Original Research

Poisson rectangular pulse (PRP) model establishment based on uncertainty analysis of urban residential water consumption patterns

Jiaxin Zhang [a, b], Dragan Savic [c, d, e], Qiang Xu [a, *], Kuo Liu [f], Zhimin Qiang [a, b]

[a] Key Laboratory of Drinking Water Science and Technology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] KWR Water Research Institute, Nieuwegein, 3430, BB, the Netherlands
[d] Centre for Water Systems, University of Exeter, Exeter, EX44QF, United Kingdom
[e] Faculty of Civil Engineering, University of Belgrade, 11000, Belgrade, Serbia
[f] Beijing Waterworks Group Co. Ltd, Beijing, 100031, China

ARTICLE INFO

ABSTRACT

The commonly used Poisson rectangular pulse (PRP) model, employed for simulating high-resolution residential water consumption patterns (RWCPs), relies on calibration via medium-resolution RWCPs obtained from practical measurements. This introduces inevitable uncertainty stemming from the measured RWCPs, which consequently impacts the precision of model simulations. Here we enhance the accuracy of the PRP model by addressing the uncertainty of RWCPs. We established a critical sampling size of 2000 household water consumption patterns (HWCPs) with a data logging interval (DLI) of 15 min to attain dependable RWCPs. Through Genetic Algorithm calibration, the optimal values of the PRP model's parameters were determined: pulse frequency $\lambda = 91$ d^{-1}, mean of pulse intensity $E(I) = 0.346$ m^3 h^{-1}, standard deviation of pulse intensity $STD(I) = 0.292$ m^3 h^{-1}, mean of pulse duration $E(D) = 40$ s, and standard deviation of pulse duration $STD(D) = 55$ s. Furthermore, validation was conducted at both HWCP and RWCP levels. We recommend a sampling size of ≥ 2000 HWCPs and a DLI of ≤ 30 min for PRP model calibration to balance simulation precision and practical implementation. This study significantly advances the theoretical foundation and real-world application of the PRP model, enhancing its role in urban water supply system management.

1. Introduction

Residential water consumption patterns (RWCPs) play a crucial role in the efficient management of urban water supply systems [1–3]. The resolution of RWCP data significantly affects their applicability [4–6]. Low-resolution RWCPs, characterized by a data logging interval (DLI) ranging from days to months, focus primarily on univariate statistics to analyze the socioeconomic effects of water consumption. In contrast, medium-resolution RWCPs, with a DLI spanning minutes to hours, reflect diurnal variations in water consumption and are commonly employed to develop hydraulic models of water distribution systems [7,8] and detect water leakages [9,10]. On the other hand, high-resolution RWCPs, featuring a

DLI at the level of seconds, permit fine flow-signature analysis [6] and are increasingly employed to develop water quality models [11]. While lower-resolution RWCPs can generally be aggregated by higher-resolution ones, but not vice versa, making high-resolution RWCPs indispensable for comprehensive water consumption analysis. In addition, the rapid development of smart metering technology has brought great convenience to water consumption measurements [12,13]. However, smart water meters have only been adopted by a subset of residential households so far, leading water authorities to often rely on medium-resolution RWCPs considering the costs for data acquisition, transmission, and storage [14,15]. Under this circumstance, simulation models have been developed to acquire high-resolution RWCPs more cost-effectively than direct measurements [16].

The Poisson rectangular pulse (PRP) model stands out as a prominent choice among simulation models, widely employed to

* Corresponding author.
E-mail address: qiangxu@rcees.ac.cn (Q. Xu).

represent random water consumption events as equivalent rectangular pulse [4,15]. Each pulse is expressed by three attributes: frequency (λ), intensity (I), and duration (D), and multiple pulses are generated by a time-dependent Poisson Arrival Process [4,15,17]. The pulse intensities are assumed to follow an exponential, log-normal, or Weibull distribution [15,18,19], while pulse durations follow an exponential, log-normal, or normal distribution [15,18,20,21]. Among these distributions, the log-normal distribution is most commonly adopted for pulse intensity and duration. On this basis, the PRP model can be expressed by five parameters: pulse frequency (λ), mean and standard deviation of pulse intensity ($E(I)$ and $STD(I)$), and mean and standard deviation of pulse duration ($E(D)$ and $STD(D)$), each holding explicit physical interpretation [22].

Prior to its practical implementation, a PRP model needs to be established, which involves calibrating and validating its parameters. The calibration of model parameters often entails minimizing the differences in certain variables (e.g., the flow rate in each time slot (slot width = DLI) (Q_i), daily water consumption volume (V)) between the measured and simulated RWCPs. Validation is performed using a goodness-of-fit test for cumulative frequency distributions (CFDs) of these characteristic variables [23,24]. In addition, the uncertainty of a measured RWCP depends on both sampling size and DLI. For example, a single-household RWCP presents discrete rectangular pulses with strong stochasticity [7,24], while a residential-community RWCP presents a bimodal curve with dominant periodicity [5,12,25]. Therefore, to improve the simulation accuracy of a PRP model, the RWCP uncertainty analysis is an inevitable prerequisite.

This study aimed to establish an accurate PRP model, intended for simulating high-resolution RWCPs, from practically measured medium-resolution RWCPs. The RWCP uncertainty was first analyzed by considering different sampling sizes and DLIs. To calibrate the PRP model, Genetic Algorithm (GA) was employed to minimize the sum of squared differences of Q_i between the measured and simulated RWCPs. Model validation was performed through a goodness-of-fit test for the CFDs of four characteristic variables: maximum flow rate (max (Q_i)), V, Q_i, and flow rate difference between adjacent time slots ($\Delta Q_i = Q_{i+1} - Q_i$) over a day. In brief, the study will substantially improve the theoretical underpinning and real-world application of the PRP model from the data perspective, thereby enhancing the management of urban water supply systems.

2. Materials and methods

2.1. PRP model establishment procedures

The PRP model was established through a series of sequential steps: (1) measuring the data pertinent to household water consumption patterns (HWCPs) with a pre-selected DLI (e.g., 15 min); (2) constructing the medium-resolution RWCPs through superimposing a certain number of HWCPs (defined as the sampling size) randomly selected for each RWCP; (3) analyzing the uncertainty of the constructed medium-resolution RWCPs; (4) calibrating the PRP model parameters; and (5) validating the PRP model simulation accuracy (Fig. 1). In this study, HWCPs represented the water consumption pattern of an individual household over a day, whereas RWCPs represented the water consumption pattern of an entire residential community over a day. Once established, the PRP model allowed for the simulation of high-resolution RWCPs, offering a valuable tool to better manage urban water supply systems.

2.2. Residential water consumption data collection

This study focused on measuring water consumption in two urban residential communities in Beijing, China. The selected communities consisted of 830 and 776 households, each equipped with smart water meters. For each household, water consumption data were measured continuously for five weekdays with a DLI of 15 min. Theoretically, a total of 8030 [i.e., (830 + 776) × 5] HWCPs could be obtained. After eliminating invalid patterns caused by malfunctions of water meters or transmission devices, 7650 useable HWCPs were finally acquired. These HWCPs were then randomly divided into a training set (4000 HWCPs) and a validation set (3650 HWCPs) for further analysis (Fig. 1).

2.3. RWCP uncertainty analysis

As shown in Fig. 1, a certain number of HWCPs (i.e., sampling size) were randomly selected from the training set and then superimposed into one RWCP. The process was repeated 100 times at a fixed sampling size to generate 100 RWCPs, and their fluctuation range for each time slot was determined as the RWCP uncertainty interval. The investigation encompassed the evaluation of 30 sampling scales, spanning from 50 to 3000 HWCPs.

To make the RWCPs comparable among different sampling scales, a normalized water consumption coefficient for the ith time slot of an RWCP (α_i) is defined as follows:

$$\alpha_i = \frac{Q_i}{\frac{1}{N}\sum_{i=1}^{N} Q_i} \tag{1}$$

where Q_i is the flow rate for the ith time slot of an RWCP (i.e., the sum of all flow rates of a certain number of HWCPs); and N is the number of time slots over a day (e.g., N = 96 at DLI = 15 min).

The RWCP uncertainty interval for the ith time slot and the jth sampling scale ($UI_{i,j}$) and the average uncertainty interval over a day
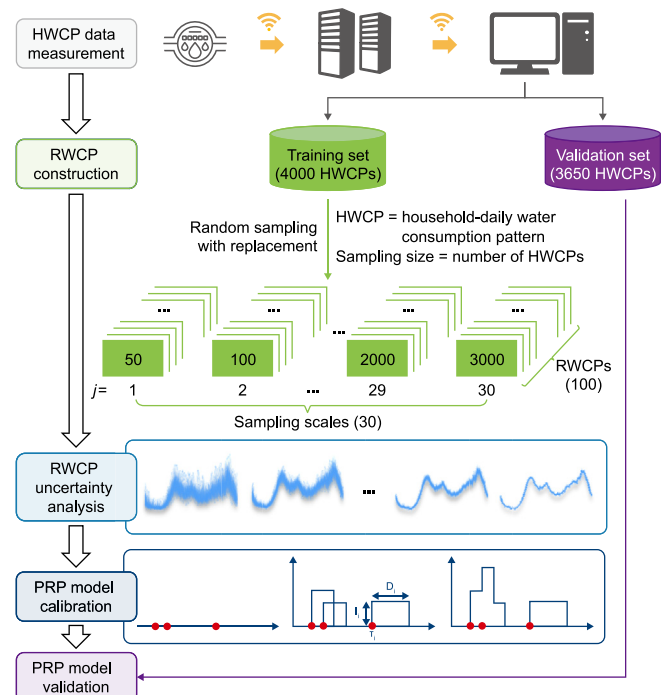

Fig. 1. Flow chart for PRP model establishment.

for the jth sampling scale ($\overline{UI}_j$) can be calculated respectively by equations (2) and (3) as follows:

$$UI_{i,j} = \max(\alpha_{i,j}) - \min(\alpha_{i,j}) \tag{2}$$

$$\overline{UI}_j = \frac{1}{N} \sum_{i=1}^{N} UI_{i,j} \tag{3}$$

where $i \in [1, N]$, $j \in [1, 30]$; and max $(\alpha_{i,j})$ and min $(\alpha_{i,j})$ are the maximum and minimum coefficients for the ith time slot and the jth sampling scale among the 100 RWCPs.

To reflect how fast the $\overline{UI}_j$ changes with the sampling size, a change rate (r) is defined as follows:

$$r_{j+1} = \frac{\overline{UI}_{j+1} - \overline{UI}_j}{NHWCP_{j+1} - NHWCP_j} \tag{4}$$

where $NHWCP_{j+1}$ and $NHWCP_j$ are the number of HWCPs for the $(j+1)$th and jth sampling scale, respectively.

2.4. PRP model description

In a PRP model, the water consumption events of a household over a day are represented equivalently by rectangular pulses generated by the Poisson arrival process. The probability P of having k pulses per day is calculated by equation (5) [26]:

$$P(k) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{5}$$

where k is a nonnegative integer; and λ is the pulse frequency.

The pulse can be simulated using a block function B as follows [26]:

$$B(I, D, \tau) = \begin{cases} I, T \in [\tau, \tau + D] \\ 0, T < \tau \text{ or } T > \tau + D \end{cases} \tag{6}$$

where T is any time point, and τ is the occurrence time point of a water consumption event.

In the block function B, both I and D are assumed to follow a log-normal distribution [15]:

$$f(I) = \frac{1}{\sqrt{2\pi}\sigma_I I} e^{-\frac{1}{2\sigma_I^2}[\ln(I) - \mu_I]^2} \tag{7}$$

$$f(D) = \frac{1}{\sqrt{2\pi}\sigma_D D} e^{-\frac{1}{2\sigma_D^2}[\ln(D) - \mu_D]^2} \tag{8}$$

where $f(I)$ and $f(D)$ are probability density functions of I and D, respectively; μ_I and σ_I are the mean and standard deviation of $\ln(I)$, respectively, which can be calculated from E(I) and STD(I) by equations (9) and (10); and μ_D and σ_D are the mean and standard deviation of $\ln(D)$, respectively, which can be calculated from E(D) and STD(D) by equations (11) and (12) [27]:

$$\mu_I = \ln[E(I)] - \frac{1}{2}\ln\left[1 + \frac{STD(I)^2}{E(I)^2}\right] \tag{9}$$

$$\sigma_I = \sqrt{\ln\left[1 + \frac{STD(I)^2}{E(I)^2}\right]} \tag{10}$$

$$\mu_D = \ln[E(D)] - \frac{1}{2}\ln\left[1 + \frac{STD(D)^2}{E(D)^2}\right] \tag{11}$$

$$\sigma_D = \sqrt{\ln\left[1 + \frac{STD(D)^2}{E(D)^2}\right]} \tag{12}$$

Equations (7) and (8) are used to generate random k sets of I and D (i.e., k pulses). For each pulse in the ith time slot, its occurrence probability is proportional to Q_i, while its occurrence time point is generated randomly. Then, the k pulses generated are superimposed to form an HWCP [19,28]. Consequently, at any time point, the flow rate can be calculated as the sum of the intensities of all pulses occurring.

2.5. PRP model establishment

The PRP model was established after calibration and validation with the medium-resolution RWCPs measured (Fig. 1). According to the DLI for water consumption data measurement in this study (i.e., 15 min), the initial PRP model (uncalibrated, Section 2.4) was used to simulate a desired number of HWCPs, and then RWCPs were formed by superimposing different sets of HWCPs (Section 2.3). For model calibration, an objective function (F), which sums the squared differences of the maximum, mean, and minimum flow rates between the measured and simulated RWCPs (each 100), was minimized to obtain the five parameters (λ, E(I), STD(I), E(D), and STD(D)) by using GA [29]:

$$F = \sum_{i=1}^{N} \left[\left(Q_i^{\max} - SQ_i^{\max}\right)^2 + \left(Q_i^{\mean} - SQ_i^{\mean}\right)^2 + \left(Q_i^{\min} - SQ_i^{\min}\right)^2 \right] \tag{13}$$

where $Q_i^{\max}$, $Q_i^{\mean}$, and $Q_i^{\min}$ are the maximum, mean, and minimum flow rates for the ith time slot of measured RWCPs, respectively; and $SQ_i^{\max}$, $SQ_i^{\mean}$, and $SQ_i^{\min}$ are the maximum, mean, and minimum flow rates for the ith time slot of simulated RWCPs, respectively.

Thereafter, the PRP model was validated at both HWCP and RWCP levels by four characteristic variables (max(Q_i), V, Q_i, and ΔQ_i). Specifically, a goodness-of-fit test was performed for the CFDs of these characteristic variables between the measured and simulated HWCPs (each 1000) or RWCPs (each 100) through comprehensive consideration of the mean error (ME), the root of mean square error (RMSE) and coefficient of determination (R^2). To compare the results at different sampling sizes, ME and RMSE were expressed as relative values to the average measured characteristic variables (Text S1).

3. Results

3.1. RWCP uncertainty analysis

Six out of all 30 sampling scales, in correspondence to the six representative sampling sizes of 50, 200, 500, 1000, 2000, and 3000 HWCPs, were selected to investigate the daily variations of α_i, as shown in Fig. 2. As mentioned above, the RWCP uncertainty interval (UI) was referred to the vertical fluctuation range of α_i, representing the vertical distance between the upper and lower envelop curves. Results indicate that the UI value exhibited greater values during morning and night peak hours but decreased at midnight. Moreover, as the sampling size increased, the UI
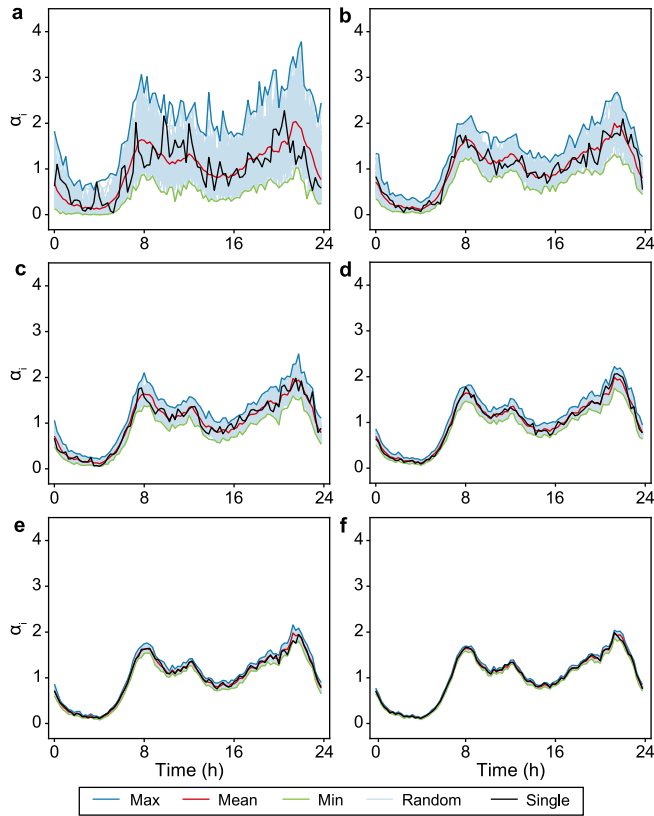
Fig. 2. Variations of α_i over a day at representative sampling sizes: **a**, 50; **b**, 200; **c**, 500; **d**, 1000; **e**, 2000; **f**, 3000 HWCPs.

Fig. 3. Variations of $\overline{UI}$ and r with sampling size.

gradually diminished. For example, the UI fluctuated largely from 0.53 (at 2:30) to 3.01 (at 22:00) at 50 HWCPs, but slightly from 0.03 (at 3:00) to 0.21 (at 22:00) at 3000 HWCPs. In addition, the "single" α_i curve (black) deviated drastically from the "mean" α_i curve (red) at 50 HWCPs, while the two curves nearly converged at 3000 HWCPs. These observations highlight that increasing the sampling size effectively reduced the RWCP uncertainty.

To investigate the relationship between sampling size and the rate of UI change, the variations of both $\overline{UI}$ and r are presented in Fig. 3. Results indicate that $\overline{UI}$ decreased monotonically from 1.60 to 0.10 as the sampling size increased from 50 to 3000 HWCPs. Meanwhile, the absolute value of r ($|r|$) decreased from 0.960% to 0.007% correspondingly. To minimize the negative impact of the measured RWCPs' uncertainty on the calibration of the PRP model to be established, the thresholds for a dependable RWCP were thus proposed as follows: $\overline{UI} < 0.2$ and $|r| < 0.01\%$, as a trade-off between model simulation accuracy and data acquisition cost. As depicted in Fig. 3, a critical sampling size of 2000 HWCPs and a DLI of 15 min were required to meet the above thresholds.

3.2. PRP model calibration

The procedures for PRP model calibration are described in Fig. S1. The objective function (F) at the critical sampling (or simulating) size (i.e., 2000 HWCPs) was minimized by using the GA. After our computational trials, the population size in each generation of the GA was set to ten individuals. Each individual contained five genes corresponding to the five model parameters (λ, E(I), STD(I), E(D), and STD(D)), and each gene was represented by a binary string. The selection, crossover, and mutation probabilities of the GA were set as 0.8, 0.7, and 0.01, respectively [30].
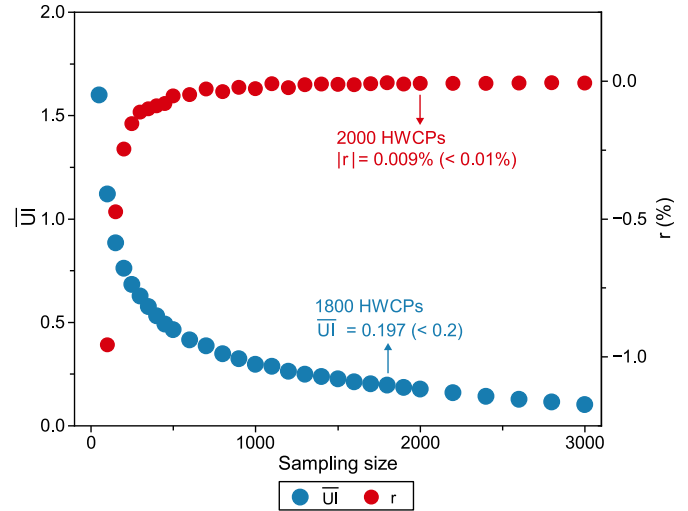
To ensure efficient convergence of the GA, the initial range of each parameter should be set appropriately. By referring to literature reported values (12 sets), the maximum and minimum values of each parameter were selected as its initial range boundaries: λ (d^{-1}) $\in$ [47, 121], E(I) (m^3 h^{-1}) $\in$ [0.248, 0.544], STD(I) (m^3 h^{-1}) $\in$ [0.194, 0.378], E(D) (s) $\in$ [36, 64], and STD(D) (s) $\in$ [49, 115] (Table S1). The GA was terminated if the residual error of F (i.e., $F_i - F_{i+1}$) of the best individual was less than 10^{-4} for 20 consecutive generations or the total number of generations reached 500. In this study, an optimal solution was generally achieved within 60 generations, and the optimal parameters were determined as follows: $\lambda = 91$ d^{-1}, E(I) = 0.346 m^3 h^{-1}, STD(I) = 0.292 m^3 h^{-1}, E(D) = 40 s, and STD(D) = 55 s.

3.3. PRP model validation

3.3.1. HWCP level

The CFDs of max (Q_i), V, Q_i, and ΔQ_i between the measured and simulated HWCPs (each 1000) are compared in Fig. 4. Note that these HWCPs were not superimposed into one or more RWCPs; rather, the examination was carried out exclusively at the HWCP level. Results indicate that the measured and simulated HWCPs agreed well for max (Q_i), Q_i, and ΔQ_i, but a relatively larger deviation was observed for V.

For max (Q_i), the measured values ranged from 0 to 0.72 m^3 h^{-1}, most of which (94%) fell between 0.04 and 0.40 m^3 h^{-1}; and the simulated values ranged from 0.04 to 0.80 m^3 h^{-1}. For Q_i, the measured and simulated curves almost overlapped; the measured values had the same range as those of max (Q_i), with 98.3% of which below 0.12 m^3 h^{-1} and 61.8% of which being zero (i.e., no water use); the simulated values had a slightly broader range (i.e., 0−0.80 m^3 h^{-1}), with 43.7% of which being zero. For ΔQ_i, which was simply derived from Q_i, the measured and simulated curves also overlapped as expected; the simulated values had a slightly broader range (i.e., from −0.72 to 0.72 m^3 h^{-1}) than the measured ones (i.e., from −0.64 to 0.60 m^3 h^{-1}). For V, the measured values ranged from 0 to 0.73 m^3, with only a small portion of which (0.2%) was zero (i.e., vacant homes); the simulated values ranged from 0.15 to 0.62 m^3. For max (Q_i), Q_i, and ΔQ_i, the simulated curves all had a slightly broader range than the measured ones, while for V, the simulated curve had a notably narrower range than the measured one.

The goodness-of-fit test results for the CFDs of these characteristic variables between the measured and simulated HWCPs are
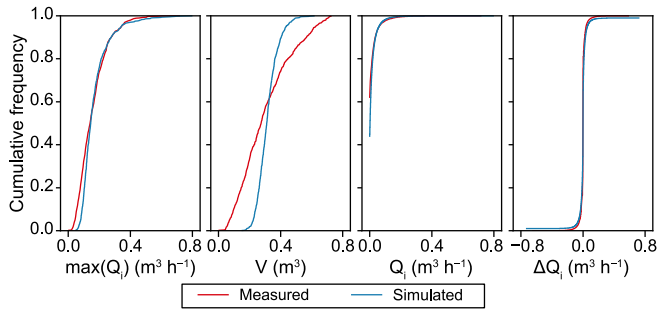
Fig. 4. Comparisons of CFDs of max (Q_i), V, Q_i, and ΔQ_i between measured and simulated HWCPs (1000 HWCPs each).

Table 1
Goodness-of-fit test for CFDs of max (Q_i), V, Q_i, and ΔQ_i between measured and simulated HWCPs (1000 HWCPs each).

| Characteristic variable	ME (%)	RMSE (%)	R^2
max (Q_i) (m³ h⁻¹)	−1.44	4.99	0.99
V (m³)	−3.45	26.32	0.88
Q_i (m³ h⁻¹)	0.06	0.74	0.98
ΔQ_i (m³ h⁻¹)	0.26	2.67	1.00

listed in Table 1. For max (Q_i), Q_i, and ΔQ_i, the calibrated PRP model showed good performance in simulation (|ME| ≤ 1.44%, RMSE ≤4.99%, R^2 ≥ 0.98); while for V, the simulation error was comparatively larger (|ME| = 3.45%, RMSE = 26.32%, R^2 = 0.88). This was also commonly observed in previous studies [22,26]. Overall, after being calibrated with the measured RWCPs (sampling size = 2000 HWCPs and DLI = 15 min), the PRP model could be validated at the HWCP level with acceptable simulation accuracies for the characteristic variables.

3.3.2. RWCP level

The daily variations of Q_i at six representative simulating sizes (50, 200, 500, 1000, 2000, and 3000 HWCPs) of the calibrated PRP model are shown in Fig. 5. For comparison, the max$_{mea}$, mean$_{mea}$, and min$_{mea}$ represented the upper envelope, mean, and lower envelope curves of the measured RWCPs (100), respectively; and the mean$_{sim}$ and random$_{sim}$ represent the mean curve and the random distribution region of the simulated RWCPs (100), respectively. Results indicate that the simulated mean curve agreed well with the measured one at each simulating size. Furthermore, the goodness-of-fit test for the CFDs of the upper envelope, mean, and lower envelope curves showed good agreement between the measured and simulated RWCPs (|ME| ≤ 8.89%, RMSE ≤10.85%, R^2 ≥ 0.94) (Table S2).

Fig. 5 Moreover, the CFDs of max (Q_i), V, Q_i, and ΔQ_i between the measured and simulated RWCPs (each 100) are compared in Fig. S2. For max (Q_i), the simulated curve almost overlapped with the measured one at 50 or 200 HWCPs, and slightly shifted to the right at other simulating sizes; for Q_i and ΔQ_i, the measured and simulated curves overlapped at each simulating size; and for V, the simulated curve had a notably narrower range than the measured one. This result (at the RWCP level) is similar to that shown in Fig. 4 (at the HWCP level).

The goodness-of-fit test results for the CFDs of these characteristic variables between the measured and simulated RWCPs are listed in Table S3. For max (Q_i), Q_i, and ΔQ_i, the calibrated PRP model showed quite good performance in simulation (|ME| ≤ 17.53%, RMSE ≤21.31%, R^2 ≥ 0.89); while for V, the simulation error was comparatively larger (|ME| = 29.14%, RMSE = 52.02%, R^2 = 0.72),
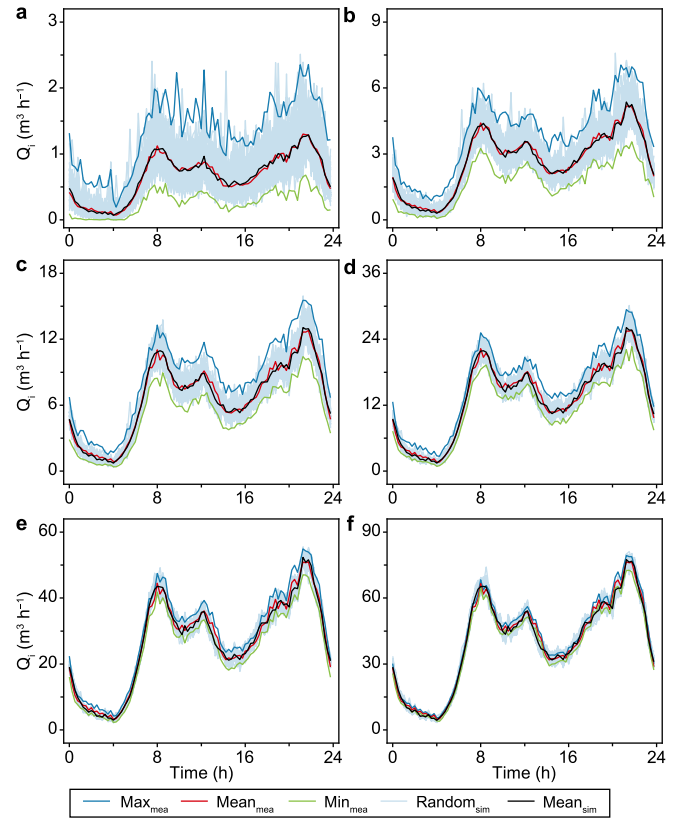
Fig. 5. Variations of Q_i over a day at representative simulating sizes of calibrated PRP model: **a**, 50; **b**, 200; **c**, 500; **d**, 1000; **e**, 2000; **f**, 3000 HWCPs.

similar to that observed in a previous study [31]. Overall, the calibrated PRP model could also be validated at the RWCP level, although with a certain decrease in the simulation accuracy compared to that at the HWCP level.

It should be noted that Beijing has a considerably higher residential density (i.e., a residential community usually containing 500−3000 households) than most cities worldwide. Hence, the PRP model established in this study purposely considered larger simulating sizes (e.g., 3000 households) than those established in other countries (≤500 households) [22,23,26]. Table S3 shows that the PRP model calibrated at the sampling size of 2000 HWCPs could well simulate the RWCPs with a simulating size of ≤2000 HWCPs, but the simulation error increased to some extent for the RWCP with a higher simulating size (i.e., 3000 HWCPs) in terms of max (Q_i) and V.

4. Discussion

4.1. Impact of DLI on RWCP uncertainty

Apart from the widely adopted 15-min time interval for DLI, water authorities also employ other DLIs, such as 30 and 60 min, to measure the RWCP. Therefore, the impact of DLI on RWCP uncertainty was investigated, as shown in Fig. S3. Results indicate at a fixed sampling size, the $\overline{UI}$ and |r| values decreased with an increasing DLI. For example, the maximum $\overline{UI}$ values were 1.60, 1.31, and 1.08 at the DLIs of 15, 30, and 60 min at the sampling size of 50 HWCPs, respectively. As the sampling size increased, the difference of $\overline{UI}$ between different DLIs gradually decreased. Due to the "leveling effect", prolonging the DLI could reduce the RWCP uncertainty, but at the cost of a decreased resolution of the RWCPs.

Table 2
PRP model parameters calibrated with water consumption data at different DLIs.

DLI (min)	λ (d^{-1})	E(I) (m^3 h^{-1})	STD(I) (m^3 h^{-1})	E(D) (s)	STD(D) (s)
15	91	0.346	0.292	40	55
30	85	0.364	0.324	36	56
60	77	0.248	0.220	56	92

Based on the desired thresholds ($\overline{UI} < 0.2$ and $|r| < 0.01\%$) for dependable RWCPs as aforementioned (Section 3.1), the critical sampling sizes were 2000, 1800, and 1600 HWCPs in correspondence to the DLIs of 15, 30, and 60 min, respectively (Fig. S3). The relationship between the DLI and the sampling size could guide water authorities for field measurements of dependable RWCPs while avoiding excessive data collection.

4.2. Comparison of PRP model parameters calibrated at different DLIs

The PRP model parameters were also calibrated at the DLIs of 30 and 60 min, as listed in Table 2. Results indicate that the parameters calibrated at a DLI of 15 min closely resembled those at a DLI of 30 min, but deviated more from those at a DLI of 60 min. Specifically, the calibration at a DLI of 60 min yielded a smaller λ and E(I) but larger E(D) than those at the other two DLIs. Consequently, the PRP model calibrated at a longer DLI would generate water consumption events with a lower frequency and intensity but a longer duration. However, the water consumption volumes represented by the pulses (E(I) × E(D)) were similar at the three different DLIs (i.e., 3.84, 3.64, and 3.86 L, respectively), which is consistent with the actual situation of a household.

For the measurement of RWCPs, a short DLI leads to a high resolution but at a high cost for data collection and analysis [6], while a long DLI leads to a lower resolution and sometimes even distortion. Therefore, considering the practical circumstance, a DLI of ≤30 min is recommended to calibrate the PRP model.

4.3. Comparison of simulation accuracies of PRP model calibrated at different sampling sizes

To further investigate the impact of RWCP uncertainty on model simulation accuracy, the PRP model was calibrated at four additional sampling sizes (50, 500, 1000, and 3000 HWCPs). Fig. 6 shows the variations of ME and RMSE for the CFDs of max (Q_i), V, and Q_i at representative simulating sizes (i.e., 1, 50, 200, 500, 1000, 2000, and 3000 HWCPs). The goodness-of-fit test results clearly demonstrate that the PRP model calibrated at a smaller sampling size would induce a larger simulation error, and the error was amplified with increasing simulating size.

The PRP model, calibrated at the sampling size of 2000 HWCPs in this study, exhibited good accuracy for all the representative simulating sizes. In addition, it was found that the PRP model calibrated at a certain sampling size (e.g., 100 HWCPs) performed well in simulating RWCPs with an equal or lower size (i.e., ≤100 HWCPs) but was less effective in simulating RWCPs with a higher size (i.e., >100 HWCPs). Therefore, it is recommended that the PRP model, once calibrated, should better be applied "downwards".

5. Conclusions

The primary aim of this research was to establish the PRP model more accurately based on the uncertainty analysis of RWCPs practically measured. Based on the outcomes of the analysis, several key findings can be summarized as follows.

- The reduction of RWCP uncertainty can be effectively attained by increasing the sampling size or DLI. To establish a RWCP with high confidence, the study proposed the adoption of the
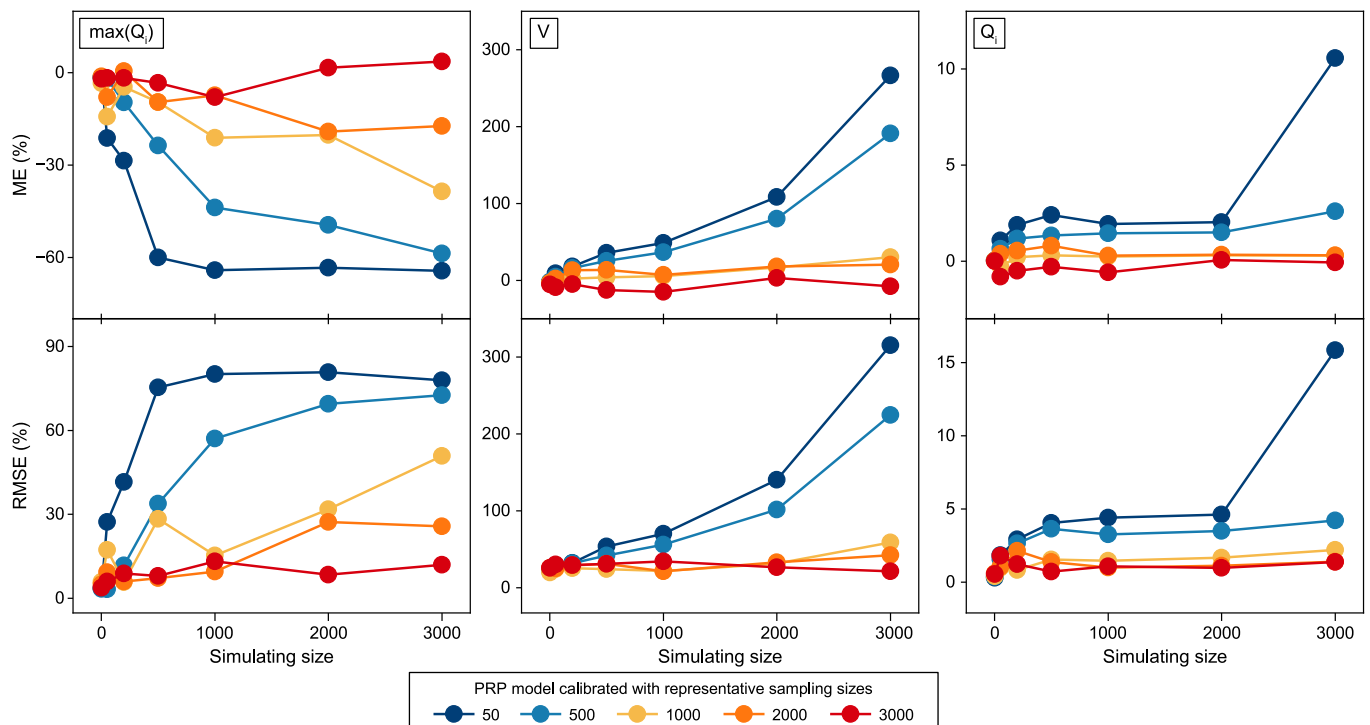

Fig. 6. Variations of ME and RMSE for CFDs of max (Q_i), Q_i, and ΔQ_i with simulating size of PRP model calibrated separately with representative sampling sizes.

following threshold criteria: $\overline{UI} < 0.2$ and $|r| < 0.01\%$. On this basis, the critical sampling sizes were determined as 2000, 1800, and 1600 HWCPs in correspondence to the DLIs of 15, 30, and 60 min, respectively.

- Through PRP model calibration, we identified the optimal parameters at a DLI of 15 min. The calibrated values for the parameters were as follows: $\lambda = 91$ d^{-1}, $E(I) = 0.346$ m^3 h^{-1}, $STD(I) = 0.292$ m^3 h^{-1}, $E(D) = 40$ s, and $STD(D) = 55$ s. Notably, these optimized parameters showed a resemblance to those obtained at a DLI of 30 min, whereas they exhibited more substantial deviations from those at a DLI of 60 min. Hence, a DLI of ≤ 30 min is recommended to calibrate the PRP model.

- To ensure the simulation accuracy, a sampling size of ≥ 2000 HWCPs is recommended to calibrate the PRP model. Once calibrated at a certain sampling size, the PRP model should be better applied "downwards" to simulate RWCPs with an equal or smaller size.

CRediT author contribution statement

Jiaxin Zhang: Methodology, Software, Writing - Original Draft. **Dragan Savic:** Writing - Review & Editing. **Qiang Xu:** Conceptualization, Writing - Review & Editing, Supervision. **Kuo Liu:** Resources. **Zhimin Qiang:** Writing - Review & Editing, Supervision, Funding Acquisition.

Declaration of competing interest

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

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2023.100317.

References

[1] K. Wang, E.G.R. Davies, Municipal water planning and management with an end-use based simulation model, Environ. Model. Software 101 (2018) 204–217.

[2] L. Chen, H.X. Yan, J.R. Yan, J.Y. Wang, T. Tao, K.L. Xin, S.P. Li, Z.H. Pu, J. Qiu, Short-term water demand forecast based on automatic feature extraction by one-dimensional convolution, J. Hydrol. 606 (2022) 127440.

[3] T. Salloom, O. Kaynak, W. He, A novel deep neural network architecture for real-time water demand forecasting, J. Hydrol. 599 (2021) 126353.

[4] S.G. Buchberger, L. Wu, Model for instantaneous residential water demands, J. Hydraul. Eng. 121 (3) (1995) 232–246.

[5] A. Di Nardo, M. Di Natale, R. Gargano, C. Giudicianni, R. Greco, G.F. Santonastaso, Performance of partitioned water distribution networks under spatial-temporal variability of water demand, Environ. Model. Software 101 (2018) 128–136.

[6] E. Clifford, S. Mulligan, J. Comer, L. Hannon, Flow-signature analysis of water consumption in nonresidential building water networks using high-resolution and medium-resolution smart meter data: two case studies, Water Resour. Res. 54 (1) (2018) 88–106.

[7] R. Gargano, F. Di Palma, G. de Marinis, F. Granata, R. Greco, A stochastic approach for the water demand of residential end users, Urban Water J. 13 (6) (2015) 569–582.

[8] Y. Zhao, Y.J. Yang, Y. Shao, Y. Lee, T. Zhang, Demand-driven spatiotemporal variations of flow hydraulics and water age by comparative modeling analysis of distribution network, J. Water Resour. Plann. Manag. 144 (12) (2018) 04018074.

[9] Y.P. Wu, S.M. Liu, X. Wu, Y.F. Liu, Y.S. Guan, Burst detection in district metering areas using a data driven clustering algorithm, Water Res. 100 (2016) 28–37.

[10] X.T. Wang, G.C. Guo, S.M. Liu, Y.P. Wu, X.Y. Xu, K. Smith, Burst detection in district metering areas using deep learning method, J. Water Resour. Plann. Manag. 146 (6) (2020) 04020031.

[11] E.J.M. Blokker, J.H.G. Vreeburg, S.G. Buchberger, J.C. van Dijk, Importance of demand modelling in network water quality models: a review, Drink. Water Eng. Sci. 1 (2008) 27–38.

[12] J.E. Pesantez, E.Z. Berglund, N. Kaza, Smart meters data for modeling and forecasting water demand at the user-level, Environ. Model. Software 125 (2020) 104633.

[13] E.B. Rizzo, F.A. Cousin, R.M. Lucca, S.R. Lautenschlager, Autonomous metering system for monitoring water consumption, AQUA Water Infrastruct. Ecosyst. Soc. 70 (6) (2021) 797–810.

[14] K. Muniina, C. Maksimovic, N. Graham, A novel approach for estimating urban water end use characteristics of cities in the developing world, Urban Water J. 14 (7) (2017) 750–757.

[15] S.G. Buchberger, G.J. Wells, Intensity, duration, and frequency of residential water demands, J. Water Resour. Plann. Manag. 122 (1) (1996) 11–19.

[16] E. Creaco, P. Kossieris, L. Vamvakeridou-Lyroudia, C. Makropoulos, Z. Kapelan, D. Savic, Parameterizing residential water demand pulse models through smart meter readings, Environ. Model. Software 80 (2016) 33–40.

[17] E. Creaco, R. Farmani, Z. Kapelan, L. Vamvakeridou-Lyroudia, D. Savic, Considering the mutual dependence of pulse duration and intensity in models for generating residential water demand, J. Water Resour. Plann. Manag. 141 (11) (2015) 04015031.

[18] R. Guercio, R. Magini, I. Pallavicini, Instantaneous residential water demand as stochastic point process, Water Resour. Manag. 48 (2001) 129–138.

[19] V.J. Garcıa, R. Garcıa-Bartual, E. Cabrera, F. Arregui, J. Garcıa-Serra, Stochastic model to evaluate residential water demands, J. Water Resour. Plann. Manag. 130 (5) (2004) 386–394.

[20] V.H. Alcocer-Yamanaka, V.G. Tzatchkov, F.I. Arreguin-Cortes, Modeling of drinking water distribution networks using stochastic demand, Water Resour. OR Manag. 26 (2012) 1779–1792.

[21] V.H. Alcocer-Yamanaka, V. Tzatchkov, S.G. Buchberger, Instantaneous water demand parameter estimation from coarse meter readings, in: Proceedings of the 8th Annual Water Distribution Systems Analysis Symposium, 2006. Cincinnati, Ohio, USA.

[22] E.J.M. Blokker, S.G. Buchberger, J.H.G. Vreeburg, J.C. van Dijk, Comparison of water demand models: PRP and SIMDEUM applied to Milford, Ohio, data, in: Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, Kruger National Park, South Africa, 2008.

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

[24] S. Alvisi, N. Ansaloni, M. Franchini, Generation of synthetic water demand time series at different temporal and spatial aggregation levels, Urban Water J. 11 (4) (2014) 297–310.

[25] Y.Y. Zhao, J. Yang, Y. Shao, Y. Lee, T.Q. Zhang, Demand-driven spatiotemporal variations of flow hydraulics and water age by comparative modeling analysis of distribution network, J. Water Resour. Plann. Manag. 144 (12) (2018) 04018074.

[26] E. Creaco, E.J.M. Blokker, S.G. Buchberger, Models for generating household water demand pulses: literature review and comparison, J. Water Resour. Plann. Manag. 143 (6) (2017) 04017013.

[27] S.G. Buchberger, Z.W. Li, PRPsym: A Modeling System for Simulation of Stochastic Water Demands, Proceedings of World Environmental and Water Resources Congress, Tampa, Florida, USA, 2007, 2007.

[28] S. Alvisi, M. Franchini, A. Marinelli, A stochastic model for representing drinking water demand at residential level, Water Resour. Manag. 17 (2003) 197–222.

[29] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison Wesley Publishing Company, Boston (USA), 1989.

[30] Y.J. Lei, S.W. Zhang, X.W. Li, Y. Lei, MATLAB Genetic Algorithm Toolbox and its Application, Xi Dian University Press, Xi'an (China), 2014 (In Chinese).

[31] F. Di Palma, R. Gargano, F. Granata, R. Greco, The overall pulse model for water demand of aggregated residential users, Procedia Eng. 186 (2017) 483–490.