. This value clearly grows as Δt increases, attaining values close to the instantaneous demand variance for $\Delta t = \Delta t_{max}$. Note that the missed variability would become even closer to the instantaneous values if the maximum time interval was increased. The last row of pictures in Fig. 4 shows how the apparent variance changes for different time intervals. As demonstrated before, rows 2 and 3 are complementary. They validate that not considering pulse overlap when computing demand variability is a fine simplification in the analytical approach.

In the overall, Fig. 4 shows that the analytical approach works well for Benthuizen case study. It is important to highlight that Monte Carlo simulations are associated with a greater computational effort. The average computational time for the analytical approach is 177.9 s (time required to compute all statistical properties for all Δt at the full test area for each time) in an Intel Core i7-6700 CPU 3.40 GHz 16GB RAM desktop computer (using Matlab R2016a), as opposed to the 11126 s (approximately 3 hours) required to run the equivalent Monte Carlo simulation in the same machine.

Analytical model results

The S-shaped curves in the second and third rows of Fig. 4 are interesting from a practical point of view. They show that the apparent variance over Δt, which is comparable to the variance of measurement readings provided by a meter at a specific location, reduces as the size of the time interval increases. This means that less variability is perceived by the metering and/or monitoring system when considering low sampling rates (i.e. high Δt). The reduction in the apparent variance is associated with an increase in the missed variance: the greater the time interval, the more information that is lost by assuming such a sampling rate and not a higher frequency. As the apparent variance curve can be directly obtained from the missed variance curve and the instantaneous demand variance, from now own we will specifically focus on the S-shaped curve of the missed variance evolution with Δt. This is the key element to connect registered variability with total variability. Next, the properties of this curve are going to be analysed.

A non-dimensional version of the missed variance curves in Fig. 4 (second row) can be computed by dividing the missed variance for different time intervals by the instantaneous demand variance, which is constant under the steady state assumption:

$$\text{Relative missed variance } = \frac{\sigma^2_{z_{t,\Delta t,x}}}{\sigma^2_{t,x}} \tag{34}$$

Fig. 5 shows the relative missed variance for each individual household (in grey) and the full neighbourhood (in black) at 8:30. This figure shows that the shape of the S-curve remains approximately the same no matter the spatial aggregation level. Note that the graph has only been represented at one time because an homogeneous proportion of end-uses has been assumed in this implementation, so its shape remains similar regardless of the time of day. The curve helps to better understand and/or extrapolate the missed variability at a specific location, as it will be discussed in the "Implications" section.

Analytical model results can further be discussed in order to better understand temporal scale effects. For example, a pseudo coefficient of variation ($CV^*_{t,\Delta t,x}$) can be computed in order to assess the relative importance of the missed variance ($\sigma^2_{z_{t,\Delta t,x}}$) over the apparent average ($\mu_{m_{t,\Delta t,x}}$):

$$CV^*_{t,\Delta t,x} = \frac{\sqrt{\sigma^2_{z_{t,\Delta t,x}}}}{\mu_{m_{t,\Delta t,x}}} \tag{35}$$

Note that $\sigma^2_{z_{t,\Delta t,x}}$ and $\mu_{m_{t,\Delta t,x}}$ are not strictly comparable, but their ratio can give a good idea of the mean variability of water demands as long as the mean of the average over the interval remains approximately constant (as it is the case). Fig. 6 shows the evolution of this pseudo coefficient of variation with varying Δt at three different times. These three curves show that as Δt increases, so does the pseudo coefficient of variation. They also illustrate that greater coefficients of variation are obtained for minimum flows (night period), and the curve flattens (i.e. $CV^*_{t,\Delta t,x}$ reduces) as the flow increases. These curves can be normalized by dividing the pseudo coefficient of variation by the instantaneous coefficient of variation at that time ($CV_{t,x}$):

$$CV_{t,x} = \frac{\sqrt{\sigma^2_{t,x}}}{\mu_{t,x}} \tag{36}$$

As it happened with the relative missed variance, the shape of the normalized pseudo coefficient of variation curves would remain similar for different times because it is representative of the proportion of end-uses here assumed.

Conversely, spatial scale effects on interval statistical properties could be assessed. However, as the pseudo coefficient of variation presented in this paper (Eq. 35) tends to the instantaneous coefficient of variation (Eq. 36) for sufficiently long time intervals, results would be very similar to those in Díaz and González (2020). Such study showed that fitted lines of cumulative coefficients of variation vs number of inhabitants or households in a double logarithmic scale have a slope of approximately -0.5 regardless of the time of day being considered. This is due to the fact that when analysing an entity that includes N independent elements with the same mean, variance and coefficient of variation, it can be assumed that the total coefficient of variation is equal to the individual coefficient of variation multiplied by $\frac{1}{\sqrt{N}}$, i.e. a -0.5 slope in a double logarithmic scale. This is exact for homogeneous cases (like in Magini et al. 2008), but Díaz and González (2020) showed that it is still true even in this heterogeneous case study. These same authors also showed that the -0.5 power law can be assumed to compute demand uncertainty from mean instantaneous demand values in absence of better data. The same would apply in this work for sufficiently long Δt values.

Just to give an idea of the potential of a combined spatial and temporal scale effect analysis, Fig. 7 shows the cumulative pseudo coefficient of variation for different Δt and levels of spatial aggregation at Benthuizen case study. The first row of figures shows that the pseudo coefficient of variation is highly affected by the number of households. Note that Fig. 6 is equivalent to the curve in Fig. 7 for the total number of households at the corresponding time. Coefficients of variation considerably increase as the number of households reduces. The same happens in the second row of figures, which assess the spatial scale effect according to the number of inhabitants. These laws could be similarly used in order to estimate the variability of water demand for any particular level of spatial aggregation x, time interval Δt and time of day t.

Implications

Results obtained in this work have several implications in real practice. On the one hand, they can be used to estimate how demand uncertainty would vary for different temporal and spatial scales, thus helping to design a suitable metering and/or monitoring system for a particular network.

Let us imagine that a flow meter with a $\Delta t = 3600$ s sampling rate is located at a branch of the water distribution system at Benthuizen test area that provides water to $x = 60$ households (i.e. spatial aggregation level). Each flow reading will then average the water demand of downstream households every hour, so that the apparent average and apparent variance at different times t can be computed based on measurement records. For the sake of illustration, it will be here assumed that measurement records provide an apparent average at 8:30 of 0.4 l/s ($\mu_{m_{t,\Delta t,x}} = 0.4$ l/s) and an apparent variance equal to 0.005 $\mathrm{l}^2/\mathrm{s}^2$ ($\sigma^2_{m_{t,\Delta t,x}} = 0.005\ \mathrm{l}^2/\mathrm{s}^2$) at the selected location. If the relative missed variance curve is known, the relative missed variance for $\Delta t = 3600$ s can be estimated: $\sigma^2_{z_{t,\Delta t,x}}/\sigma^2_{t,x} \approx 0.93$ for $\Delta t = 3600$ s according to Fig. 5. The relative apparent variance can be computed as 1 minus the relative missed variance, because apparent and missed terms are complementary. Making an analogy with Eq. (34), the combined use of flow records and this S-shaped curve could provide the instantaneous demand variance:

$$\sigma^2_{t,x} = \frac{\sigma^2_{m_{t,\Delta t,x}}}{1 - \sigma^2_{z_{t,\Delta t,x}}/\sigma^2_{t,x}} \left(= \frac{\text{from records}}{\text{from S-shaped curve}} \right) \tag{37}$$

In this particular example $\sigma^2_{t,x} = 0.071\ \mathrm{l}^2/\mathrm{s}^2$. This highlights the interest of the approach, because there is no way to measure instantaneous demand variance unless using a high frequency recording monitoring system, many times unaffordable for practical issues. Once the instantaneous demand variance is obtained, the absolute value of the missed variance can be derived from Eq. (14) in order to quantify the non-recorded variability ($\sigma^2_{z_{t,\Delta t,x}} = 0.066\ \mathrm{l}^2/\mathrm{s}^2$ in this example). Also, the mean instantaneous demand (which can be here taken as the apparent average $\mu_{m_{t,\Delta t,x}} = 0.4$ l/s, i.e. the average of available flow measurements) can be combined with the instantaneous demand variance as in Eq. (36) to compute the instantaneous coefficient of variation ($CV_{t,x} = 0.67$). Knowing that this coefficient of variation relates to the number of independent units under study (i.e. number of households and/or inhabitants) with a -0.5 slope in a double logarithmic scale, it would be possible to assess the uncertainty when considering a different spatial and/or temporal aggregation level. Note that the analytical approach here presented assumes mutually independent behaviours among end-uses and end-users (i.e. demand nodes are not correlated to each other). The reader may refer to Díaz and González (2020) for a detailed explanation of these independence hypotheses. In reality, cross correlation gains importance when aggregating demand over time and space. This may have consequences in some applications, like estimating peak coefficients. This is a subject for further research.

Apart from quantitative results, it is worth qualitatively discussing how current measurement strategies may or may not be suitable for some specific types of analysis. It is clear that water demand variability in relative terms increases as approximating the terminal branches of any water supply system. Therefore, coefficients of variation are higher in the outer fringes, and lower in upstream pipes that deliver water to populated areas. This means that higher temporal resolution is needed near homes, but it can be relaxed in the water mains (Tessendorff 1972). Fig. 4 shows that as Δt grows, the apparent variance reduces. This implies that even though in the outer fringes demand coefficients of variation increase considerably, coefficients of variation would be underestimated if computed only based on apparent variance. Even though coefficients of variation are known to increase in the outskirts of the network, measurement policies are precisely conceived the other way around: volumetric remote meters (which measure usually every hour or even less frequently) are used at the entrance to each household, and flow meters (which may measure every minute) are located in water mains. These sampling rates obey different reasons. They are mainly the result of different measurement technologies, but in some other cases they are oriented to some specific uses, like detecting undeclared manoeuvres or identifying leakage. However, the traditional sampling rate scheme may not be suitable for some specific analyses. For example:

- If pressures at terminal branches of the network are important in a supply system, it will be difficult (or nearly impossible) to accurately characterize their variability with traditional sampling strategies.
- If water quality must be analysed in the outskirts of the network, the existing metering scheme may be sufficient to estimate the mean value of water velocity. This would enable a first approach to water quality assessment, like computing water age. However, all

the processes that relate non-linearly to water velocity would be limited by a common measurement strategy.

- Any methodology conceived to take into account all measurements' uncertainty in order to monitor the state of the system may be limited, as it is the case of state estimation techniques (Díaz 2017). With traditional sampling frequencies, estimations will approximate average values, but the associated uncertainty is going to remain significantly large if temporal resolution is not increased on a distribution level. Moreover, in this type of monitoring applications, it is needed to consider the scale difference between average measurements (e.g. from flow or volumetric meters) and instantaneous readings (e.g. from pressure meters). The approach here presented may help to make compatible the different nature and resolution of measurements, at least in terms of their input uncertainty, which propagates to the estimates (Díaz et al. 2016a).

These scenarios are just mentioned here in order to illustrate how the present methodology can contribute to improve the monitoring and/or management of water systems, but they are out of the scope of this paper. Note that even though demand variability has been widely discussed on a scientific level, its importance has not yet affected the engineering and/or metrological practice. Only by delving into the scaling laws that govern water demands in realistic scenarios, can practitioners be motivated to shift towards this new paradigm of high-resolution temporal and spatial scales.

CONCLUSIONS

This paper analyses temporal scale effects in water supply system demands, which affect monitoring performance, thanks to a novel analytical approach to stochastic demand modelling. The proposed method keeps improving the conceptualization of the well-known SIMDEUM model. Until recently, Monte Carlo simulations implemented in SIMDEUM were the only way of computing high-resolution stochastic demand patterns based on survey parameters. Díaz and González (2020) lately developed an analytical approach that provides instantaneous mean demand and variance thanks to independence hypotheses among end-uses and end-users. The methodology here presented goes one step further and provides an analytical formulation for computing statistical properties of water demands over a time interval, which is of major importance when designing monitoring requirements.

This work differentiates the "apparent" or registered statistical properties that can be computed based on the available records of a meter with a specific sampling rate from the internal oscillations that take place over the time interval, which are not recorded by the metering device whatsoever, i.e. they are "missed" statistical properties. The proposed approach not only enables fast computation of the apparent average and variance (which can be compared to measurement records), but also of the missed variance associated with a particular sampling rate. Results obtained for Benthuizen case study are interesting for different reasons. On the one hand, they make explicit that sampling rates (i.e. temporal scales) associated with specific metering or monitoring systems condition the degree to which the network behaviour and performance can be assessed. Results obtained for this realistic case study may therefore be useful for metering and/or monitoring design in this or other similar networks. On the other hand, results show that there is correspondence between instantaneous and apparent values, and this relationship may be used to characterize the missed variance. This may be useful to rapidly understand or even estimate demand variability thanks to the obtained scaling laws when there is absence of better data. Care must be taken when using records collected with different sampling rates, because they represent different signals with different behaviours. Therefore, the method gives some guidelines for progressive incorporation of high-resolution water demand measurements or estimations in engineering practice.

DATA AVAILABILITY STATEMENT

Some or all data, models, or code used during the study were provided by a third party (Benthuizen details). Direct requests for these materials may be made to the provider as indicated in the Acknowledgements.

ACKNOWLEDGEMENTS

The authors want to thank Dr. Mirjam Blokker for providing the Benthuizen case study details. The authors would also like to thank the financial support provided by the Spanish

Ministry of Science and Innovation - State Research Agency (PID2019-111506RB-I00 / AEI / 10.13039/501100011033) and Junta de Comunidades de Castilla-La Mancha (SBPLY/19/180501/000162, co-financed by European FEDER funds).

REFERENCES

Blokker, E., Beverloo, H., Vogelaar, J., Vreeburg, J., and van Dijk, J. (2011a). "A bottom-up approach of stochastic demand allocation in a hydraulic network model: a sensitivity study of model parameters." *J. Hydroinf.*, 13(4), 714–728.

Blokker, E., Buchberger, S., Vreeburg, J., and van Dijk, J. (2009). "Comparison of water demand models: PRP and SIMDEUM applied to Milford, Ohio, data." *Water Distribution Systems Analysis 2008*, K. Van Zyl, ed., American Society of Civil Engineers.

Blokker, E., Pieterse-Quirijns, E., Vreeburg, J., and van Dijk, J. (2011b). "Simulating nonresidential water demand with a stochastic end-use model." *J. Water Resour. Plann. Manage.*, 137(6), 511–520.

Blokker, E. and Vreeburg, J. (2005). "Monte Carlo simulation of residential water demand: A stochastic end-use model." *World Water and Environmental Resources Congress, Alaska, USA*.

Blokker, E., Vreeburg, J., and van Dijk, J. (2010). "Simulating residential water demand with a stochastic end-use model." *J. Water Resour. Plann. Manage.*, 136(1), 19–26.

Buchberger, S. and Nadimpalli, G. (2004). "Leak estimation in water distribution systems by statistical analysis of flow readings." *J. Water Resour. Plann. Manage.*, 130(4), 321–329.

Buchberger, S. and Wells, G. (1996). "Intensity, duration and frequency of residential water demands." *J. Water Resour. Plann. Manage.*, 122(1), 11–19.

Buchberger, S. and Wu, L. (1995). "Model for instantaneous residential water demands." *J. Hydraul. Eng.*, 121(3), 232–246.

Creaco, E., Blokker, M., and Buchberger, S. (2017). "Models for generating household water demand pulses: Literature review and comparison." *J. Water Resour. Plann. Manage.*, 143(6), 04017013.

Díaz, S. (2017). “Comprehensive approach for on-line monitoring water distribution systems via state estimation related techniques.” Ph.D. thesis, Univ. of Castilla-La Mancha, Spain.

Díaz, S. and González, J. (2020). “Analytical stochastic microcomponent modelling approach to assess network spatial scale effects in water supply systems.” *J. Water Resour. Plann. Manage.*, 146(8), 04020065.

Díaz, S., González, J., and Mínguez, R. (2016a). “Uncertainty evaluation for constrained state estimation in water distribution systems.” *J. Water Resour. Plann. Manage.*, 142(12), 06016004.

Díaz, S., Mínguez, R., and González, J. (2016b). “Stochastic approach to observability analysis in water networks.” *Ingeniería del Agua*, 20(3), 139–152.

Díaz, S., Mínguez, R., and González, J. (2018). “Topological state estimation in water distribution systems: Mixed-integer quadratic programming approach.” *J. Water Resour. Plann. Manage.*, 144(7), 04018026.

Filion, Y., Karney, B., Moughton, L., Buchberger, S., and Adams, B. (2008). “Cross correlation analysis of residential demand in the city of Milford, Ohio.” *Water Distribution Systems Analysis Symposium 2006*, S. G. Buchberger, R. Clark, W. Grayman, and J. Uber, eds., American Society of Civil Engineers.

Haan, C. (1977). *Statistical methods in hydrology*. The Iowa State University Press / Ames, Iowa, USA.

Kossieris, P. and Makropoulos, C. (2018). “Exploring the statistical and distributional properties of residential water demand at fine time scales.” *Water*, 10(10), 1481.

Kossieris, P., Tsoukalas, I., Makropoulos, C., and Savic, D. (2019). “Simulating marginal and dependence behaviour of water demand processes at any fine time scale.” *Water*, 11(5), 885.

Kumar, S. M., Narasimhan, S., and Bhallamudi, S. M. (2008). “State estimation in water distribution networks using graph-theoretic reduction strategy.” *J. Water Resour. Plann. Manage.*, 134(5), 395–403.

Magini, R., Pallavicini, I., and Guercio, R. (2008). “Spatial and temporal scaling properties of water demand.” *J. Water Resour. Plann. Manage.*, 134(3), 276–284.

Rodriguez-Iturbe, I. (1986). "Scale of fluctuation of rainfall models." *Water Resour. Res.*, 22(9S), 15S–37S.

Tessendorff, H. (1972). "Problems of peak demands and remedial measures." *Proc., 9th Congress of Int. Water Supply Association, Int. Standing Committee on Distribution Problems*, Subject n. 2, s10–s14.

Vertommen, I., Magini, R., and Cunha, M. (2015). "Scaling water consumption statistics." *J. Water Resour. Plann. Manage.*, 141(5), 04014072.

Vertommen, I., Magini, R., Cunha, M., and Guercio, R. (2012). "Water demand uncertainty: the scaling laws approach." *Water supply systems analysis: selected topics*, A. Ostfeld, ed., IntechOpen, 105–129.

List of Tables

TABLE 1. Frequency, duration and intensity parameters for the Netherlands according to Blokker et al. (2010)

		Frequency	Duration			Intensity		
		Mean number of openings per day and inhabitant	Distribution	Mean (s)	Variance (s^2)	Distribution	Mean (l/s)	Variance (l^2/s^2)
End-use type	End-use subtype	μ_{N_u}		μ_{d_u}	$\sigma^2_{d_u}$		μ_{i_u}	$\sigma^2_{i_u}$
Kitchen tap	Consumption	4.73*	Lognormal	16	20.8	Uniform	0.083	0.0023
	Doing dishes	3.15*	Lognormal	48	62.4	Uniform	0.125	0.0052
	Washing hands	3.15*	Lognormal	15	19.5	Uniform	0.083	0.0023
	Others	1.58*	Lognormal	37	48.1	Uniform	0.083	0.0023
Bathroom tap	Washing and shaving	1.35	Lognormal	40	52	Uniform	0.042	0.0006
	Brushing teet	2.75	Lognormal	15	19.5	Uniform	0.042	0.0006
Outside tap	Garden	0.33	Lognormal	300	390	Uniform	0.1	0.0033
	Other	0.11	Lognormal	15	19.5	Uniform	0.1	0.0033
WC	9L	6	Fixed	216	-	Fixed	0.042	-
	9L with water saving	6	Fixed	108	-	Fixed	0.042	-
	6L	6	Fixed	144	-	Fixed	0.042	-
	6L with water saving	6	Fixed	72	-	Fixed	0.042	-
Bathtub	-	0.044	Fixed	600	-	Fixed	0.2	-
Shower	No water saving	0.7	Lognormal	510	255	Fixed	0.142	-
	With water saving	0.7	Lognormal	510	255	Fixed	0.123	-
Dishwasher	-	0.3	Fixed	21/cycle**	-	Fixed	0.167	-
Washing machine	-	0.3	Fixed	75/cycle**	-	Fixed	0.167	-

*Frequency for the kitchen tap is per household per day

**4 cycles over 7200s

Source: Data from Blokker et al. (2010).

List of Figures

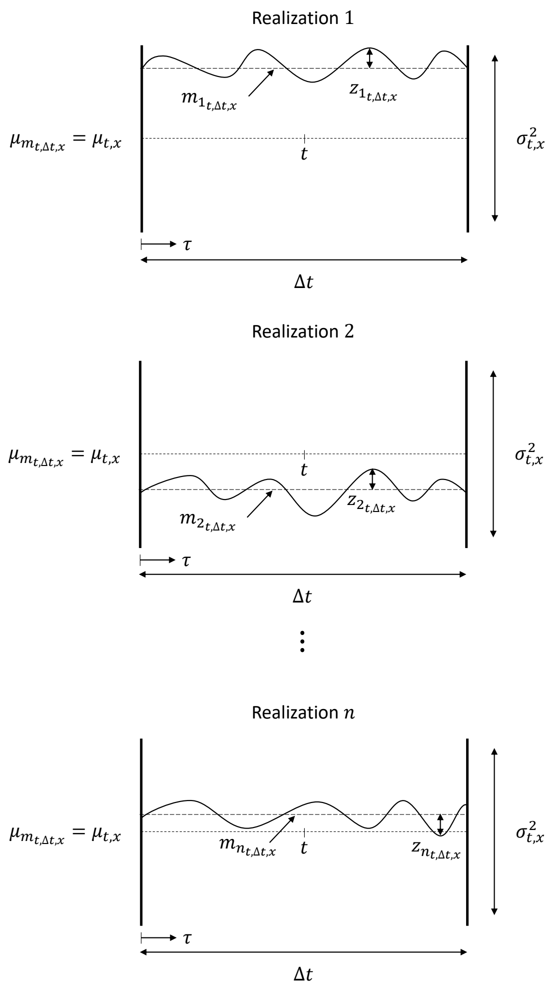

Fig. 1. Instantaneous vs interval statistical properties of water demand at a time t and spatial aggregation level x.

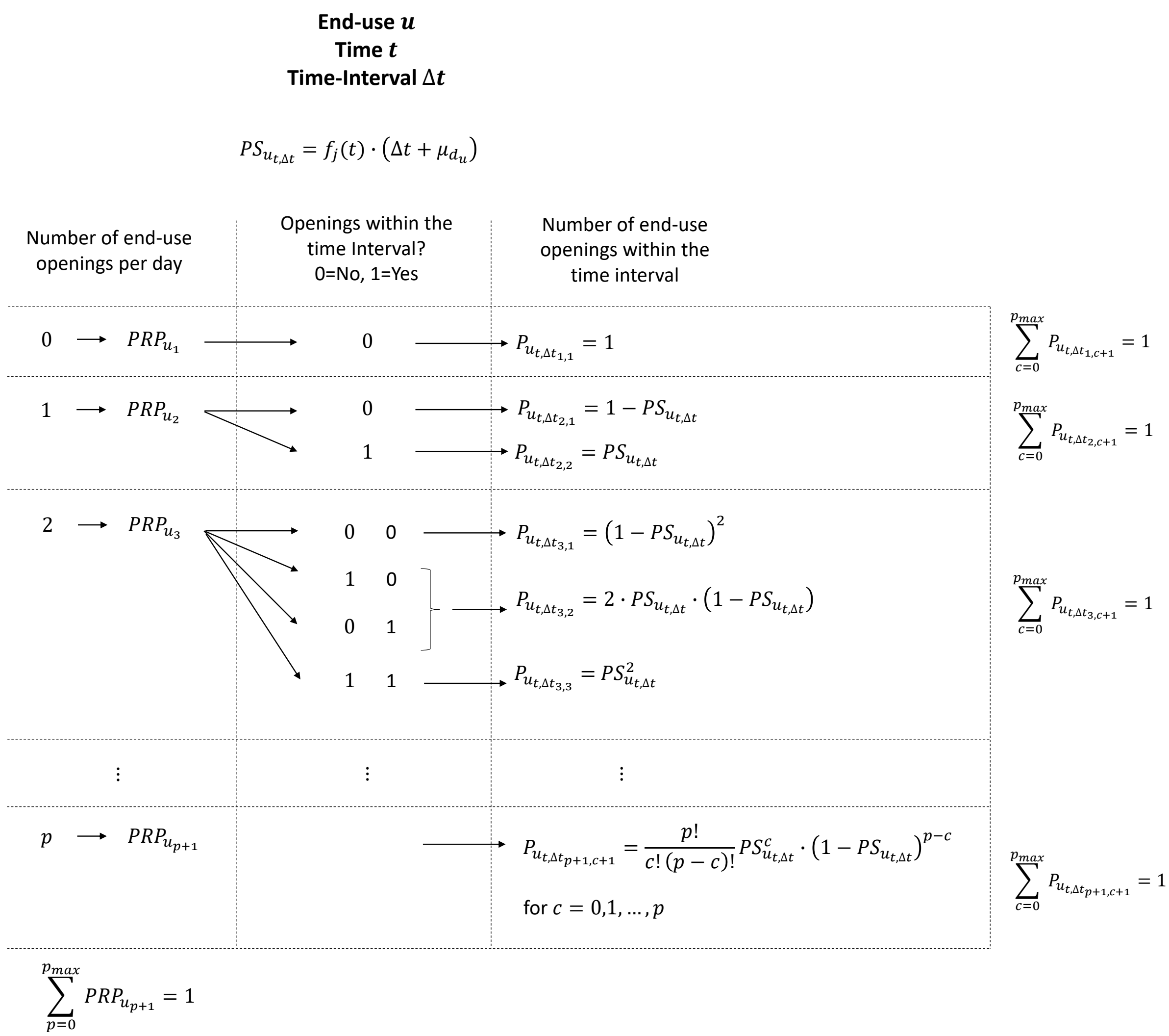

Fig. 2. Construction process for probability matrix $P_{u_{t,\Delta t}}$.

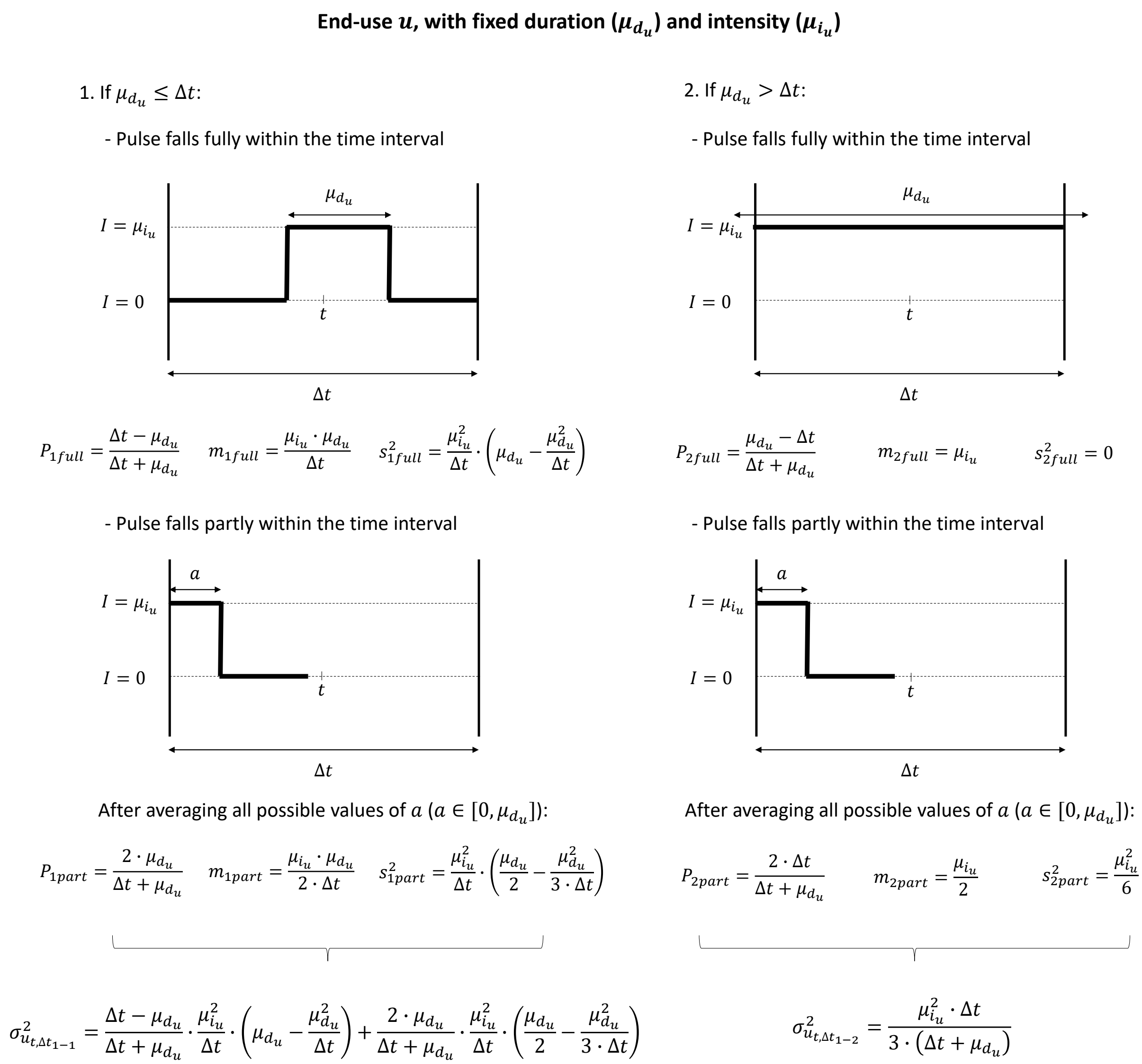

Fig. 3. Scenarios for computing the missed variance of one pulse over a time interval: 3-left for $\mu_{d_u} \leq \Delta t$ and 3-right for $\mu_{d_u} > \Delta t$. Simplification for a fixed duration and intensity end-use.

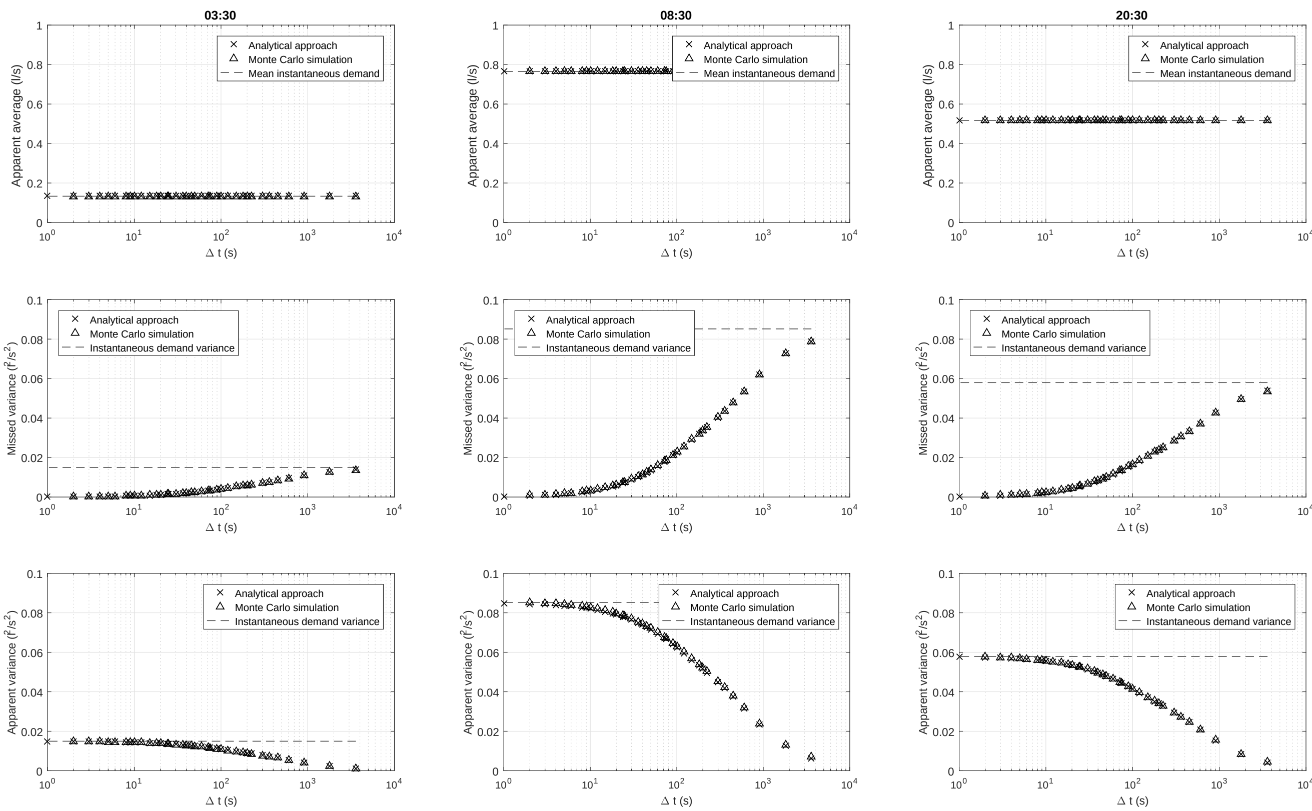

Fig. 4. Statistical properties of water demands over different time intervals (in rows, 4-top apparent average, 4-center missed variance and 4-bottom apparent variance) at three different times (in columns) for the full Benthuizen test area: analytical approach vs Monte Carlo simulation.

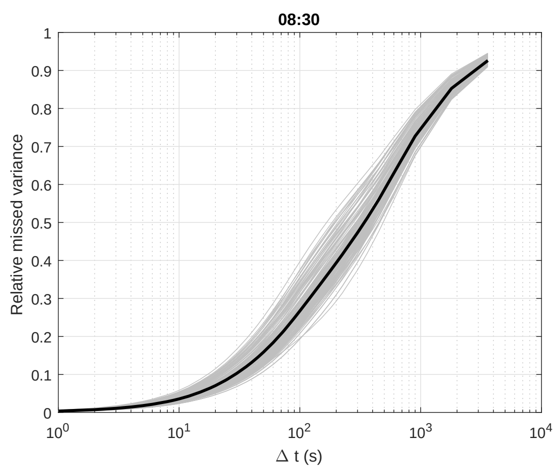

Fig. 5. Relative missed variance with respect to instantaneous demand variance for different time intervals at 8:30: individual households (grey) vs full Benthuizen neighbourhood (black).

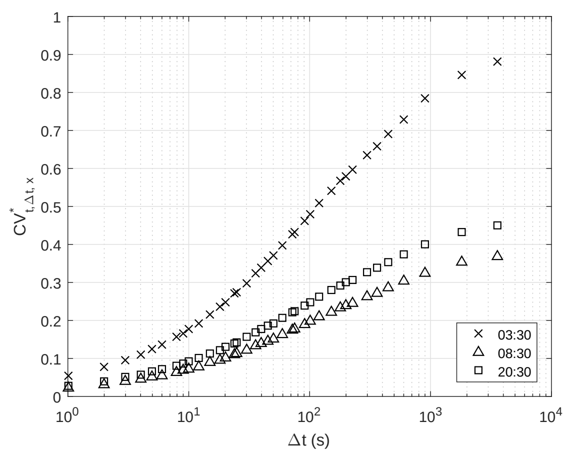

Fig. 6. Evolution of the pseudo coefficient of variation with Δt at three different times at Benthuizen case study.

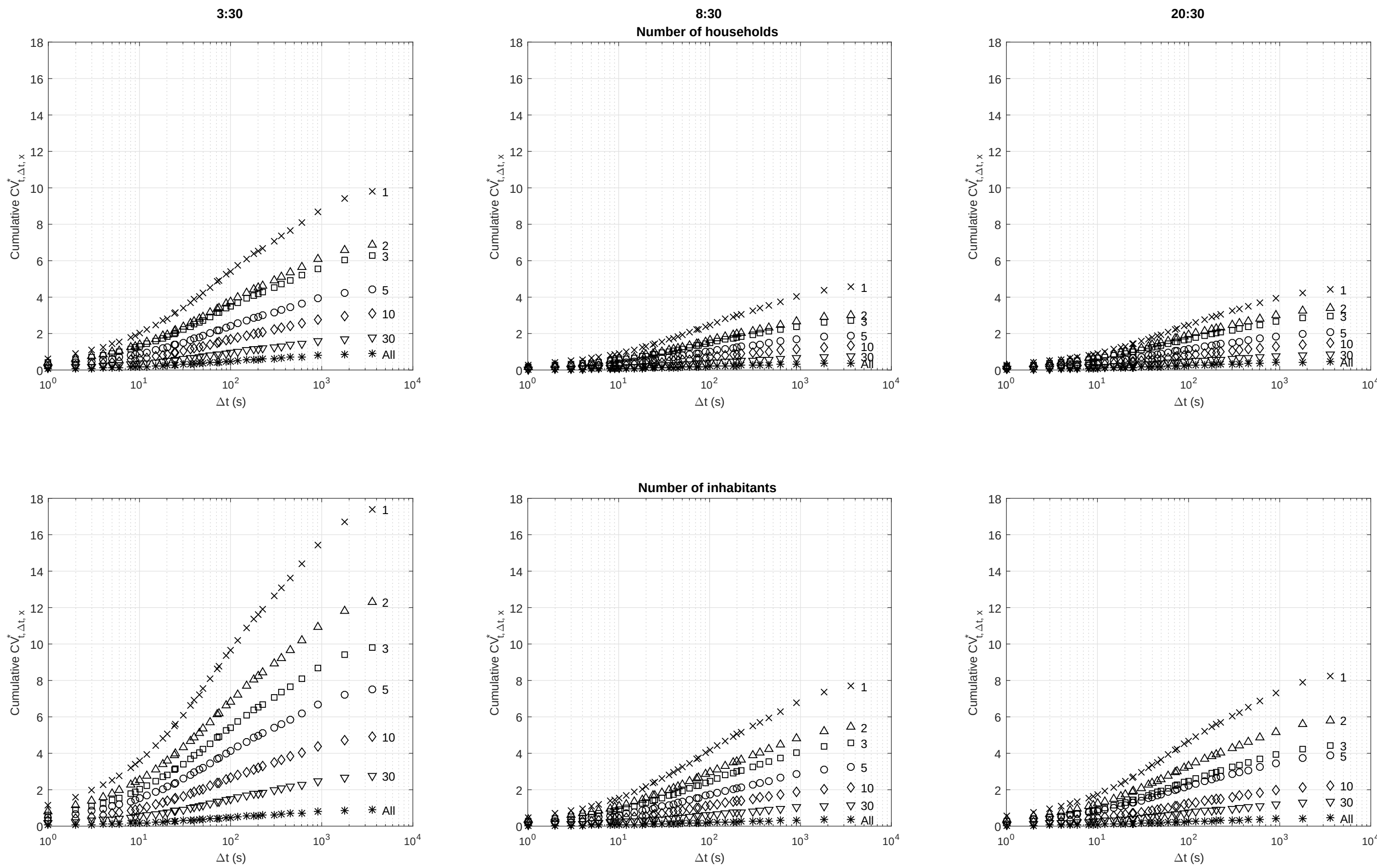

Fig. 7. Cumulative pseudo coefficient of variation over Δt vs time-interval for different times (in columns) and levels of spatial aggregation (in rows, 7-top number of households, 7-bottom number of inhabitants) at Benhtuizen case study.