INTRODUCTION

Hot water systems are the primary source of building energy consumption in both developed and developing countries. In China, hot water accounts for approximately 23.4% of total domestic energy consumption (excluding heating energy consumption) in residential areas. In Brazil, electric hot water for showers accounts for 24% of a typical household's energy consumption. In other countries, such as the US and various nations in the European Union, domestic hot water consumption accounts for approximately 14%-25% of the total energy consumption in residential buildings. Because the energy consumption of a hot water system is relatively large, special attention should be paid to reducing building energy consumption.

With the aim of saving energy in hot water systems, many previous studies have been conducted. In general, there are two ways to reduce energy consumption: (a) improving the efficiency of the primary equipment, and (b) optimizing the control method to balance the supply and demand more effectively, thereby reducing the energy consumption during an operation period. Many studies have been conducted in the former area using higher-efficiency boilers, heat pumps, and solar heaters. In recent studies, combined heating...
equipment such as the solar assisted heat pump (SAHP) \cite{11} has been proposed to improve the equipment efficiency further.\cite{12}

Although an increased efficiency of the primary equipment can decrease the amount of energy consumption in a hot water system, this decrease often creates a bottleneck in reality, when most of the equipment has been updated to achieve better efficiency.\cite{13} Most people tend to neglect the energy saving potential during an operation period, despite the fact that a significant amount of energy can still be saved. A study from the Netherlands revealed that if the occupants do not conduct their tasks in a highly efficient way, that is, one that supports the intended design, even a high-performance system can consume much more energy than expected.\cite{14}

A properly designed control method is considered an effective means of energy saving during an operation period. Several control methods have been proposed for hot water systems, aiming at minimizing the energy consumption or energy cost. Among the reported studies, predictive control\cite{15} has stood out as an effective method for a hot water system to achieve good energy and cost saving results. Michael et al\cite{17} used a model predictive control method of a price-based demand response for heat pump utilization. The costs were reduced by approximately 5\%, whereas the morning peak of district heating and evening peak of the electric grid were relatively reduced. Rasmus et al\cite{16} used economic model predictive control (EMPC) and obtained electricity cost savings of up to 25\%-30\%.

A proper prediction of the energy load is an important task for predictive control, which is formidable\cite{17} because the future is unknown and uncontrollable. With the development of data science,\cite{18} data-driven prediction methods for building energy consumption have become a research hotspot in recent years.\cite{19} Compared with traditional predictive control methods, a data-driven method is more precise and achieves a better match in individual situations.\cite{19} Catalina et al\cite{20} developed a regression model to predict monthly heating demand for a single-family residential sector. Raza et al\cite{21} reported the superiority of support vector machines (SVMs) over other methodologies such as artificial neural networks (ANNs). Comprehensive reviews introducing data mining methodologies and practical applications regarding energy consumption prediction in buildings can be found in.\cite{22}

Occupant behavior has a significant impact on the results of future energy demand prediction and has become an important aspect in reducing thermal losses from energy inefficient buildings.\cite{23} Kazmi et al\cite{13} further indicated that obtaining a better understanding of occupant behavior only increases in importance as a necessary first step. However, for most data-driven predictive methods, a prediction is usually conducted on an aggregated energy consumption. This provides the model certain advantages by extracting personal behavior from the aggregated data, then analyzing and putting them into practical use. Kazmi et al\cite{13} proposed an occupant data-driven optimization model to operate an air-source heat pump at the highest possible efficiency, and the results indicate a 20\% reduction in energy consumption using a computationally inexpensive heuristic approach. Aki et al\cite{19} used support vector regression (SVR) to predict hot water demand in a bathtub as well as many other end users and proposed a management system in a practical environment.

In the case of a domestic hot water system, better control of the remaining hot water in a hot water storage tank can help reduce heat loss, which accounts for up to 20\% of the total energy consumption in a DHW system, and therefore reduce the amount of energy consumption. This is practically useful considering its application to not only new devices but also to the millions of devices that have already been installed in households.\cite{23} A better balance between supply and demand will help decrease the remaining hot water in the tank and therefore realize greater energy saving.\cite{24}

This study aims to reduce the energy consumption of a DHW system during an operation period using a predictive control method and to achieve a better balance between supply and demand. This includes two main parts: proposing a predictive control method and calculating the amount of energy saving. In the first part, a predictive control method is raised based on a novel data-driven algorithm for predicting the DHW demand. This algorithm takes each occupant’s showering behavior as the first reference by using SVM. Future hot water demand is calculated by summing each occupant’s individual prediction results, which differs from other prediction methods that directly predict the aggregated energy consumption or demand. In the second part, an energy consumption model of a DHW system is built to estimate the energy-saving effect by adopting the proposed predictive control method.

Moreover, this research is grounded on real data. Detailed user behavior data are often difficult to collect, thus making a behavior analysis and real-data-based prediction difficult to conduct. This paper fills the gap
regarding the use of occupant behavior characteristics to predict future energy demand. Therefore, results from the simulation are more convincing and more suitable for a practical application.

The novel contributions of this paper not only lie in a data-driven algorithm for hot water demand prediction, but also in a complete predictive control method designed for a DHW system. This includes an algorithm for using the occupant behaviors to calculate the future hot water demand and a hot water supply strategy for the entire system. Such a control method can be easily applied in a similar hot water system with little equipment retrofit, little effect on user comfort, and considerable energy savings.

This paper is divided into three main sections: methodology, evaluation, and results and analysis. Among them, the methodology section describes the method for predicting the shower habits of each user, a way to calculate domestic hot water demand using the prediction results, and a related hot water supply strategy. In the evaluation section, a model of a domestic hot water system is described, and an evaluation of the proposed predictive control method is conducted through a simulation. The results and analysis section describes the prediction and simulation results and provides a comparative analysis with other control strategies.

2  |  METHODOLOGY

The predictive control method for a hot water system is developed in the following steps:

1) Establish a predictive model to learn and predict user shower habits hourly.
2) Calculate future hot water demand using the prediction results of user shower habits.
3) Propose a domestic hot water supply strategy based on future hot water demand.

The first part is the core of the entire predictive control method. The problem can be defined as judging whether a user will take a shower in the next hour and can be modeled as a binary classification problem.

2.1  |  Data description

Because water control units (WCUs), as shown in Figure 1, are widely used in many well-controlled DHW systems, particularly in dormitories and public bathrooms for individual charge, they can be used to collect hot water consumption data. An integrated circuit (IC) card is used to control the switch of the showerhead. WCUs will automatically record the IC number, start time (including the month, day, hour, minute, and second), end time, and other relevant information. The start time is defined as the shower time, whereas the end time is used to calculate the duration of a shower. Water consumption per shower is calculated based on the duration of the shower, at a water flow speed of 500 L/hour (8.33 L/minute).²⁵

Seven occupants living in Shanghai, China (31°01′N 121°25′E) volunteered to provide shower records for this research. General information of each occupant is listed in Table 1. A total of 1250 shower records were collected from February 1 to November 30, 2017. Each record includes 10 labels: the IC number, shower time (including the day and hour), hot water consumption, temperature, humidity, wind speed, solar radiation, weekday, and time interval between two showers. The IC number is used to identify the occupant. The amount of hot water consumption is used to generate the predicted water demand. The other eight labels are used as elements to predict the shower habits of occupants.

A raw dataset is prewashed before training. Low-quality data, in which the water consumption is less than 10 L or larger than 300 L, are removed. Because the tourist season is in August, all shower records from this month were also removed to unify the time stamp. The after-treated dataset was then divided into a training set, validation set, and test set, as shown in Table 2.
The shower habits of seven users are analyzed to find the main influencing factors. Fuentes et al reported that daily and seasonal variables (i.e., shower time/hour and month) are the most significant factors. In addition, the time interval between showers is also considered in this study. Further analyses of the shower habits of all users along with the parameters were conducted, the results of which are shown in Figure 2.

From Figure 2A, it can be observed that the fraction of shower incidents is related with the time interval for most users. Fraction peaks happen in a time interval of between 20 and 30 h, which means the users take a shower every day (24 hours). It can be noted that users 5 and 7 have similar patterns but differ from the others. Instead of having a single peak, they have several small peaks, which are at time intervals of 1, 2, 3, or 4 days (24, 48, 72, or 96 hours). This shows that the users take a shower every several days, but not at a fixed number.

The relation between the fraction of incidents and the shower time (hour) was investigated, and the results are shown in Figure 2B. It was observed that most users tend to take a shower after 18:00, whereas some tend to take a

| User | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total | Share |
|------|---|---|---|---|---|---|---|------|-------|
| Training | 94 | 126 | 118 | 72 | 38 | 100 | 37 | 585 | 46.8% |
| Validation | 32 | 43 | 43 | 29 | 29 | 29 | 29 | 224 | 17.92% |
| Test | 74 | 86 | 78 | 60 | 37 | 82 | 24 | 441 | 35.28% |
| Total | 200 | 255 | 239 | 161 | 104 | 211 | 80 | 1250 | 100% |

FIGURE 2 Shower habit analysis of each user. (A: Fraction of shower incidents with Time Interval; B: Fraction of shower incidents with Hour; C: Fraction of shower incidents with Month; D: Average hot water consumption change with Month)
shower both in the morning and in evening. Similar shower habits have also been reported by Fuentes et al.25

The relationship between hot water consumption and the seasonal variable is investigated. Hot water consumption is strongly related with the month mainly because of the seasonal changes in the environmental temperature.25 Many researchers have reported a trend in which hot water consumption will increase as the environmental temperature decreases.26 Although the value varies based on the different regions and climate conditions, a similar trend is shown. George et al.27 conducted a high-resolution measurement in 119 houses in Canada and found a +6.8% variation of hot water consumption in cold periods and a −0.96% variation in warm periods. Ahmed et al. conducted onsite data measurements in 185 apartments in Finland and found a +15.3% variation in cold periods and −17.4% in warm periods.28

The increase in hot water consumption in cold months is mainly due to the increase in hot water consumption per shower.29 This point is also supported by the data collected in this study. The relationship between the fraction of shower incidents and month is shown in Figure 2C. It can be seen that the number of shower incidents does not change significantly by month for most users, whereas some users (users 4 and 5) have more shower incidents in warm months (May to October) than in cold months (November to April of the next year). The average hot water consumption of each month with respect to the average hot water consumption of the entire year is shown in Figure 2D. It can be observed that hot water consumption per shower changes with the season, and in warm months, all users tend to use less water per shower, whereas in cold months the opposite trend occurs. This explains why hot water consumption is higher in winter than in summer. Detailed calculations were also conducted, the results of which are listed in Table 3, which will be used in calculating the water demand as a seasonal coefficient.

Based on the information above, the time interval and shower time are treated as the main factors in predicting a user’s shower habits, whereas the month is treated as a minor factor with a small weight during this process. However, the month is treated as a significant factor when calculating domestic hot water demand, as described in Section 2.5.

### 2.2 Mathematical methodology

The prediction of a user’s future shower status can be simplified as a binary classification problem. Hourly data of a user is labeled as 1 when the user has a shower, and 0 when the user does not. The SVM algorithm is used in this work to learn the shower habits of users. SVM is a well-developed data mining technology derived from the statistical learning theory introduced by Vapnik.30 The SVM model is based on a series of kernel functions, which convert the original, low-dimensional data to a higher feature space in an implicit manner.31 It has recently been used successfully in various areas including solving complex problems of pattern recognition, classification, regression analysis, and forecasting and shows a greater performance compared to other data mining technologies such as a neural network32 and a statistical data processing method.33

A brief introduction to the principle of SVM is provided below. For a given set of instance-label pairs \((x_i, y_i)\), \(i = 1, \ldots, l\),

\[
y_i(x_i) = w^T x_i + b
\]

where \(x_i \in \mathbb{R}^n\), \(y_i \in \{1, -1\}\). Training vectors \(x_i\) are converted into a higher dimensional space by the function \(\phi(x)\). To maximize the slack between hyperplanes, the optimization problem of the SVM can be defined as follows:

\[
\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i
\]

subject to

\[
\begin{align*}
y_i &\geq 1 - \xi_i \\
\xi_i &\geq 0
\end{align*}
\]

where \(w\) and \(b\) are the weight and constant term, respectively, and can be used to describe the relationship between \(x\) and \(y\). Here, \(\xi\) is called a slack variable, and \(C > 0\) is the penalty parameter. For nonlinear classification problems, the vectors \(x_i\) can be mapped into a nonlinear classifier using a Kernel trick.

\[
K(x_i, x_j) = \phi(x_i)^T \phi(x_j)
\]

where \(K\) is called a kernel function. The four popular kernel functions are linear, polynomial, radial basis function (RBF), and sigmoid functions. Among them, the RBF function shows a better performance in many nonlinear cases because of its computational effectiveness, convenience, reliability, ease of adaptation to optimize other adaptive techniques, and adaptability in handing complex parameters.32 The definition of the RBF kernel function is as follows:

\[
K(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2}
\]

where \(\gamma > 0\) is called a kernel parameter.

The dataset may still not be linearly separable in a converted space, that is, it may not be able to use a linear decision boundary to divide all samples. An SVM classifier will

| Season          | User 1  | User 2  | User 3  | User 4  | User 5  | User 6  | User 7  |
|-----------------|---------|---------|---------|---------|---------|---------|---------|
| Warm (May-October) | −0.70%  | −5.14%  | −9.60%  | −12.07% | −15.77% | 1.43%   | −11.98% |
| Cold (November-April) | 0.84%   | 6.17%   | 11.52%  | 14.49%  | 18.92%  | 14.92%  | 14.38%  |
not converge under this circumstance. In this case, the kernel function, slack variable $\xi$, and penalty parameter $C$ should be adjusted to change the soft margin for a decision boundary.\textsuperscript{30}

In this study, the prediction is conducted hourly. At hour $t$, if there is shower record for user $u$, the shower status $\psi_{u,t}$ equals 1. Otherwise, $\psi_{u,t}$ equals 0. The prediction output is the predicted shower status $\psi'_{u,t}$ for user $u$ at time $t$, that is, 0 or 1.

### 2.3 Evaluation criteria

The root mean square error (RMSE) is used to evaluate the future hot water demand. For a binary classification problem, criteria that are widely adopted in building energy prediction such as the coefficient of variation (CV), mean absolute percentage error (MAPE), and RMSE \textsuperscript{22} are not suitable in this case. The accuracy of the binary classification is defined as follows:

$$a = \frac{N_{\text{true}}}{N_{\text{false}}}$$  \hspace{1cm} (6)

where $a$ is the accuracy, $N_{\text{true}}$ is the number of data in which the prediction matches the experimental result, and $N_{\text{false}}$ is the number of data in which the prediction does not match.

However, because users usually take a shower 1.7 times a day,\textsuperscript{25} it indicates that most of the shower data are zeros during most hours, and the occupant’s behavior dataset is lopsided. This may cause a problem because the classifier will overfit zero data while underfitting data with a value of 1. This sabotages the utility of the prediction because the shower time is the focus of this work and is important to predict the hot water demand.\textsuperscript{34}

Compared with accuracy, confusion matrix (CM, shown in Table 4) is a more intuitive description of the training result for a binary classification problem. The confusion matrix consists of four types of classification results: true positive (TP), true negative (TN), false negative (FN), and false positive (FP). In this problem, FN indicates that the model failed to predict an occupant’s shower incident, which may lead to missing hot water demand of the occupant. FP indicates that the model predicts that the occupant will take a shower, but in fact the occupant does not, which causes more energy waste in the ST. These two types of false classifications should all be avoided as much as possible, but in comparison, the effect caused by FN is more grievous than that of FP.

The relation between accuracy and CM is shown as below:

$$N_{\text{true}} = TP + TN$$  \hspace{1cm} (7)

### 2.4 Training process

To reduce the effects of a lopsided dataset, random undersampling is used in building the training set of the predictive model, which means only some of the records from a large-quantity class contribute to building a predictive model. A sampling ratio of 1:1 to 1:10 was chosen and treated as one of the SVM parameters to conduct the optimization.\textsuperscript{34} The optimized result is 1:1.

Predictive models for each occupant are built individually. There are three parameters used in the SVM model: the RBF parameter $\gamma$, penalty parameter $C$, and termination criterion tolerance $\varepsilon$. Here, $\varepsilon$ is set to 0.001 while the SVM model is tested using a cross validation to find the best combination of the other two parameters, $\gamma$ and $C$. The optimal parameters are selected and listed in Table 5. The optimal training results of each occupant are shown in Table 6. The training results show that the combination of accuracy and recall is a better evaluation criterion than using them individually.

Further experiments were carried out to find the minimum dataset size for prediction. The prediction model is built from

\begin{table}[h]
\centering
\caption{SVM parameters for binary classification of shower status (separate prediction)}
\begin{tabular}{|l|c|}
\hline
\textbf{SVM type}               & \textbf{Binary classification} \\
\hline
Feature number               & 2     \\
Output                       & 0 or 1 \\
Sampling ratio               & 1:1    \\
Kernel type                  & RBF   \\
$\gamma$                     & 0.01   \\
$C$                          & [3 15] \\
$\varepsilon$                & 0.001  \\
\hline
\end{tabular}
\end{table}

$N_{\text{false}} = TP + TN$  \hspace{1cm} (8)

It is worth noting that occupant comfort is regarded as being of most importance, and the new control method should try not to disturb the user’s shower habits. Under this consideration, recall ($r$) is used as the other important evaluation criteria along with accuracy. Recall is a commonly used criterion to deal with lopsided binary classification problems, showing the portion of TP records in all actual positive incidents. This is defined as follows:

$$r = \frac{TP}{TP + FN}$$  \hspace{1cm} (9)

The evaluation criterion is defined as the weighted average of the recall and accuracy during the training process, the corresponding weights of which are 0.7 and 0.3. Recall is also used in the predictive control method to check when the model deviates with the real data provided in section 3.2.
the same validation dataset. The minimum dataset sizes for each person are shown in Figure 3, and the result shows that the required dataset size for prediction is not the same among different people. For users 2, 3, 4, and 6, an algorithm only needs 10 days of data collection (approximately 7-10 shower incidents) to obtain a prediction model with the required criteria, whereas 90 days of data collection (approximately 30-50 shower incidents) are needed for users 1, 5, and 7.

This result is useful when new users arrive. Here, the system needs to learn the new shower habit first, which requires time to collect the user’s data and run a test. The duration of this process depends on the user, but under favorable circumstances 10 days is sufficient to meet the requirement to generate the new user's prediction model. For some users, it may take up to 90 days or even longer. During this period, the system will continuously prepare adequate hot water for the new user.

2.5 Calculation of domestic hot water demand

The future domestic hot water demand can be calculated from the average water consumption \( V_u \) and standard deviation \( \sigma_u \) with the prediction model of the user’s shower behavior using the following equation:

\[
V_{d,t+1} = \sum_{u=1}^{7} \psi_{u,t+1} \ast (V_u + \sigma_u) \ast (1 + \varsigma_u) \ast \varphi_u, \quad u = 1, \ldots, 7
\]

where \( \psi_{u,t+1} \) is the predicted shower status, \( V_u \) is the average hot water consumption, \( \sigma_u \) is the standard deviation, \( \varsigma_u \) is the seasonal coefficient, and \( \varphi_u \) is the adjustment coefficient.

The values of \( V_u \) and \( \sigma_u \) for each user are listed in Table 7. A variation analysis is conducted for each user in both cold (November to February) and warm (May to October) seasons. The results are also shown in Table 7. Here, \( \varphi_u \) reflects the variation of each user, and ranges from 0.5 to 1.2. This value is obtained by optimizing the result during the validation process to satisfy the hot water demand by assuring that the remaining hot water in the ST is always above zero to obtain 100% hot water supply assurance, as shown in Figure 4. The optimal values for \( \varphi_u \) are listed in Table 7.

2.6 Hot water supply strategy

The prediction model of hot water demand can be used to control a DHW system more effectively. The scheme of this process can be seen in Figure 5. When \( \psi_u \) turns from zero to 1, the DHW system produces hot water to prepare a line for user \( u \) in the hot water storage tank (ST, see Figure 5A at time 1). When user \( u \) uses hot water (see Figure 5A at time 3), the DHW system stops producing hot water for the user's share. The hot water remaining in the ST becomes part of the next-hour hot water preparation. When another user uses hot water, this margin will be offset (see Figure 5B at time 2) by the system. This
The process ends when user $u$ consumes hot water or $\psi_u$ turns from 1 to zero, and the remaining hot water in the ST is released.

3 | EVALUATION

The proposed predictive control method was simulated based on a dormitory DHW system located in Shanghai Jiao Tong University. Using the parameters of the DHW system model for seven users, the control method was applied with a time step of 1 hour from September 1 to November 30, 2017.

3.1 | Description of DHW system

The hot water system designed to evaluate the proposed prediction control method is shown in Figure 6. The design of the system conforms to GB 50015-2003 (2009) and satisfies the need of the maximum capacity (105 L for each user per day) as presented by Fuentes et al. The parameters of the DHW system are shown in Table 8.

A water boiler and hot water storage tank (ST) are chosen as the two main pieces of equipment in the proposed DHW system. The incoming water is heated to the water supply temperature by the water boiler and stored in the ST. The ST takes energy to keep the water at the water supply temperature. A detailed design of the ST is used for simulation accuracy, and the geometric dimensions of the ST are as shown in Table 9. The value of the heat conductivity coefficient of each layer and the outside convection coefficient were adopted from a study by Kiyan et al. By connecting all original sensors such as the liquid level meter and temperature sensors to the controller, the water control unit (Figure 1) is implemented to each showerhead in the shower room. Therefore, the shower data of the occupants can be collected by the controller and can be used to conduct the water demand prediction to proceed with the control method.

3.2 | Predictive control method for DHW system

The predictive control method contains initialization mode and system running mode. It can be expressed in Figure 7. In the initialization mode, the controller build prediction model based on occupants’ historical behavior dataset using
the method presented in section 2.2 and 2.4. In the system running mode, hot water system predicts occupant’s shower status in the next hour using the method presented in section 2.5. The amount of hot water produced by water boiler and stored in ST are controlled by the strategy proposed in 2.6. A fault tolerance mechanism is adopted in both initialization and system running period. When the occupant's behavior changes dramatically and the criteria of the built model cannot be guaranteed, this mechanism will not accept the prediction. The acceptable criteria are shown in Table 10.

3.3 | Mathematical physical model

3.3.1 | Heat loss model

Heat loss from the ST is the main heat loss source in the DHW system. In order to reveal the relationship between heat loss and water level in the ST, a more detailed heat loss model was built in this paper. The calculation of the heat loss in the ST was conducted under the following assumptions:

1. Hot water in the ST is considered to be kept at 60°C, uniformly and constantly.
2. The physical properties of air and water in the ST stays constant.
3. Outside convection coefficient stays constant.
4. Other part of DHW system is well-insulated.

Heat loss can be calculated from four parts in Figure 8. \( \Phi_{t,\text{bottom}} \) is the heat loss from the bottom. \( \Phi_{t,\text{lowerside}} \) is the heat loss from the lower side where heat transfers from the hot water through the heat-preservation wall to the environment. \( \Phi_{t,\text{upperside}} \) is the heat loss from the upper side where heat transfers from the hot air in the tank through the heat-preservation wall to the environment. \( \Phi_{t,\text{top}} \) is the heat loss from the top. Total heat loss from the tank is the summary of all heat loss:

\[
\Phi_{\text{tank}} = \Phi_{t,\text{top}} + \Phi_{t,\text{lowerside}} + \Phi_{t,\text{upperside}} + \Phi_{t,\text{bottom}} \tag{11}
\]

The four types of heat loss will be calculated individually. \( \Phi_{t,\text{bottom}} \) and \( \Phi_{t,\text{lowerside}} \) can be calculated using Fourier’s Law.36
Heat loss from the bottom of the tank can be calculated by the following:

\[ q = -\frac{dT}{dx} \]  

(12)

\[ \Phi = q \times A \]  

(13)

Heat loss from the bottom of the tank can be calculated by the following:

\[ q_{t, \text{bottom}} = \frac{T_{\text{water}} - T_{\text{amb}}}{\frac{1}{h_w} + \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3}} \]  

(14)

\[ A_{\text{bottom}} = \pi \left( \frac{d_{t, \text{inner}}}{2} \right)^2 \]  

(15)

\[ \Phi_{t, \text{bottom}} = q_{t, \text{bottom}} \times A_{\text{bottom}} \]  

(16)

Similarly, the heat loss from the lower side of the tank can be calculated by the following:

\[ q_{t, \text{lowerside}} = \frac{T_{w} - T_{\text{amb}}}{\frac{1}{h_w} + \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3}} \]  

(17)

\[ A_{\text{lowerside}} = \pi \times d_{t, \text{inner}} \times H_w \]  

(18)

\[ \Phi_{t, \text{lowerside}} = q_{t, \text{lowerside}} \times A_{\text{lowerside}} \]  

(19)

The air in the upper part is heated by the hot water below and can be simplified as finite space natural convection heat transfer. The convection in the interlayer can be determined...
For a horizontal interlayer (bottom heated), when $Gr > 2430$, natural convection appears and it becomes more intensive when $Gr$ is bigger. The horizontal interlayer finite space natural convection heat transfer can be calculated empirically by the following:

$$Gr = \frac{g \alpha (T_w - T_c) \delta^3}{v^2}$$  

(20)

For a horizontal interlayer (bottom heated), when $Gr > 2430$, natural convection appears and it becomes more intensive when $Gr$ is bigger. The horizontal interlayer finite space natural convection heat transfer can be calculated empirically by the following:

$$Nu = \begin{cases} 0.212(Gr \delta Pr)^{\frac{1}{2}}, & 1.0 \times 10^4 \leq Gr \delta \leq 4.6 \times 10^5 \\ 0.061(Gr \delta Pr)^{\frac{1}{2}}, & Gr \delta > 4.6 \times 10^5 \end{cases}$$  

(21)

The convective heat transfer coefficient can be calculated by $Nu$:

$$h_\delta = \frac{Nu \lambda}{\delta}$$  

(22)

With water temperature $T_w$ and temperature at the inside top of the tank $T_{\text{top}}$, the heat transfer from water to the top of the tank is as follows:

### TABLE 9 Parameters of ST in simulation

| Parameter               | Value | Unit | Source  |
|-------------------------|-------|------|---------|
| Shape                   | Cylinder |    | Selected |
| Inner tank diameter     | 0.8   | m    | Selected |
| Inner tank height       | 1.7   | m    | Selected |
| Shell thickness ($L_1$) | 2     | mm   | Selected |
| Insulation thickness ($L_2$) | 50 | mm   | Selected |
| Inner tank thickness ($L_3$) | 2.5 | mm   | Selected |
| $\kappa_1$ (Stainless St.- AISI 302) | 15.1 | W/m/K | Reference |
| $\kappa_2$ (Rock wool)  | 0.0045 | W/m/K | Reference |
| $\kappa_3$ (Galvanized steel) | 18  | W/m/K | Reference |
| Ambient convection coefficient ($h_{\text{amb}}$) | 10.74 | W/m²/K | Reference |

### TABLE 10 Acceptable criteria in the control method

| Criterion               | Initialization check acceptable | System running check acceptable |
|-------------------------|---------------------------------|----------------------------------|
| Check interval          | /                               | Recent 4 weeks                    |
| Recall                  | >70%                            | >70%                             |

### FIGURE 7 Flowchart of predictive control method for the domestic hot water system

### FIGURE 8 Heat loss model of the storage tank

The convective heat transfer coefficient can be calculated by $Nu$:

$$h_\delta = \frac{Nu \lambda}{\delta}$$  

(22)
Thus, heat loss from top of the tank can be developed as follows:

\[ q_{t,\text{top,in}} = h_\delta (T_w - T_{t,\text{top}}) \]  \hspace{1cm} (23)

By combining Equation (22-24), temperature at the inside top of the tank \( T_{t,\text{top}} \) can be solved:

\[ q_{t,\text{top}} = q_{t,\text{top,in}} = q_{t,\text{top,out}} \]  \hspace{1cm} (25)

Then the heat loss from the top of the tank can be calculated using:

\[ \Phi_{t,\text{top}} = q_{t,\text{top}} * A_{\text{top}} \]  \hspace{1cm} (26)

Wall temperature of an infinitesimal unit of height \( \delta \) at the inside of the tank can be calculated as:

\[ A_{\text{top}} = A_{\text{bottom}} \]  \hspace{1cm} (27)

\[ T_{t,\text{upperside},\delta} = T_w - \frac{q_{t,\text{top}}}{h_\delta} \]  \hspace{1cm} (28)

With \( T_{t,\text{upperside},\delta} \), the heat loss from the wall of this infinitesimal unit can be calculated as:

\[ q_{t,\text{upperside},\delta} = \frac{T_{t,\text{upperside},\delta} - T_{\text{amb}}}{\frac{1}{h_{\text{amb}}} + \frac{L_{t,1}}{k_{t,1}} + \frac{L_{t,2}}{k_{t,2}} + \frac{L_{t,3}}{k_{t,3}}} \]  \hspace{1cm} (29)

Heat loss from the upper side of the tank can be calculated as:

\[ q_{t,\text{upperside}} = \int_0^H q_{t,\text{upperside},\delta} d\delta \]  \hspace{1cm} (30)

\[ A_{\text{upperside}} = \pi * d_{\text{inner}} * H_w \]  \hspace{1cm} (31)

\[ \Phi_{t,\text{upperside}} = q_{t,\text{upperside}} * A_{\text{upperside}} \]  \hspace{1cm} (32)

### 3.3.2 Electric heating device

Electric boiler heats water to the water supply temperature 60°C and electric heater in ST keeps the temperature of the water 60°C in the ST. The efficiencies of the direct-flow electric water boiler and electric heater are both 95%. The temperature of the income water from the municipal water pipe is environmental temperature.

Heat of boiling water from \( t \) to \( t_0 \) in water boiler is as follows:

\[ q_w,\text{boiler,} = V_i, t * \rho_w * c_w * (T_w - T_{\text{amb}}) \]  \hspace{1cm} (33)

Energy consumption of water boiler is as follows:

\[ Q_w,\text{boiler} = \int_{t_0}^{t} q_w,\text{boiler,} dt \]  \hspace{1cm} (34)

\[ E_{\text{boiler}} = \frac{Q_w,\text{boiler}}{\eta} \]  \hspace{1cm} (35)

| Parameters   | Value       | Unit        | Source   |
|--------------|-------------|-------------|----------|
| \( v_a \)    | 16.95*10^6 | m^2/s       | Given    |
| \( \rho_a \) | 1.128       | kg/m^3      | Given    |
| \( c_a \)    | 1005        | J/kg/K      | Given    |
| \( \alpha_v \) | 3.676*10^{-3} | m/s/K  | Given    |
| \( Pr_a \)   | 0.696       | -           | Given    |
| \( T_{\text{ref}} \) | 40        | °C          | Selected |
| \( \rho_w \) | 998         | kg/m^3      | Given    |
| \( c_w \)    | 4.186       | J/kg/K      | Given    |

**TABLE 11** Physical properties of air and water used in simulation

**FIGURE 9** Flowchart of simulation process
where $V_{t,t}$ is the volume of income water at time $t$, $\rho_w$ is the density of the water and $c_w$ is the specific heat capacity of water. The parameters of air and water properties are shown in Table 11.

Energy consumption from $t$ to $t_0$ in the ST is as follows:

$$E_{heater} = \frac{\int_{t_0}^{t} \Phi_{tank} dt}{\eta}$$  \hspace{1cm} (36)
Total energy consumption of the hot water system is as follows:

\[ E_{\text{total}} = E_{\text{boiler}} + E_{\text{heater}} \]  

### 3.4 Simulation process

Many software (Matlab, Simulink, C++, Trnsys, EnergyPlus, etc.) have been used for system simulation of DHW system before. Among which, Matlab has its advantages for convenience of coding and a variety of built-in functions. LIBSVM 3.22 was used to build the predictive model and the model of DHW system was programmed in Matlab R2017b. The flowchart of simulation process is shown in Figure 9. The simulation was conducted with a time step of one hour from September 1 to November 30 (2184 hours in total). When \( t \) reached 2185, simulation ended and summed all the energy consumption value together to get the total energy consumption during simulation.

The volume of income hot water can be calculated as:

\[ V_{i,t} = \begin{cases} V_{d,t} - V_{r,t-1}, & V_{d,t} > V_{r,t-1} \\ 0, & V_{d,t} \leq V_{r,t-1} \end{cases} \]  

where the volume of remaining water in the ST can be calculated as:

\[ V_{r,t} = V_{d,t-1} - V_{i,t-1} - V_{c,t-1} \]

### RESULTS AND ANALYSIS

#### 4.1 Prediction performance

The dynamic recall change of each user during the simulation was shown in Figure 10. Recall was calculated using the latest four-week data, and the minimum limitation is 70%. The fluctuation for some users, such as users 5 and 7, was severe. Another main reason for the users’ sudden shower incidents not being predicted by the algorithm was the limited number of shower incidents. Compared with other users, users 5 and 7 had fewer shower incidents. Because the total number was small, a single incorrect prediction would influence the recall by up to 10%-15%. For other users with more shower incidents, the effect was smaller.

The hot water demand was mostly guaranteed because the predictive result obtained a high recall. The overall recall and accuracy of each user were listed in Table 12. The relationship between the recall and accuracy was repellent, that is, it was difficult to achieve the best recall and accuracy at the same time.

Compared to the training results in the model validation process mentioned in Section 2, the weighted average change in the test process varied depending on the users. Users 1, 2, and 3 were relatively stable. The changes in users 4, 5, 6, and 7 were relatively large, but still met the requirement of the prediction and did not influence the DHW system during the simulation.

#### 4.2 Comparative prediction performance

The future hot water demand was predicted using the predicted occupants’ shower status \( y^{'}_{u,t+1} \) and Equation (10) mentioned in 2.5. The prediction and actual results were shown in Figure 11.

The traditional method was to predict the aggregated hot water demand directly. This was also achieved using SVM
for regression in this study. All parameters in the SVM had been cross validated to find the best combination. The RMSE was used as the evaluation criteria for an aggregated prediction, and the main parameters for the SVM regression were as listed in Table 13. The prediction results were restricted to be larger than zero.

The aggregated prediction results and the actual results were shown in Figure 12.

From Figure 11, the results from a separate prediction were larger than actual results in most cases, whereas in Figure 12, for an aggregated prediction, the prediction results remained at the same level as the actual results. The RMSE was used to evaluate the prediction precision. The RMSE for a separate prediction was 77.63, while for that for an aggregated prediction was 58.65.

However, once the prediction result was less than the actual result, it would lead to a failure of the hot water supply and lower the hot water supply assurance (HWSA). The relationship between the RSME and HWSA was similar to the relationship between the accuracy and recall, which was described in section 2.3, and the best performance of these two evaluation criteria could not be achieved at the same time. To analyze the prediction results from the two methods, an error distribution analysis was conducted, the results of which were shown in Figure 13.

During the test process, there were 403 out of 2184 time steps that the user actually had hot water consumption and only four (0.99%) hot water failures occurred in a separate prediction. The HWSA was 99.01%. However, for the aggregated prediction, there were 313 (77.67%) occurrences, and the HWSA was 22.33%. In addition, there were 112 (27.79%) time steps in which the hot water shortage is larger than 100 L. This would result in user complaints because the hot water shortage would be so large that more than one user would be affected.

The differences in the RMSE and error distribution were due to the difference in the aims of the two prediction methods during the training process. A separate prediction used the weighted average of the accuracy and recall as the performance evaluation (see section 2.3), in which the recall could improve the HWSA by predicting all possible shower incidents. In comparison, the aggregated prediction used only the RMSE (which had the same meaning as the accuracy) as the performance evaluation. The model was trained to allow the error to be distributed on both sides of zero such that the RMSE could be the least. However, this result was not very useful or applicable in a predictive control strategy because of the low HWSA. In addition, in the calculation of domestic hot water demand, the introduction of the standard deviation, seasonal coefficient, and adjustment coefficient could improve the HWSA. These coefficients increased the hot water demand to a certain degree, which resulted in a lower RMSE.

In conclusion, the method of a separate prediction proposed in this paper had been proven as a feasible way to predict future hot water demand, with a similar RMSE performance as an aggregated prediction, but with a much better performance to ensure the HWSA; it was thus more suitable for predictive control in a domestic hot water system. A comparison between these two prediction methods was another proof that only taking the RMSE as the evaluation criteria may lead to “fake low-RMSE” results, which were not practically useful.

4.3 | Hot water supply assurance

The proposed hot water system was designed for a 24-hours hot water supply, and thus, the HWSA was an important evaluation for the comfort of the occupants. The amount of the remaining hot water was used to evaluate the HWSA during the

FIGURE 13  Error distribution of two prediction methods. (A: separate prediction; B: aggregated prediction)
The remaining hot water in the ST was well controlled by the new control method by reducing the maximum amount of remaining hot water to below 200 L. This would lead to less heat loss from the ST. As for the HWSA, there are only four time steps where the remaining hot water was less than zero, which meant the hot water supply was insufficient for the occupant demand. During the simulation, the total number of shower incidents was 441. The portion of insufficient hot water supply was less than 1%. All insufficient hot water supply incidents and insufficient values were listed in Table 14.

The influence of an insufficient hot water supply could be reduced by informing the user ahead of time and giving feedback to the control system. For example, if the remaining hot water in the ST was below 100 L and the predictive

### TABLE 14 Insufficient water supply incidents

| Time stamp | Insufficient value (L) |
|------------|------------------------|
| 1136       | 25.49                  |
| 1260       | 4.76                   |
| 1522       | 2.17                   |
| 1719       | 11.64                  |
consumption per shower for a user was 120 L, the system would alert the user in advance and let the user choose whether to continue to shower or give the system some time to prepare more hot water, in order to reduce the user dissatisfaction.

4.4 Comparative energy consumption

A traditional control method for a centralized hot water supply system with a storage tank works by filling the storage tank every day at midnight and preparing for users to consume during the daytime. It is widely used in a similar hot water system and is therefore selected as a comparison method. Using a traditional control method, the same system was set to fill up the ST to the maximum at 2:00 to 5:00, during which most of the occupants do not shower. The incoming hot water, remaining hot water, and energy consumption were calculated. The remaining hot water in the ST from the traditional control method was shown in Figure 15. This control method suited the occupants’ hot water demand well with 100% HWSA. However, using a traditional control method, a large quantity of hot water was stored in the ST to be consumed later by the occupants. This would cause more heat loss in the ST. The heat loss and energy consumption of these two methods were calculated and compared in Table 15.

The results showed that, by the adoption of the new predictive control method using the occupants’ behavior prediction, heat loss from the ST was reduced by up to 33.12%. Heat loss in the ST occupied 16.03% of the total energy consumption of the whole system when using a traditional control method. This portion reduced to 11.32% when using the new predictive control method, reducing the total energy consumption by 5.31%.

\[ E_{\text{electricity}} (\text{kWh}) = \frac{E_{\text{total}} (\text{MJ})}{3.6} \] (40)

To show the effect of this new control method more intuitively, the electricity consumption of both scenarios was also calculated using Equation (40). As the simulation runs using data from a 3 month period, the total energy consumption for seven users were reduced by 154 kWh, from 2898 to 2744 kWh. This proved that this new control method had the ability to save up to 7.33 kWh electricity per month per user compared to a traditional control method in the proposed simulation.

The cost of the implementation of such a new control method was low as the main change was the installation of a WCU for shower data collection. The cost of one WCU for one shower tap was approximately 25 to 30 US dollars including implementation. For one apartment, 2 or 3 people shared a shower tap. The electricity price in Shanghai is 0.091 US dollars/kWh. The payback period of this new control method is therefore 1.04-1.87 years.

5 CONCLUSION

In the present study, a novel predictive control method based on occupant behavior prediction using a data mining algorithm was proposed for a domestic hot water system. This method generates future hot water demand based on a prediction of each occupant’s shower behavior. A control strategy for a domestic hot water system based on the prediction results was raised and simulated in software using shower data of real occupants. The heat transfer model of a hot water storage tank was built to show the relationship between heat loss and the remaining hot water in the hot water storage tank. Finally, a comparative analysis of the prediction results and the energy consumption was carried out. The main conclusions from this paper are as follows:

1. This paper proves that SVM can be used for occupant shower behavior prediction and can obtain good results. Weighted averages of the accuracy and recall are considered as the evaluation criteria in training the predictive model and are weighted as 0.3 and 0.7, respectively. Recall is considered a primary evaluation to judge whether the predictive model is convincible for further utilization. The results show that the recall reaches above 70% for all users during the validation and testing.

2. A novel method of generating future hot water demand based on a prediction of occupant shower behaviors was proposed and validated in this paper. Each occupant’s shower habit, statistic effect, seasonal effect, and systematic optimization effect were considered as influencing factors. This well-established prediction model can achieve 100% hot water supply assurance during validation. An easily implemented hot water supply strategy was also proposed based on the prediction results.
3. This novel prediction method gives acceptable results during the test process, with a root mean square error (RMSE) of 77.63 and a hot water supply assurance (HWSA) of 99.01%. A prediction of the aggregated hot water demand was also conducted in comparison, resulting in an RMSE of 58.65 and an HWSA of 22.33%. The results show that a separate prediction is more suitable for predictive control. The main reason for the differences in the prediction results is the different choice of the evaluation criteria. This also proves that taking only the accuracy as the evaluation criteria is not always suitable, particularly for predictive control.

4. A dynamic simulation of the proposed predictive control method for a domestic hot water system was carried out using real data. Hot water supply assurance, heat loss, and total energy consumption were calculated for the two control methods. The novel control method obtains a good energy saving compared with the traditional approach, reducing heat loss from the hot water storage tank by 33.12% and reducing 5.31% of the total energy consumption, with little change in the system. Each user can save up to 7.33 kWh electricity per month through the introduction of this new control method into the DHW system.

**NOMENCLATURE**

GLOSSARY

- **a**: accuracy
- **r**: recall, %
- **V**: volume, L
- **Pr**: Prandtl number
- **Gr**: Grashof number
- **Nu**: Nusselt number
- **h**: heat transfer coefficient, W/m²/K
- **k**: conductivity coefficient, W/m/K
- **T**: temperature
- **A**: area, m²
- **Q**: heat, J
- **q**: heat flux, W/m²
- **E**: energy consumption (kWh or MJ)
- **DHW**: domestic hot water
- **ST**: hot water storage tank
- **TP**: true positive
- **TN**: true negative
- **FP**: false positive
- **FN**: false negative
- **WCU**: water control units
- **HWSA**: hot water supply assurance

SUBSCRIPTS

- **u**: user number
- **t**: time stamp
- **d**: demand
- **r**: remain
- **amb**: ambient
- **c**: consume
- **i**: income
- **ref**: reference
- **w**: water

GREEK SYMBOLS

- **η**: electric heating efficiency
- **ψ**: actual shower status, 0 or 1
- **ψ’**: predicted shower status, 0 or 1
- **ψ’’**: revised predicted shower status, 0 or 1
- **ν**: dynamic viscosity, m²/s
- **ρ**: density, kg/m³
- **c**: capacity, J/kg/K
- **α_v**: volume expansion coefficient, per K
- **Φ**: heat flux, W

**ORCID**

Lijun Yu [https://orcid.org/0000-0002-0751-3189](https://orcid.org/0000-0002-0751-3189)

**REFERENCES**

1. Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. *Energy and Buildings*. 2008;40: 394-398.
2. An J, Yan D, Deng G, Yu R. Survey and performance analysis of centralized domestic hot water system in China. *Energy and Buildings*. 2016;133:321-334.
3. Naspolini HF, Rüther R. Assessing the technical and economic viability of low-cost domestic solar hot water systems (DSHWS) in low-income residential dwellings in Brazil. *Renewable Energy*. 2012;48:92-99.
4. Bertrand A, Mastrucci A, Schüler N, Aggoune R, Maréchal F. Characterisation of domestic hot water end-uses for integrated urban thermal energy assessment and optimisation. *Appl Energy*. 2017;186:152-166.
5. Amirirad A, Kumar R, Fung AS. Performance characterization of an indoor air source heat pump water heater for residential applications in Canada. *Int J Energy Res*. 2018;42:1316-1327.
6. Hegazy AA, Diab MR. Performance of an improved design for storage-type domestic electrical water-heaters. *Appl Energy*. 2001;71:287-306.
7. Knudsen MD, Petersen S. Model predictive control for demand response of domestic hot water preparation in ultra-low temperature district heating systems. *Energy and Buildings*. 2017;146:55-64.
8. Széai I, Aldabbagh LBY, Atikol U, Hacisevki H. Performance improvement by using dual heaters in a storage-type domestic electric water-heater. *Appl Energy*. 2005;81:291-305.
9. Hepbasli A, Kalinci Y. A review of heat pump water heating systems. *Renew Sustain Energy Rev*. 2009;13:1211-1229.
21. Raza MQ, Khosravi A. A review on artificial intelligence based predictive building control via machine learning. Appl Energy. 2018;218:199-216.

22. Amasyali K, El-Gohary E, Melikoğlu M, Albostan A. Modelling and load demand forecasting techniques for smart grid and buildings. Energy and Buildings. 2016;108:697-708.

23. Kazmi H, Mehmood F, Lodeweyckx S, Driessen J. Gigawatt-hour scale savings on a budget of zero: Deep reinforcement learning based optimal control of hot water systems. Energy. 2018;144:159-168.

24. Bruni G, Cordiner S, Mulone V, Rocco V, Spagnolo F. A study on the energy management in domestic micro-grids based on Model Predictive Control strategies. Energy Convers Manage. 2015;102:50-58.

25. Fuentes E, Arce L, Salom J. A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis. Renew Sustain Energy Rev. 2018;81:1530-1547.

26. Masiello JA, Parker DS. Factors influencing water heating energy use and peak demand in a large scale residential monitoring study. Residential Buildings: Technologies, Design, Performance Analysis, and Building Industry Trends. 1992:1:157-170.

27. George D, Pearre NS, Swan LG. High resolution measured domestic hot water consumption of Canadian homes. Energy and Buildings. 2015;109:304-315.

28. Ahmed K, Pysly P, Kurniński J. Monthly domestic hot water profiles for energy calculation in Finnish apartment buildings. Energy and Buildings. 2015;97:77-85.

29. Gerin O, Bleys B, De Cuypere K. Seasonal variation of hot and cold water consumption in apartment buildings. CIBW062 Symposium. 2014:1-9.

30. Vapnik V. The nature of statistical learning theory. New York, NY: Springer Science & Business Media; 2013.

31. Ma W, Fang S, Liu G, Zhou R. Modeling of district load forecasting for distributed energy system. Appl Energy. 2017;204:181-205.

32. Sobti S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: A review. Energy Convers Manage. 2018;156:459-497.

33. Mohammadi K, Shamsirband S, Tong CW, Arif M, Petković D, Ch S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manage. 2015;92:162-171.

34. Tan P-N, Steinbach M, Kumar V. Introduction to data mining. Boston: Addison Wesley; 2005.

35. Kiyan M, Bingöl E, Melikoğlu M, Albostan A. Modelling and simulation of a hybrid solar heating system for greenhouse applications using Matlab/Simulink. Energy Convers Manage. 2013;72:147-155.

36. Bergman TL, Incropera FP, DeWitt DP, Lavine AS. Fundamentals of Heat and Mass Transfer. New Jersey: John Wiley & Sons; 2011.

37. Wanjiru EM, Sichilalu SM, Xia X. Model predictive control of heat pump water heater-instantaneous shower powered with integrated renewable-grid energy systems. Appl Energy. 2017;204:1333-1346.

38. SiLang Corporation. http://www.silang-cn.com/productDetail.asp?id=232. Accessed May 15, 2018.

39. Armstrong PM, Uapipatanakul M, Thompson I, Ager D, McCulloch M. Effects of the nature of statistical learning theory on the development of SAHP for domestic hot water. Renew Sustain Energy Rev. 2018;83:124-155.