A bottom-up approach of stochastic demand allocation in a hydraulic network model: a sensitivity study of model parameters

E. J. M. Blokker, H. Beverloo, A. J. Vogelaar, J. H. G. Vreeburg and J. C. van Dijk

ABSTRACT

An ''all pipes'' hydraulic model of a drinking water distribution system was constructed with a bottom-up approach of demand allocation. This means that each individual home is represented by one demand node with its own stochastic water demand pattern. These water demand patterns were constructed with the end-use model SIMDEUM. A sensitivity test with respect to the resulting residence times was performed for several model parameters: time step, spatial aggregation, spatial correlation, demand pattern and number of simulation runs. The bottom-up approach of demand allocation was also compared to the conventional top-down approach, i.e. a single demand multiplier pattern is allocated to all demand nodes with the base demand to account for the average water demand on that node. The models were compared to measured flows and residence times in a small network. The study showed that the bottom-up approach leads to realistic water demand patterns and residence times, without the need for any flow measurements. The stochastic approach of hydraulic modelling, with a 15 minute time step, some spatial aggregation and 10 simulation runs, gives insight into the variability of residence times as an added feature beyond the conventional way of modelling.

Key words | demand allocation, stochastic modelling, water quality modelling

E. J. M. Blokker (corresponding author)
H. Beverloo
A. J. Vogelaar
J. H. G. Vreeburg
KWR Watercycle Research Institute,
P.O. Box 1072, 3430 BB,
Nieuwegein,
The Netherlands
E-mail: *mirjam.blokker@kwrwater.nl*

J. H. G. Vreeburg
J. C. van Dijk
Delft University of Technology,
Department of Civil Engineering and Geosciences,
P.O. Box 5048, 2600 GA Delft,
The Netherlands

INTRODUCTION

The goal of drinking water companies is to supply their customers with good quality drinking water 24 hours a day. With respect to water quality, the focus has for many years been on drinking water treatment. Recently, interest in the water quality of a drinking water distribution system (DWDS) has been growing. Water age is an important aspect of water quality in a DWDS as it influences disinfectant residual, disinfection by-products, nitrification, bacterial regrowth, corrosion, sedimentation, temperature, taste and odour (EPA 2002). More specifically, the maximum water age (or residence time) is most important (Machell *et al.* 2009).

doi: 10.2166/hydro.2011.067

The key element of a water quality model for a DWDS is a detailed hydraulic model (Slaats *et al.* 2003; Vreeburg 2007), which not only takes into account the maximum flows but also the flows at the preceding time steps (Slaats *et al.* 2003; Powell *et al.* 2004; Vreeburg & Boxall 2007). A hydraulic model with an accurate simulation of the occurrence of turbulent and laminar flow and stagnant water is needed. Therefore, knowledge of the water demand on a more detailed level is essential (Blokker *et al.* 2008). The required detail in temporal and spatial scale is to be determined. This paper will investigate the required detail.

One way of improving a model's accuracy is by calibration. Calibration is usually done with pressure measurements by adjusting wall roughness coefficients and status of valves, for which several optimization techniques are available (Kapelan 2002). For this calibration a discernable head loss is required; in the periphery of the drinking water distribution system, where velocities and head losses are low, calibration on pressure is almost impossible. Jonkergouw *et al*. (2008) showed that calibration of demands can be done by using water quality measurements (in their case chlorine levels). They concluded that average daily demands can be determined with high precision, but that substantial measurement errors in the calibration data (i.e. water quality data) do not allow for an accurate calibration of the demand multiplier patterns (DMP) that construct the diurnal pattern. Pasha & Lansey (2010) have shown that water quality predictions of residual chlorine in a DWDS are very sensitive to uncertainty in demand and the bulk and wall reaction coefficients, and hardly sensitive to pipe diameter and wall roughness. Calibration of diameter and wall roughness by means of pressure measurements may therefore not be required. Water quality measurements are preferably used for calibration of reaction coefficients but not for calibration of demands.

Modelling water quality in the DWDS requires a different approach in demand allocation, where the demands show less auto- and cross-correlation and are determined on smaller temporal and spatial scales than the conventional "top-down" approach of demand allocation (Blokker *et al*. 2008). Here, top-down demand allocation means that a DMP (e.g. measured at the pumping station) is allocated to the demand nodes with a base demand to account for the average water demand on that node, thereby applying strongly spatially correlated water demand patterns among all nodes. A different way is to use a "bottom-up" approach of demand allocation. This means that unique stochastic water demand patterns are modelled for each individual home for each day of the week, and a unique water demand pattern is constructed for each demand node by summation of the individual household water demand patterns. In the traditional approach of top-down demand allocation the cross-correlation is assumed to be equal to 1 and the auto-correlation is usually high because a time step of 15 min or 1 h is used. A cross-correlation of 1 results in a limited number of flow direction reversals in a network model. A high auto-correlation means that the flow over the day is relatively constant and the model will show no periods with stagnant water and possibly a limited period of turbulent flow. In case the actual flows are not strongly correlated, flow direction reversals (in looped networks) and periods of stagnancy and turbulent flows will occur. A traditional approach in demand allocation may therefore underestimate maximum residence times and dispersion.

The hypothesis is that a bottom-up approach of demand allocation results in a model with realistic demands, which show more resemblance with real demands with respect to instantaneous peak values and diurnal variability, and therefore leads to realistic residence times. The hypothesis is tested by comparing this bottom-up approach against the traditional top-down approach and to measurement results of a tracer study with a conservative compound. The bottom-up demand allocation was done with the use of the end-use model SIMDEUM (Blokker *et al*. 2010b). This paper presents a sensitivity test for several model parameters. The influence on modelled residence time was tested of the following parameters: demand pattern time step, spatial aggregation of demands, spatial correlation of demand patterns and the shape of the demand pattern. The bottom-up approach is a stochastic modelling approach and each simulation will give different results. The number of simulation runs that is required was also investigated. For this purpose a small DWDS was selected as a test area. The DWDS was operated in its normal looped layout. In order to reduce the effect of mixing, the DWDS was also operated in a branched layout. In order to limit the measurement time and the effect of stagnant water on tracer dispersion, an extra flow was generated in the DWDS. A second study (Blokker *et al*. 2010a) is concerned with the practical implications of the bottom-up approach of demand allocation in a real DWDS.

A bottom-up approach of demand allocation leads to larger hydraulic models with more nodes, more pipes and more numerous demand patterns. Using a smaller time step, means that simulations take longer. As the demand patterns are the results of a Monte Carlo simulation, multiple simulation runs are required to understand the variability of the results. This study aims to understand what level of detail is required and what model simplifications and reductions are acceptable. In this way the total simulation time can be controlled.

METHODS AND MATERIALS

Generic methodology

A small distribution network was selected as a test area. In this network, the total water demand was measured and a tracer study was performed to determine the residence time towards three locations in the network. The network was operated in two different ways, viz. in a looped and a branched layout, and a continuous flow of 400 L/h was extracted.

An "all pipes" hydraulic model was constructed with a bottom-up approach of demand allocation of individual and unique stochastic demand patterns. A second model was constructed with a conventional top-down approach of demand allocation with a common DMP. The model results were compared with respect to the measured flow and residence time. A sensitivity test was done for the models. Demand patterns with various time steps were applied; demand patterns were allocated at the household connection and aggregated on the modelled pipe ends; different sets of demand patterns were applied.

The network

The selected network is situated in the town of Benthuizen in the west of the Netherlands (near The Hague). The network was built in the mid-1970s and consists of 580 m of Ø100 mm asbestos cement pipes, 380 m of Ø110 mm, Ø63 mm and Ø50 mm PVC pipes, and 70 m of Ø80 mm lined cast iron pipes and supplies 144 homes (Figure 1). The pipes have well defined internal diameters and wall roughnesses. The annual water use in the network was determined from the water meter readings of 2004 of the Water Company Dunea. On average, 314 L per home per day was registered. This was confirmed by flow measurements in 2004 on a district metered area (DMA) of ca. 1200 homes (Beuken *et al.* 2006) which encloses the network under study. The 2006 study also showed that this network has no leaks. The supply area Vlietregio, in which Benthuizen is located, supplies ca. 16,000 (mainly residential) connections; its flow is continuously measured. The measured water demand patterns of this supply area of the period 21–30 July 2007 are indicated by DMP_{PS}, where PS stands for pumping station.

The drinking water was distributed without any disinfectant, as is common in the Netherlands. A tracer study with NaCl was performed between 24 and 30 July 2007. Some valves were permanently closed during the measurement period to isolate the area from the rest of the DWDS. Two other valves (0033 and 0035) were operated to set the network layout to either a branched or a looped system. The valves were closed from Tuesday 24 to Thursday 26 July; they were open from Friday 27 to Sunday 29 July. Because dispersion could have a large effect on water quality modelling (Li 2006) the Taylor dispersion in the network was limited by applying an additional demand of 520 L/h. This measure ensured the absence of stagnant water and an average Reynolds number of 5500, i.e. a turbulent flow during most of the day.

Measurement setup for the tracer study

Four measurement locations were selected (Figure 1). Location 1 is located at the entrance of the isolated test area. Locations 2, 3 and 4 are located on the central Ø100 mm AC main. Note that in the branched network layout, the water travels from location 1, to 2, to 3 and then to 4. In the looped network layout, this is not necessarily the case.

Sodium chloride (NaCl) was used as a tracer and the electrical conductivity was measured; from these measurements, the residence time was determined. NaCl has several advantages for use as a tracer, namely at a well measurable dosage, it causes no disruption or health risk to customers; it yields results of good accuracy and it is low-cost (Skipworth *et al.* 2002). At location 1, NaCl was dosed to a fixed concentration in order to raise the electrical conductivity (EC, in mS/m) by a measurable amount: EC $\approx$44 mS/m without dosage, and EC $\approx$58 mS/m with dosage. The tracer was dosed in pulses of four hours on and four hours off. This means that, per day, six positive and six negative step inputs were induced.

In order to reach a fixed concentration, the flow was measured (Endress + Hauser Promay W) and the dosage was controlled. A static mixer ensured a constant concentration of the tracer over the pipe cross-section. To overcome the head loss through the static mixer and to establish a fixed head, a pump was placed at location 1. The instantaneous flow was logged every minute; this resulted in five full days of flow

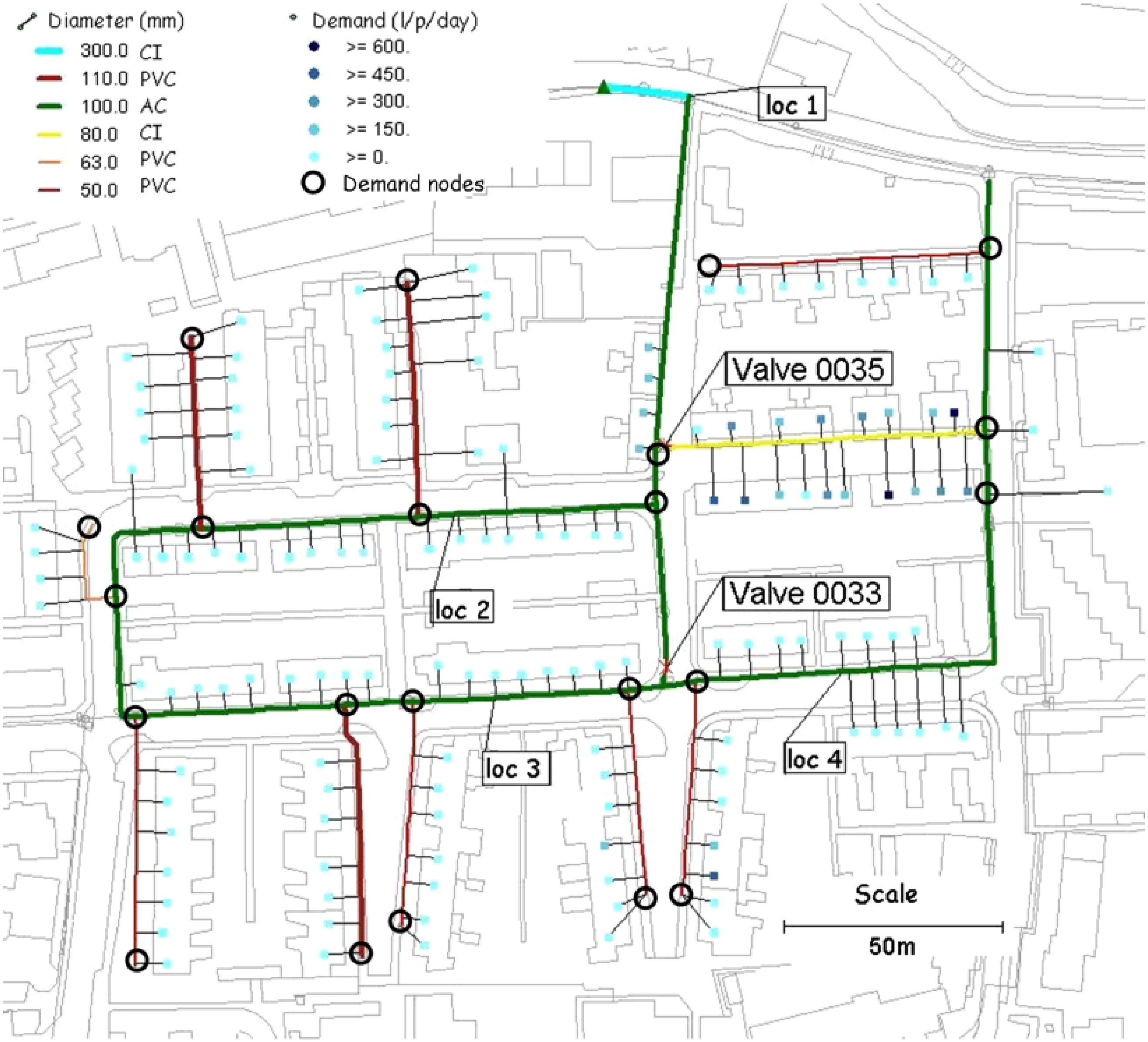

Figure 1 | Network layout. Water enters at location 1; a continuous demand of 400 L/h was extracted at loc. 4. Valves 0033 and 0035 are closed in the branched layout and open in the looped layout. Valve 0035 is placed at the 80 mm CI main.

measurements at location 1 (3 weekdays, 2 weekend days). The measured flow patterns were converted into an average DMP for weekdays and one for weekend days, which is denoted DMP_{REF}.

At all four locations, the EC was measured (LIQUISYS M CLM223). The monitoring systems required a continuous 40 L/h extraction at the measurement locations. An extra 400 L/h was extracted at location 4 to ensure turbulent flows during most of the day.

The residence time between location 1 and 2, 3 and 4 respectively was determined from the time between the centres of the ascending and descending tails (at ca. 51 mS/m) of the measured EC pulses. Because dispersion was limited, the pulses retained their shapes. Figure 2 shows for the branched layout that at 15:12 the residence time between location 1 and 2 was equal to 0.5 h; at 18:37 the residence time between location 1 and 2 was equal to 0.3 h. The residence time varies over the day and between days.

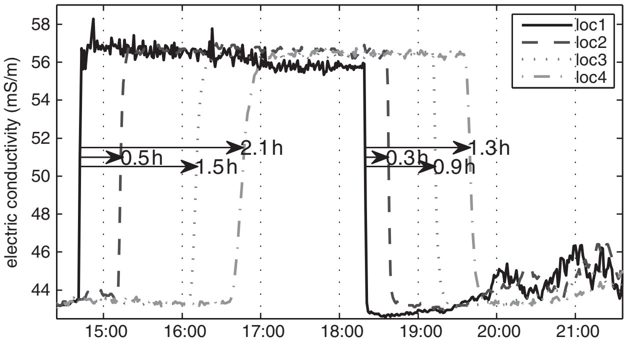

Figure 2 | EC at the measurement locations (Tuesday 24 July 2007) and the travel times between locations 1 and 2, 3, and 4.

This variation is considered in both the measurements and the hydraulic model.

Hydraulic model and demand allocation

Wallingford's InfoWorks® was used as a hydraulic network model solver. Basically, two models were constructed that are distinguished by demand allocation. $\mathrm{Model_{TD}}$ is the model with the top-down approach of demand allocation; $\mathrm{Model_{BU}}$ is the model with the bottom-up approach of demand allocation.

Each of the 144 homes was defined as a customer point (the InfoWorks entity that comprises water demand) and these customer points were connected to nodes on the mains. The hydraulic model $\mathrm{Model_{TD}}$ has consolidated demand nodes at pipe ends and junctions (the open circles in Figure 1). The hydraulic model $\mathrm{Model_{BU}}$ has unique individual demand nodes for all homes; the demand nodes are located at the stop taps of the homes (depicted by the connecting lines between demand nodes and distribution mains in Figure 1).

The measurement locations 1, 2, 3 and 4 were assigned a continuous extraction of 40, 40, 40 and 400 L/h respectively. The demand allocation at the customer points was done in two ways:

- In $\mathrm{Model_{TD}}$ an identical DMP was allocated to all customer points with a correction factor to account for the average demand per day. The utilized DMP are (a) $\mathrm{DMP_{PS}}$, (b) $\mathrm{DMP_{REF}}$ and (c) $\mathrm{DMP_{SIM}}$ (Table 1). The average demand of each customer point was assigned based on the water meter reading of 2004.
- In $\mathrm{Model_{BU}}$ a unique stochastic water demand pattern was assigned to each individual customer point. This is described in more detail in the next section.

In accordance with the measurement period, the branched system was modelled with weekday patterns; the looped system was modelled with weekend patterns.

The water demand patterns of the individual homes were generated on a 1 s time base. The generated water demand patterns were time averaged over different time scales before assigning them to the individual customer points in the hydraulic model. The hydraulic model was run with a hydraulic time step equal to the pattern time step, which was set to 1 min, 15 min and 1 h in different computer runs.

Table 1 | Overview of Demand Multiplier Patterns in $\mathrm{Model_{TD}}$

	$\mathrm{DMP_{PS}}$	$\mathrm{DMP_{REF}}$	$\mathrm{DMP_{SIM}}$
Name	DMP of pumping station	Reference DMP of test area	Simulated DMP of test area
Origin	Supply area Vlietregio	Test area	$\mathrm{Model_{BU}}$ of test area with SIMDEUM demand patterns
# homes	16,000	144	144
Time	21–30 July 2007	24–29 July 2007	n.a.
# weekdays	6	3	10
# weekend days	4	2	10
Original time step	5 min, average flows	1 min, instantaneous flows	1 s
Remark		Continuous flows for monitoring systems were subtracted	

The model was not calibrated on pressure; the selected network has well known pipe lengths, diameters and wall roughnesses and a fixed head at the entry point.

Water demand pattern generation

The end-use model SIMDEUM (Blokker *et al.* 2010b) was used as a water demand pattern generator. SIMDEUM input consists of information on the number and age of residents, the residents' sleep–wake rhythm and possession of, and behaviour with respect to, water-using appliances. Generic Dutch data were used for the water-using appliances and time use (Blokker *et al.* 2010b). Specific census data of the town Benthuizen were used; this information is available in the Netherlands by postal code area (CBS; Table 2).

For each of the 144 homes, 20 unique water demand patterns on a time scale of 1 s were generated with SIMDEUM (Blokker *et al.* 2010b); 10 patterns for weekdays and 10 patterns for weekend days. The patterns were then temporally aggregated (over 1 min, 15 min and 1 h) and divided by the average daily demand as obtained from the water meter readings. This led to 10 DMPs for each home. These DMPs do not necessarily have an average of 1 because they were divided by the average daily demand and not the average of the specific simulated demand.

A table was constructed which cross references customer point identification, average daily demand (L/day) and the demand category identification number (ID). This table was imported into the InfoWorks model. One demand category ID is linked to each of the 144 homes, and remains the same for all 10 different patterns. The demand multiplier patterns were exported into a text file of a specific InfoWorks format; a so called ".ddg" file. For each "demand category", information on the demand category ID, the number of multipliers, the time step and multiplier per time step are written to the file. This led to 20 computer generated .ddg files containing 144 DMPs per time scale. For the three different time scales (1 min, 15 min and 1 h) this means that 60 .ddg files were created. These .ddg files were imported for the different model scenarios.

Table 2 | Specific Benthuizen input data into SIMDEUM; data of postal code 2731 (1230 homes) in 2006 (CBS)

Resident type		Value
Households	One-person households	22%
	Households without children	30%
	Households with children	48%
	Average household size	2.8
Age distribution	0 to 15 years old	23%
	15 to 25 years old	14%
	25 to 45 years old	26%
	45 to 65 years old	27%
	65 years and older	11%

Water quality model

The hydraulic model was run with the water quality option enabled. This allowed for the determination of the residence time, which InfoWorks calls "water age". To determine the residence time the simulation run was set to 48 hours, where the diurnal demand patterns were repeated at hour 24 to 48. Because the residence times in this network do not exceed 24 hours, the diurnal patterns of the second day were not altered. One model run took less than one minute.

Sensitivity analysis and model validation

Twelve different scenarios were modelled (Table 3), which are related to network layout (branched and looped), different hydraulic time steps (1 min, 15 min and 1 h) and different demands. Scenarios 1, 3 and 5 were analysed to determine the influence of temporal scale. Scenarios 7, 9 and 11, and 8, 10 and 12 respectively, were studied to determine the influence of DMP. Scenarios 3 and 9, and 4 and 10 respectively, were examined to determine the influence of the top-down approach and the bottom-up approach of demand allocation.

For each scenario, the $\mathrm{Model}_{\mathrm{TD}}$ was run once and the system flow and residence time at three locations were determined. The $\mathrm{Model}_{\mathrm{BU}}$ was run 10 times with 10 different sets of stochastic water demand patterns. The resulting system flow (and corresponding $\mathrm{DMP}_{\mathrm{SIM}}$) is the averaged pattern of the 10 resulting patterns; the resulting residence time at the three locations is determined by the average and the 95% confidence interval of the 10 simulations. This 95% confidence interval is due to variation, not to uncertainty, and is

Table 3 | Model scenarios

Scenario	Hydraulic model	Layout	DMP	Specifics	Time step	# runs
1	$Model_{BU}$	branched	N.A.	weekday	1 min	10
2	$Model_{BU}$	looped	N.A.	weekend	1 min	10
3	$Model_{BU}$	branched	N.A.	weekday	15 min	10
4	$Model_{BU}$	looped	N.A.	weekend	15 min	10
5	$Model_{BU}$	branched	N.A.	weekday	1 h	10
6	$Model_{BU}$	looped	N.A.	weekend	1 h	10
7	$Model_{TD}$	branched	DMP_{REF}	weekday	15 min	1
8	$Model_{TD}$	looped	DMP_{REF}	weekend	15 min	1
9	$Model_{TD}$	branched	DMP_{PS}	weekday	15 min	1
10	$Model_{TD}$	looped	DMP_{PS}	weekend	15 min	1
11	$Model_{TD}$	branched	DMP_{SIM}	weekday	15 min	1
12	$Model_{TD}$	looped	DMP_{SIM}	weekend	15 min	1

determined by the average $+/-$ two times the standard deviation.

It was tested if 10 runs is enough to get a good view of mean and standard deviation of the residence times over the day. The difference between $(\mu+2\sigma)$ of the residence time after N – 1 simulations and $(\mu+2\sigma)$ after N simulations reveals how large the effect of an extra simulation run is. To calculate the effect for $N=10$, $(\mu+2\sigma)_{10}$ is compared to $(\mu+2\sigma)_9$. Because the 10 data points (i.e. calculated residence times) are the results of a Monte Carlo simulation, the order of the 10 data points is random. To account for this effect, the ten different subsets consisting of nine data points are considered as possible results for $N=9$. At each time step, the maximum difference (MD) between $(\mu+2\sigma)_{10}$ and $(\mu+2\sigma)_9$ as a percentage of μ_{10} is calculated:

$$MD = \frac{MAX(\mu+2\sigma)_9 - (\mu+2\sigma)_{10}}{\mu_{10}} \tag{1}$$

An MD of less than 5% is considered to be small enough to assume that 10 simulation runs suffice.

Additionally, a Kolmogorov–Smirnov (KS) test was performed on the resulting residence time per time step on each of the three locations to verify that the data are normally distributed; the mean and variance of the normal distribution were estimated from the data. If the data are normally distributed, the mean and standard deviation can be used to demonstrate the results of several simulation runs.

For the model validation, the modelled DMP_{SIM} and the measured DMP_{PS} and DMP_{REF} were compared. The diurnal pattern allows a visual assessment of how well the models resemble reality. To not leave it at a visual assessment, the resemblance of the DMPs (with, by definition, an average of 1) is quantified with the help of the auto- and cross-correlation of the DMPs. The cross-correlation between the DMPs quantifies how well the modelled DMPs fit the measured DMP_{REF}. The auto-correlation of the DMPs shows whether the DMPs have similar temporal variability.

The measured residence time at three locations and different times on the day was compared to the modelled residence time in the two network modes. The difference between (the average of) the model and the measurement is expressed by the Mean Error (ME), Root Mean Square Error (RMSE), and coefficient of determination R^2. The absolute values of ME and RMSE are expressed in hours; the relative values are percentages of the measured residence times. Also, the percentage of the model values that differ less than 10 minutes from the measured value is calculated. For the $Model_{TD}$, this percentage is calculated for the average modelled values. For the $Model_{BU}$, this percentage is calculated for the average modelled values and for the 95% confidence interval of the 10 different runs.

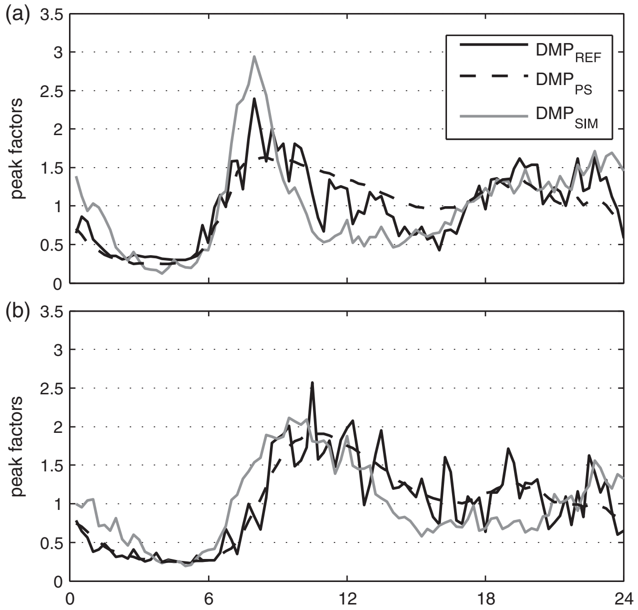

Figure 3 | Measured and simulated normalised DMP at 15 minutes time step on a) weekdays and b) weekend days.

RESULTS AND DISCUSSION

Demand multiplier pattern

Figure 3 shows the diurnal patterns. On weekdays, DMP_{PS} does not show a very distinct morning peak while DMP_{REF} and DMP_{SIM} do. The start of low night use for DMP_{SIM} is later than for DMP_{REF} and DMP_{PS}. For weekend days, all DMP are similar. DMP_{REF} is more spiky, because it is only based on 2 to 3 days of measurements of instantaneous flows.

A further analysis of the auto- and cross-correlation (not illustrated here) showed that for weekdays, the auto-correlation of DMP_{REF} is represented slightly better by the auto-correlation of DMP_{SIM} than by the auto-correlation of DMP_{PS}. This means that the variability of the flow into the network is predicted better by DMP_{SIM} than by DMP_{PS}. For weekend days, DMP_{SIM} has a much better agreement to DMP_{REF} than DMP_{PS} does. The cross-correlation between DMP_{PS} and DMP_{REF} is slightly higher (84% on weekdays, 90% on weekend days) than the cross-correlation between DMP_{SIM} and DMP_{REF} (79% on weekdays, 75% on weekend days).

Residence time – sensitivity analysis

At each time step, the maximum difference (*MD*, Equation (1)) between $(\mu + 2\sigma)_{10}$ and $(\mu + 2\sigma)_9$ as a percentage of μ_{10} is calculated. Table 4 shows at how many time steps *MD* is smaller than 5%. It also shows the average of *MD* over all time steps. It shows that in the branched layout, 10 simulations runs lead to a stable result, i.e. a less than 5% difference of $\mu + 2\sigma$ of the residence time between the ninth and tenth simulation run. This conclusion cannot be drawn for the looped layout at location 4. Especially with a short time step of 1 minute, 10 simulation runs does not give a stable result yet and more simulation runs are required. A KS test was performed on the resulting residence time (10 data points) per time step (96 time steps at the hydraulic time scale of 15 min) on each location. The null hypothesis was that the data are normally distributed and the test was performed at the 5% significance level. At location 2 in the branched layout with a 15 min time step, the KS test showed that at 91 time steps (95%, Table 5) the null hypothesis could not be rejected. Therefore, it is assumed that in this case the resulting residence time at each time step is normally distributed and that ten runs are enough to get information on the mean and standard deviation. This assumption appears to be valid in all cases,

Table 4 | Relative maximum difference (*MD*, Equation (1)) between $\mu + 2\sigma$ of ninth and tenth simulation run per time step

		Time steps where $MD < 5\%$			Average *MD*		
Location	**Layout**	**1 min**	**15 min**	**1 h**	**1 min**	**15 min**	**1 h**
2	Branched	100%	100%	100%	1.9%	1.7%	1.0%
3	Branched	100%	100%	100%	1.4%	1.3%	0.9%
4	Branched	100%	100%	100%	1.2%	1.2%	0.9%
2	Looped	100%	100%	100%	1.9%	1.7%	1.2%
3	Looped	99%	100%	96%	2.0%	1.9%	1.5%
4	Looped	83%	91%	96%	3.9%	3.6%	2.6%

Table 5 | Percentage of modelled residence time results per time step that is normally distributed according to Kolmogorov–Smirnov test on 10 data points per time step

		Time scale		
Location	**Layout**	**1 min**	**15 min**	**1 h**
2	Branched	95%	95%	100%
3	Branched	95%	95%	100%
4	Branched	96%	98%	88%
2	Looped	95%	92%	96%
3	Looped	92%	93%	96%
4	Looped	82%	76%	50%

except for location 4 in the looped layout, where the normal distribution can be confirmed for less than 90% of the time. In this case ten runs of the $\mathrm{Model_{BU}}$ may not be enough. For this case study, the number of simulation runs was limited to ten.

Figures 4–6 show the modelled and measured residence time over the day for the different scenarios; Table 6 summarizes the statistics. Depending on the network layout and the measurement location, the maximum residence time is reached between 5:00 and 9:00 a.m., which is related to the low night use. The fast decrease in residence time after the maximum is related to the peak in demand in the morning. The 95% confidence interval of the residence time in the

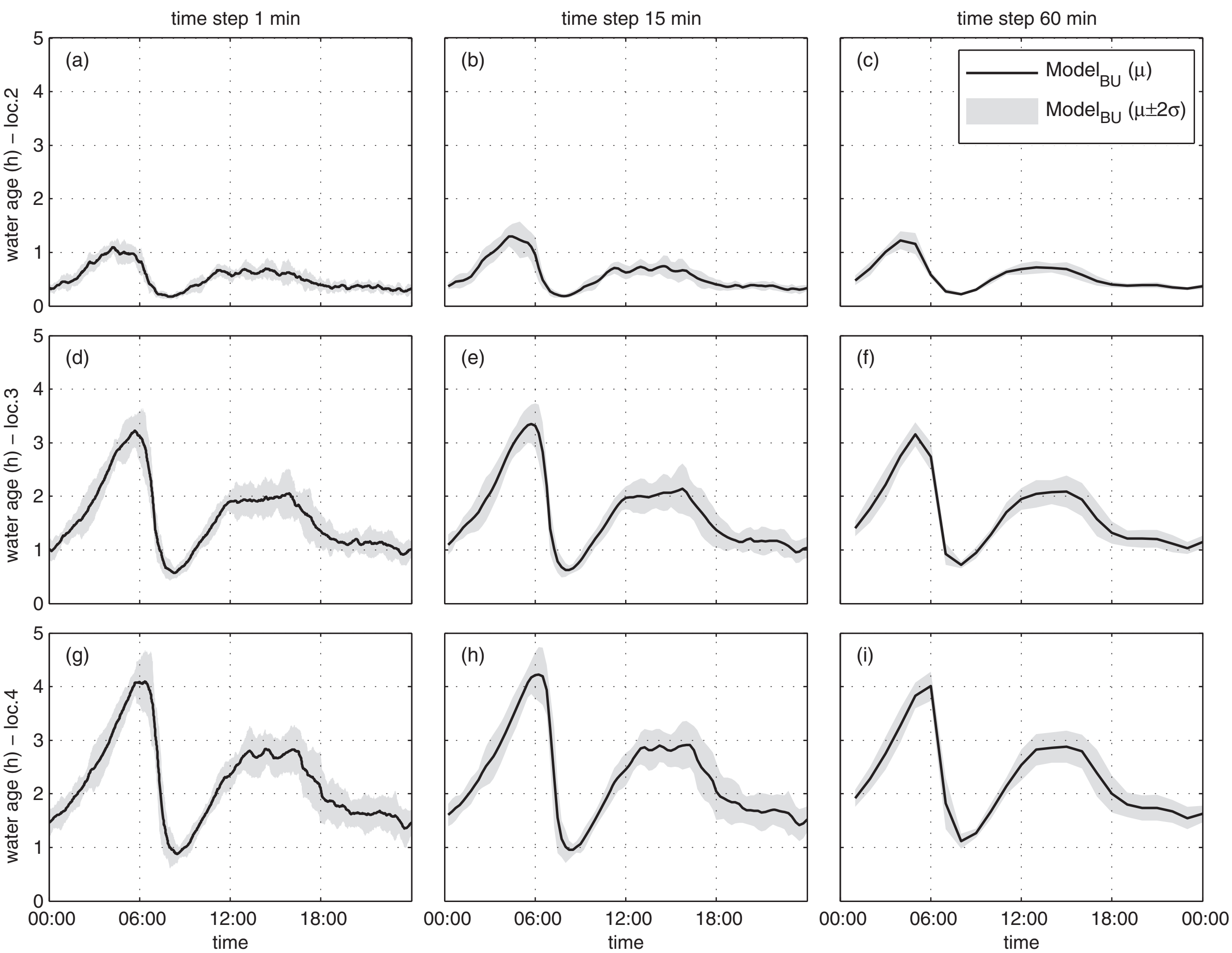

Figure 4 | Modelled residence time with $\mathrm{Model_{BU}}$ in branched network layout (scenarios 1, 3 and 5) on locations 2 (a-c), 3 (d-f) and 4 (g-i) with a time scale of 1 min (a, d, g), 15 min (b, e, h) and 1 h (c, f, i).

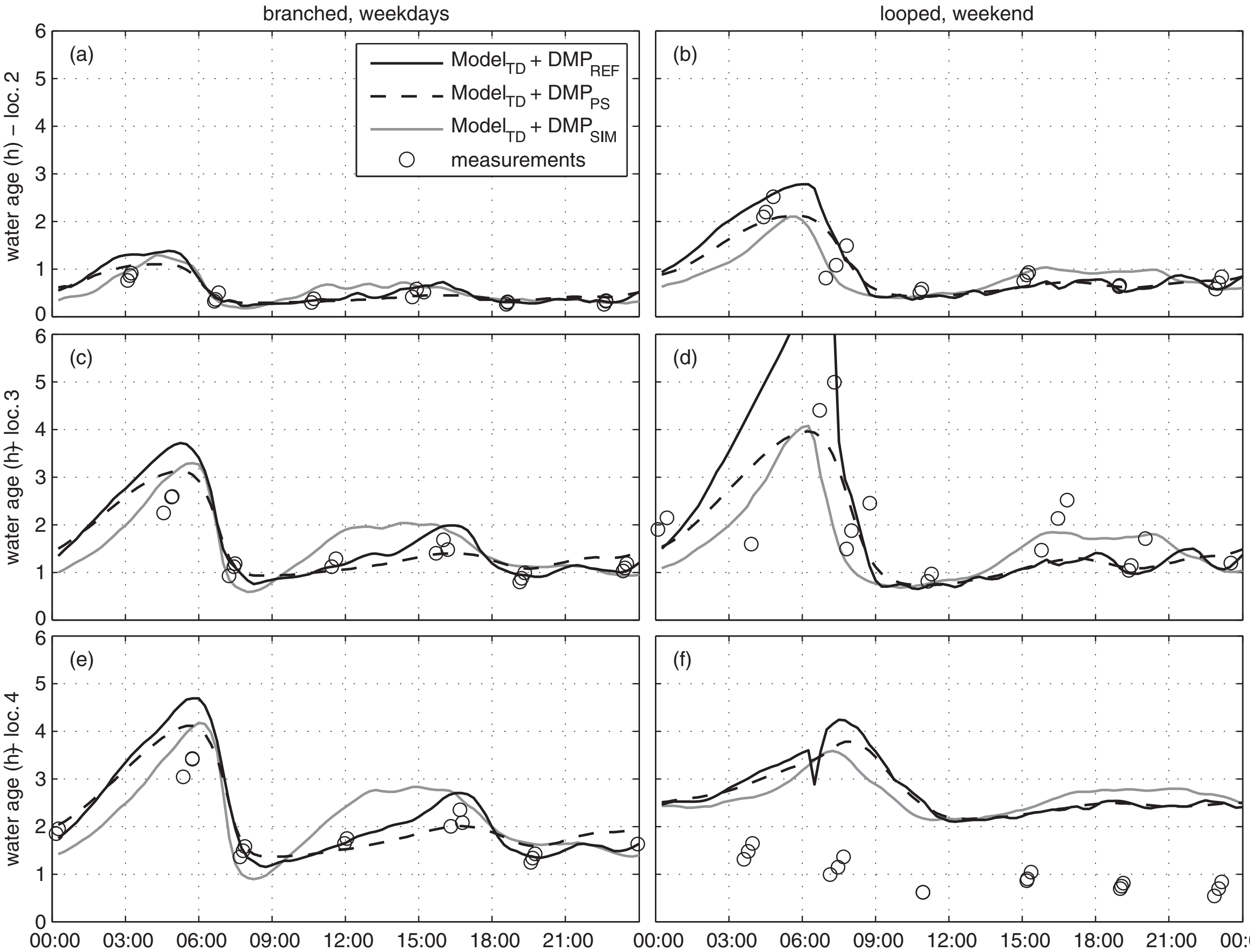

Figure 5 | Measured and modelled residence time with Model$_{TD}$ and different DMP in branched (a, c, e; scenarios 7, 9 and 11) and looped (b, d, f; scenarios 8, 10 and 12) network layout on locations 2 (a-b), 3 (c-d) and4 (e-f) with a time scale of 15 min.

Model$_{BU}$ is the largest for location 4 in the looped network layout; this is due to local conditions and flow direction reversals (Figure 6(f)).

The effect of the model's temporal scale was determined by comparing the resulting residence time from the Model$_{BU}$ with hydraulic time steps of 1 min, 15 min and 1 h (Figure 4). The difference in results is due to time averaging only. For determining residence time in this particular case, a 15 minute time scale is accurate enough. A shorter time step (1 minute) does not lead to a different 95% confidence interval of residence times. A longer time step (1 h) leads to too much time averaging; the minimum residence time is higher and the maximum residence time is lower than with a time step of 15 minutes. Therefore, the 15 minute time scale was used in the remaining analysis of Figures 5 and 6.

The effect of the model's spatial scale can be determined by comparing the resulting average residence time from the Model$_{BU}$ with a hydraulic time steps of 15 min (Figure 4) and the resulting residence time from the Model$_{TD}$ + DMP$_{SIM}$ (Figure 5). These two models vary in spatial correlation of the demands and in the location of the demand nodes, i.e. the Model$_{BU}$ has its demand nodes distributed along the pipes; the Model$_{TD}$ has its demand nodes and the ends of the pipes. In the branched layout (compare average of Figures 4(b), (e) and (h) and black solid line in Figure 5(a), (c) and (e)) there is no difference between the average of the Model$_{BU}$ and the Model$_{TD}$. In the looped layout (compare average of Model$_{BU}$ in Figures 6(b), (d) and (f) and black solid line in Figures 5(b), (d) and (f)) there is a small difference at locations 3 and 4: the Model$_{TD}$ results in a smoother line and shows a lower

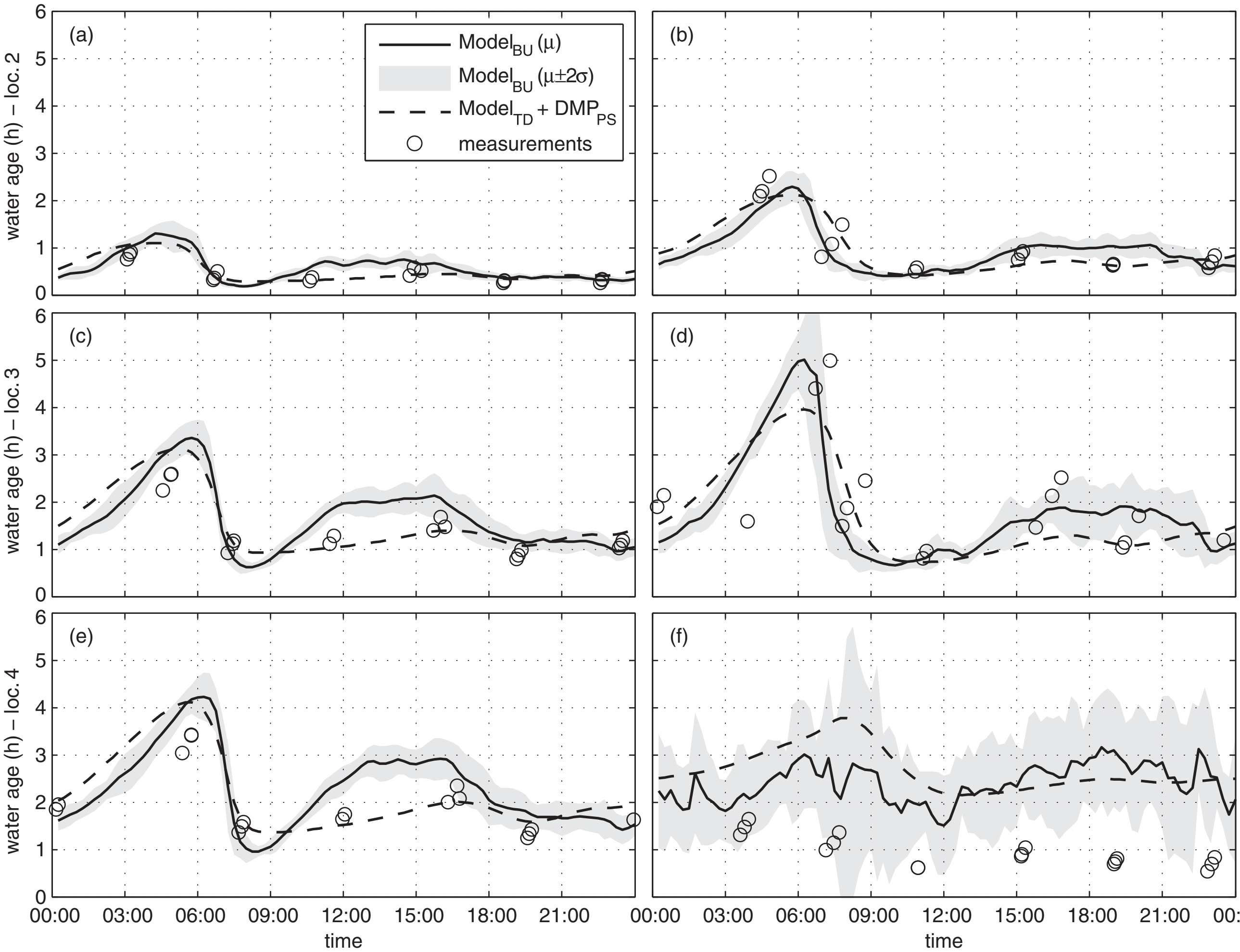

Figure 6 | Measured and modelled residence time (Model$_{TD}$ + DMP$_{PS}$ and Model$_{BU}$) in branched (a, c,e; scenarios 3 and 9) and looped (b, d, f; scenarios 4 and 10) network layout on locations 2 (a–b), 3 (c–d) and 4 (e–f) with a time scale of 15 min.

residence time in the morning at location 3 and a higher residence time in the morning at location 4. Spatial scale therefore has a limited effect on the mean and 95% confidence interval of calculated residence times at the measurement locations. The effect of spatial scale is eminent at the periphery of the network, where only a few homes are connected.

The effect of the Model$_{TD}$'s DMP was determined by comparing the results from Model$_{TD}$ + DMP$_{REF}$, DMP$_{PS}$ and DMP$_{SIM}$ respectively (Figure 5). An effect of the different water demand patterns on residence time was expected from Figure 3. The effect of the DMP is apparent as the three different DMPs lead to different residence times. The measured residence time is most often predicted best by the Model$_{TD}$ + DMP$_{PS}$, i.e. for location 3 and 4 in the branched layout and for location 2 and 3 in the looped layout (Table 6). Sometimes the Model$_{TD}$ + DMP$_{SIM}$ works best, i.e. for location 2 in the branched layout and for location 4 in the looped layout (Table 6). The DMP that was actually measured does not lead to the best results when applied in the Model$_{TD}$ + DMP$_{REF}$.

Residence time – model validation

The difference between the conventional approach (Model$_{TD}$ + DMP$_{PS}$) and the new approach (Model$_{BU}$) of demand modelling is shown in Figure 6 and Table 6. The two models predict the residence time with comparable ME and RMSE

Journal of Hydroinformatics | 13.4 | 2011

Table 6 | Difference between measured and modelled residence time of Figures 5 and 6

			Model$_{BU}$ (SIMDEUM)			Model$_{TD}$ (DMP$_{PS}$)			Model$_{TD}$ (DMP$_{REF}$)		
			loc. 2	loc. 3	loc. 4	loc. 2	loc. 3	loc. 4	loc. 2	loc. 3	loc. 4
Branched	Scenario			3			9			7	
	Sample size		17	17	17	17	17	17	17	17	17
	ME	Absolute (h)	0.05	0.22	0.25	0.05	0.15	0.21	0.10	0.26	0.33
		Relative (%)	10.9	16.0	12.9	11.9	11.0	10.5	21.83	19.02	16.59
	RMSE	Absolute (h)	0.11	0.40	0.53	0.13	0.33	0.42	0.21	0.53	0.65
		Relative (%)	24.2	29.0	27.0	28.0	23.9	21.1	45.87	38.33	32.79
	R^2 (%)		72	50	42	63	66	64	N.A.	12	13
	Within 10 min deviation(%)	Compared to mean	88	18	6	88	35	41	76	59	59
		Compared to 95% c.i.	100	76	47	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Looped	Scenario			4			10			8	
	Sample size		17	17	17	17	17	17	17	17	17
	ME	Absolute (h)	−0.07	−0.37	1.46	−0.02	−0.28	1.76	0.04	0.07	1.89
		Relative (%)	−7.0	−18.5	151.8	−2.2	−13.8	182.7	4.21	3.47	196.19
	RMSE	Absolute (h)	0.37	1.01	1.64	0.35	0.86	1.86	0.41	1.19	2.03
		Relative (%)	34.9	50.4	170.2	33.1	43.0	193.2	39.14	59.57	210.88
	R^2 (%)		66	22	N.A.	69	44	N.A.	57	N.A.	N.A.
	Within 10 min deviation(%)	Compared to mean	29	23	0	59	17	0	47	18	0
		Compared to 95% c.i.	88	47	35	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

and R^2. Both models predict the residence time in the branched layout with an ME and RMSE of less than 30%. Both models perform poorly for the looped layout (RMSE > 30%) and especially for location 4 (Figure 6(f)) where they significantly overestimate the residence time. The R^2 is never above 0.72 which suggests that neither model performs very well. The values of ME, RMSE and R^2 mainly have a meaning in comparing the two models. The absolute values are not easy to interpret because the measured residence times at specific moments on the day are different for consecutive days and therefore show variance. Presumably, the residence time is normally distributed. However, the range of measured data is compared to the model averages. The 95% confidence interval of the Model_{BU} presents many more data points within 10 min from the measured residence time than the average of both the Model_{BU} and the Model_{TD}. This shows the added value of the Model_{BU}.

The bottom-up modelling approach is probabilistic in nature and offers a new perspective for assessing water quality in the drinking water distribution system. The test case showed that, especially on location 4 in the looped network layout, the variability of residence time between days is expected to be very high with the maximum residence time 2.5 times as large as the average residence time, or the minimum residence time 2.5 times as small as the average residence time. This suggests that it would be very difficult to use the tracer measurements at this location for calibration purposes. Averaging of water demand and residence time prediction may lead to misinterpretation of water quality data. The model's sensitivity is also related to the variability of residence times, not so much to the average residence times. A hydraulic model with a demand pattern time step of 15 minutes and limited spatial aggregation of demand patterns leads to good results.

The stochastic approach of hydraulic modelling gives insight into the variability of residence times as an added feature beyond the conventional way of modelling. The conventional Model_{TD} has a higher auto- and cross-correlation of flows than the actual flows in the network. This results in the Model_{TD} underestimating the flow direction reversals, stagnant flows and thus maximum residence times. Because Machell *et al.* (2009) have argued that the maximum residence time is much more important than the average residence time, the Model_{BU} has benefits in determining residence time.

Practical application

The study of a bottom-up approach of demand allocation in this small test area showed that this approach leads to a good understanding of average and variance of residence times. This comes at a cost. Compared to the conventional top-down approach of demand allocation, the new model approach leads to larger hydraulic models with more nodes, more pipes and more numerous demand patterns. In the test model, an extra node, pipe and demand pattern were added for each individual home. Using a smaller time step means that simulations take longer; a time step of 1 minute instead of 1 hour, approximately leads to a 60 times longer simulation. In the test area this means that simulations were still very quick. To determine the variance of the residence time, multiple simulation runs are required. This also means a longer total simulation time. With the still increasing computer capacity, the problem of more demanding hydraulic models will diminish. Another practical issue is that the analysis of the set of results cannot be done in a hydraulic network solver, but needs to be done elsewhere.

A sensitivity analysis showed that for the specific purpose of determining residence times for the selected locations, there is no need to run the hydraulic model at as small a time step as one second; a 15 minute time step suffices. Also, some spatial aggregation is permitted; it is not required that each home has its own demand node in the hydraulic model. The optimal number of simulations can iteratively be determined by comparing the results of the Nth and (N-1)th simulations and test that the results are normally distributed. In this study the optimal number of simulations was not determined; instead it was tested if ten simulations were sufficient. For most locations in this case study, a stochastic modelling approach of ten different simulations is enough to get a good feel for the 95% confidence interval. These findings can help to limit the model increase.

In this study some simplifications were introduced. Firstly, an additional demand was extracted in order to keep dispersion low. Secondly, in the branched layout no mixing occurred. In the looped layout some mixing occurred, but probably full mixing can be assumed because there only T-junctions in this network (Ho *et al.* 2006). And thirdly, there are no significant leaks in this network. These simplifications allowed for the sensitivity analysis. A similar modelling approach was done for a network with 1,000 homes and without the first two of the mentioned simplifications (Blokker *et al.* 2010a). The construction of an "all pipes" network model meant an effort. This effort was not specifically done for the study. Most Dutch water companies are migrating to using all pipes network models as they can automatically be generated from the GIS systems. Filling the model with the appropriate water demand patterns and running the simulations was automated and took little effort. The study in this more practical network showed that a bottom-up approach of demand allocation in real networks is feasible without the need for calibration on water demands, and leads to a realistic insight into average and variance of residence times. It therefore has an advantage over the conventional top-down approach.

In a network with leakage, the bottom-up approach of demand allocation can also be used. Leaks must then be added as separate demands. Giustolisi *et al.* (2008) show how this can be done. In this case, leakages should be modeled as pressure-dependent demands. The leakage patterns depend on the hydraulic status of the network and on pipe-specific leakage parameters.

CONCLUSION

A bottom-up approach of demand allocation (i.e. water demand patterns are modelled per individual home and subsequently the individual water demand patterns are summed to obtain the water demand patterns at demand nodes) leads to a total flow that is predicted at least as well as the flow from the commonly used top-down approach model. Furthermore, the bottom-up approach leads to promising results in predicting residence time in a small distribution network. The individual demand patterns are obtained from the end-use model SIMDEUM without the need for any flow measurements, nor for calibration of demands. Some specific census data was collected and used as input to SIMDEUM; most of the input data can be re-used from earlier studies.

The water demand patterns are constructed per individual home and on a per second basis. For the purpose of residence time prediction at locations with a number of households behind it, it is acceptable to use time-averaging, and use a hydraulic time step of 15 minutes and "spatial-averaging" by summing numerous individual water demand patterns into one demand node. For most locations in this case study, ten simulation runs is enough to get an understanding of the expected mean and variance of residence times over the day. A comparison between the results of N simulations and $N-1$ simulations and a Kolmogorov–Smirnov test can be used to verify how many simulations are required.

A stochastic approach in demand and water quality modelling results in more insight into the variability of residence times. A detailed demand allocation with stochastic demand patterns will improve the water quality modelling, especially in the periphery of the drinking water distribution system.

SYMBOLS AND ABBREVIATIONS

DMP	Demand multiplier patterns; see Table 1 for subscripts
DWDS	Drinking water distribution system
EC	Electrical conductivity
KS test	Kolmogorov–Smirnov test
MD	maximum difference
ME	mean error
$Model_{BU}$	Model with bottom-up approach of demand allocation
$Model_{TD}$	Model with top-down approach of demand allocation
RMSE	root mean square error
R^2	coefficient of determination
μ	mean
σ	standard deviation

REFERENCES

Beuken, R. H. S., Lavooij, C. S. W., Bosch, A. & Schaap, P. G. 2006 Low leakage in the Netherlands confirmed. *Proc. 8th Ann. Int. Symp. Water Distribution System Analysis*, Cincinnati, OH.

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. *Drink. Water Eng. Sci.* **1**(1), 27–38.

Blokker, E. J. M., Vreeburg, J. H. G., Beverloo, H., Klein Arfman, M. & van Dijk, J. C. 2010a A bottom-up approach of stochastic demand allocation in water quality modelling. *Drink. Water Eng. Sci.* **3**(1), 43–51.

Blokker, E. J. M., Vreeburg, J. H. G. & van Dijk, J. C. 2010b Simulating residential water demand with a stochastic end-use model. *J. Water Resour. Plann. Manag.* **136**(1), 19–26.

CBS http://statline.cbs.nl, December 2007

EPA 2002 *Effects of water age on distribution system water quality*, prepared by AWWA with assistance from Economic and Engineering Services, Inc. EPA White paper.

Giustolisi, O., Savic, D. & Kapelan, Z. 2008 Pressure-driven demand and leakage simulation for water distribution networks. *J. Hydraul. Engng* **134**(5), 626–635.

Ho, C. K., Orear, L., Wright, J. L. & McKenna, S. A. 2006 Contaminant mixing at pipe joints: Comparison between laboratory flow experiments and computational fluid dynamics models. *Proc. 8th Ann. Int. Symp. Water Distribution System Analysis*, Cincinnati, OH.

Jonkergouw, P. M. R., Khu, S.-T., Kapelan, Z. S. & Savic, D. A. 2008 Water quality model calibration under unknown demands. *J. Water Resour. Plann. Manag.* **134**(4), 326–336.

Kapelan, Z. 2002 Calibration of water distribution system hydraulic models. *Ph.D. Thesis.* University of Exeter, p. 334.

Li, Z. 2006 Network water quality modeling with stochastic water demands and mass dispersion. *Ph.D. Thesis*, University of Cincinnati, p. 165.

Machell, J., Boxall, J., Saul, A. & Bramley, D. 2009 Improved representation of water age in distribution networks to inform water quality. *J. Water Resour. Plann. Manag.* **135**(5), 382–391.

Pasha, M. F. K. & Lansey, K. 2010 Effect of parameter uncertainty on water quality predictions in distribution systems – case study. *J. Hydroinf.* **12**(1), 1–21.

Powell, J., Clement, J., Brandt, M. R. C., Holt, D., Grayman, W. & LeChevallier, M. 2004 *Predictive Models for Water Quality in Distribution Systems*. AWWARF Report 91023F, Denver, CO.

Skipworth, P. J., Machell, J. & Saul, A. J. 2002 Empirical travel time estimation in a distribution network. *Water Maritime Engng* **154**(1), 41–49.

Slaats, P. G. G., Rosenthal, L.P.M., Siegers, W. G., van den Boomen, M., Beuken, R. H. S. & Vreeburg, J. H. G. 2003 *Processes involved in the generation of discolored water*, AWWARF Report 90966F, Denver, CO.

Vreeburg, J.H.G. 2007 Discolouration in drinking water systems: a particular approach. *Ph.D. Thesis*. Kiwa Water Research, The Netherlands.

Vreeburg, J. H. G. & Boxall, J. B. 2007 Discolouration in potable water distribution systems: A review. *Water Res.* **41**(3), 519–529.

First received 6 May 2010; accepted in revised form 14 October 2010. Available online 18 January 2011