


Article

Enhancing Residential Water End Use Pattern Recognition Accuracy Using Self-Organizing Maps and K-Means Clustering Techniques: *Autoflow v3.1*

Ao Yang ^{1,*}, Hong Zhang ^{1,2}, Rodney A. Stewart ^{1,2}  and Khoi Nguyen ^{1,2}

¹ School of Engineering and Built Environment, Griffith University, Queensland 4222, Australia; hong.zhang@griffith.edu.au (H.Z.); r.stewart@griffith.edu.au (R.A.S.); k.nguyen@griffith.edu.au (K.N.)

² Cities Research Institute, Griffith University, Queensland 4222, Australia

* Correspondence: ao.yang@griffithuni.edu.au; Tel.: +61-7-5552-8778

Received: 30 July 2018; Accepted: 6 September 2018; Published: 10 September 2018



Abstract: The aim of residential water end-use studies is to disaggregate water consumption into different water end-use categories (i.e., shower, toilet, etc.). The authors previously developed a beta application software (i.e., *Autoflow v2.1*) that provides an intelligent platform to autonomously categorize residential water consumption data and generate management analysis reports. However, the *Autoflow v2.1* software water end use event recognition accuracy achieved was between 75 to 90%, which leaves room for improvement. In the present study, a new module augmented to the existing procedure improved flow disaggregation accuracy, which resulted in *Autoflow v3.1*. The new module applied self-organizing maps (SOM) and K-means clustering algorithms for undertaking an initial pre-grouping of water end-use events before the existing pattern recognition procedures were applied (i.e., ANN, HMM, etc.) For validation, a dataset consisting of over 100,000 events from 252 homes in Australia were employed to verify accuracy improvements derived from augmenting the new hybrid SOM and K-means algorithm techniques into the existing *Autoflow v2.1* software. The water end use event categorization accuracy ranged from 86 to 94.2% for the enhanced model (*Autoflow v3.1*), which was a 1.7 to 9% improvement on event categorization.

Keywords: water end-use; K-means clustering; self-organizing maps; *Autoflow*; water consumption

1. Introduction

The urban population is expected to increase by 54 to 66% by 2050 [1]. Population growth and higher per capita water usage presents a challenge to future water demand [2]. Moreover, the number of mega-cities with more than 10 million inhabitants will grow to over 40 by 2030 [3], which will create significant water demand in certain locations [4]. The combination of growing urban development and climate change will have a significant impact on water security [5]. Therefore, it is important to conduct systems modeling to examine whether water supply sources are adequate to meet future demand scenarios [6].

Recently, a number of studies have demonstrated that water end-use analysis can be utilized to inform policy and practices for urban water systems [7]. Specifically, water end-use studies can provide essential water consumption information regarding when, where, how, and why residential users consume water in the home [8,9]. However, the accurate disaggregation of large amounts of water flow data into different end-use categories is challenging. Therefore, this is the focal point of this water end-use study.

Over the last few years, numerous smart water metering programs have been applied to a number of cities worldwide to monitor residential water consumption [10]. There are presently two main

approaches being applied for undertaking the disaggregation of residential smart meter water data: (1) decision tree algorithm-applied by Trace Wizard[®] [11] and Identiflow[®] [12] and (2) machine learning algorithms-applied by HydroSense [13], BuntBrainForEndUses[®] [14], REU2016 [15], and *Autoflow* [5].

Autoflow employs a hybrid combination of pattern recognition algorithms (i.e., Hidden Markov Model and Dynamic Time Warping algorithm) and data mining techniques (i.e., event probability analysis) to learn distinct flow signature patterns for each water end-use category [5,7,16,17]. This approach overcomes the drawbacks of other existing models to address the water end-use classification problem. For Trace Wizard[®], this software is strongly dependent on the experience of the analyst and two weeks of flow data requires two hours of work by the analyst [7]. In addition, the accuracy of Trace Wizard[®] is reduced when more than two events occur concurrently [18]. Identiflow[®] has a higher reported accuracy than Trace Wizard[®] [7], but its accuracy heavily depends on the physical features for each fixture/appliance that was input into the software [10]. The drawback of HydroSense is that it needs a large number of obtrusive pressure sensors (i.e., 33 sensors in a single household) connected to each water device [13] to be effective.

The first version of *Autoflow v1.1* provided recognition accuracy greater than 90% for a set of mechanised appliances (e.g., clothes washer, dishwasher, toilet, etc.) while behaviorally influenced water end-use categories (e.g., shower, bathtub, and irrigation) was below 70%. It achieved an average overall pattern recognition accuracy across all categories of 85% [17]. *Autoflow v2.1* achieved an overall water end use pattern recognition accuracy of 90% and better handled combined event disaggregation [5]. While *Autoflow v2.1* accuracy is getting close to the targeted 95% recognition accuracy required for a commercial application, there are some deficiencies with this version related to the initial clustering of discrete events into the most likely water end-use category before more detailed pattern recognition algorithms are applied.

To address this deficiency in *Autoflow v2.1*, the present study developed a hybrid model combining Self-Organizing Maps (SOM) and K-means clustering techniques in order to improve the existing data clustering process. SOM has been applied in water resources problems [19], water quality analysis [20], and pattern analysis [21]. Moreover, SOM and K-means are combined and applied in market segmentation [22] and water distribution systems [23]. In this paper, water end-use pattern recognition accuracy improvements related to the addition of the newly developed SOM+K-means computational procedure has been evaluated (i.e., *Autoflow v3.1* compared to *Autoflow v2.1*). The overall clustering process was divided into two main stages: (1) weight matrix estimation using SOM based on the water end-use events features and (2) clustering all of the unclassified water end-use events according to the weight matrix.

The following sections of the paper outline the study background, the enhanced classification model, the model validation and discussion, and finally a description on how this research is an integral part of the urban water industry. The Background (Section 2) briefly introduced the importance of conducting water end-use studies, the difference between the conventional water metering process and advanced process, the existing models for autonomous water end-use classification, an overview of the existing versions of *Autoflow*, and persuasion for the latest version of *Autoflow*. Section 3 provides a detailed application of SOM and K-means, which has been applied as a pre-grouping process in the proposed version *Autoflow v3.1* to improve the recognition accuracy of single water end-use event analysis. Afterward, two independent verification processes, limitations, and future research directions are provided in the model validation section (Section 4). Lastly, Section 5 summarizes the *Autoflow* software and outlines the importance of its widespread implementation for the urban water industry.

2. Background

2.1. Conventional Water Metering Process

Water end-use studies have been increasingly used by water businesses globally in the past decade to understand water demand characteristics and trends for their customers. Moreover, water end use

data underpins future demand forecasts for a city and the design of effective demand management strategies. Reliable accounting and management of water demand is now expected from customers in the current digital and sustainability conscious era. Smart metering and other sensor technologies now provides the opportunity to collect big data on water usage within each segment of an intelligent water network. However, useful analysis of this big data especially what is required to disaggregate residential water consumption into discrete water end use categories previously required extensive human resources. This resource intensive water end use data analysis process needed to be automated to make it feasible for viable widespread implementation. The authors addressed this problem by designing an autonomous water end use analysis software tool (*Autoflow*) that will enable future smart meters to be equipped with this firmware or cloud linked software that will provide water end use information to each and every residential customer as well as urban water professionals. Such near real-time water end use information will significantly enhance the management of future urban water resources.

Water consumption data is recorded manually on a monthly or even half-yearly basis. Current water billing systems typically use only two to four data points to describe a whole year's worth of water consumption data of a household. Kilolitre is often the current resolution of a report, which are counted by conventional water meters that do not have the ability to record when (i.e., the time of day) and where (e.g., tap, toilet) the consumption takes place [24]. The current water metering system has many limitations since it does not provide real-time or continuous water flow data and does not provide flow data of a sufficient resolution to allow for water end-use event disaggregation. Continuous water flow data (i.e., minute to hourly) is essential for real-time network modeling and optimized water infrastructure modeling and engineering [25–27]. Water end use data is useful for water demand management [28], enhanced water infrastructure planning [25], managing water peak demand [29], and understanding water-energy linkages [30,31].

2.2. Existing Autonomous Water End Use Classification Models

With the advent of smart meters providing greater volumes of high-resolution water flow data, researchers have begun to develop advanced models to autonomously or semi-autonomously classify collected real-time and continuous water consumption data into different water end-use categories. There are two main categories of models including descriptive models and predictive models [10]. Descriptive models analyze the observed water consumption behaviors of users and these models can focus on consumption patterns, according to the resolution of water flow data (e.g., [25,26,32,33]). Predictive models focus on solving the problem of water demand (e.g., [5,28,34,35]). Nguyen [16] provides a comprehensive critique of the strengths and weaknesses of each of the existing models. This critique suggested that the utilization of smart water meters at the property boundary, which is supported by firmware that could autonomously disaggregate water flow data into discrete water end use categories by applying machine learning techniques, was the most feasible approach to realize a vision of end-use data being delivered to customers and utilities.

2.3. Water End-Use Classification Process Using Autoflow

Many applications have been developed to automate this costly process including HydroSense, Identiflow, and Trace Wizard, but these existing software have many limitations preventing their widespread implementation. *Autoflow* was developed to address the deficiencies evident in these applications. *Autoflow* software is now being utilized by five water utilities in Australia. *Autoflow* enabled the accurate and reliable use event disaggregation of water consumption data, which is useful for understanding when, where, and how various categories of residential consumers use water in their homes. The primary functions of the software include, but are not limited to, collecting and transferring water consumption data through a smart water metering system and analyzing the data and producing a wide range of water end use reports, which can be accessed by relevant stakeholders (i.e., consumers, water utilities, etc.). Anonymous summary reports can also be compiled for government departments

to assist them in formulating a range of policies (e.g., incentives, outdoor water restrictions, behavioral marketing, etc.).

The *Autoflow* offers a robust pattern recognition procedure through the hybrid combination of existing popular techniques. *Autoflow v1.1* software combined Hidden Markov Model (HMM), Artificial Neural Networks (ANN), and the Dynamic Time Warping (DTW) algorithm (Figure 1). The first version was suitable for mechanical water end-use category analysis such as clothes washers, dishwashers, and toilets [7,16]. *Autoflow v2.1* included an additional search process using the Dynamic Time Warping algorithm. This enhanced version marginally increased the classification accuracy of mechanized end use categories [5].

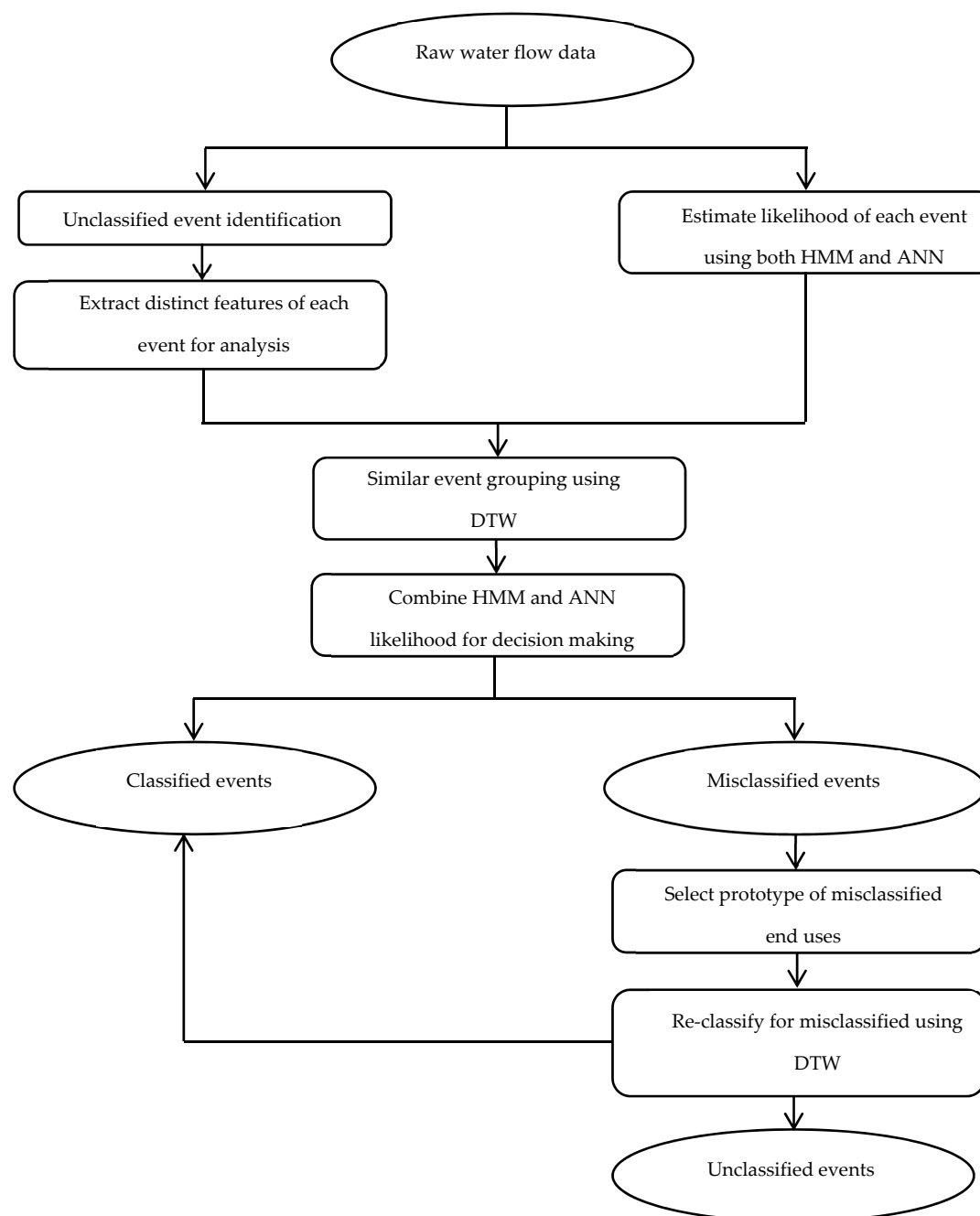


Figure 1. Overall event classifications procedures.

The goals of the design of the *Autoflow* processing algorithms is to provide both accuracy and efficiency. The first version of *Autoflow* (i.e., *v1.1*) achieved an average overall pattern recognition accuracy across all categories of 85%. The overall accuracy of *Autoflow v2.1* was more than 90% and had a 4.9% and 8.0% improvement for single event and combined event categorization, respectively. In terms of efficiency, the *Autoflow v2.1* can recognize patterns in almost 100 events in one second, which is considerably faster than an expert using Trace Wizard. *Autoflow v2.1* also provided an illustrative user interface offering a user-defined water consumption web page to view daily, weekly, and monthly consumption tables as well as charts on consumers' water demand across major end-use categories (e.g., clothes washer, shower, irrigation, etc.). The system also provides customer alerts regarding whether they have leaks rather than waiting for the present slow feedback process from the traditional metering process [5,7,16].

Given that mechanized end use categories such as clothes washer, dishwasher, and toilet usually consist of similar and repeated patterns, one of the initial *Autoflow* analysis modules completed a grouping process where similar patterns were initially classified as belonging together before subsequent detailed pattern recognition procedures were completed. *Autoflow v1.1* and *v2.1* adopted DTW as the main technique to achieve this task. However, recent investigations indicated that DTW had some limitations since it grouped similar patterns based on a pre-determined distance threshold between the unclassified events and the reference event. Two events would be considered similar if their distances to a third event (i.e., reference event) were close. However, this distance value is merely a dimensionless number that represents the shape difference between two given patterns and, in some cases, two completely different events could have the same distance as the reference event. As a result, DTW could assign those two events into the same initial category even though their water end use patterns are completely different. While this incorrect classification was often discovered in subsequent analysis, it increased the data processing effort and resulted in higher rates of erroneous classification.

2.4. Overview of Applied Techniques

Apart from DTW, many other popular clustering techniques have been established and widely applied including SOM [36], K-means clustering [37], and Artificial Neural Network (ANN) [38] to address various complex pattern matching problems such as hand writing, speech recognition, fingerprint recognition, surface water level, and seabed liquefaction predictions [39,40]. By investigating many of the above mentioned models, it was determined that the SOM and K-means clustering were highly suitable approaches for clustering similar water end use events together as required in this study.

The SOM is a type of ANN with an unsupervised learning process [36]. This approach attempts to replicate the function of the brain when it receives visual and memory stimulation and is useful for efficiently identifying common patterns and clustering them together.

The SOM network consists of an input layer and an output layer (Kohonen layer). The output layer is designed as a two-dimensional arrangement of neurons that have an ability to project a high dimensional input to a low-dimensional grid by maintaining the topological order. In this layer, each neuron represents a cluster. In addition, the input layer of nodes is connected to the output nodes and own a weight vector respectively. Euclidean distance and cosine distance are two major criteria between the weight vector of output nodes and input nodes.

The SOM algorithm has the advantage of producing intuitive results [41], but it has two limitations related to its static architecture and capabilities to characterize hierarchical relations within the data [42]. The K-means algorithm is a well-established statistical technique for unsupervised iterative data cluster analysis. SOM and K-means clustering were the utilized techniques for enhancing *Autoflow v2.1*, which is described in the next section.

3. Enhanced Water End-Use Classification Model

Autoflow v2.1 adopted the DTW technique for grouping similar end-use events. However, DTW was heavily dependent on a threshold value, which defined the similarity between two samples. For *Autoflow v3.1*, DTW was replaced by the hybrid SOM+K-means clustering analysis module in order to improve residential water end-use pattern recognition accuracy. The following sections describe how these techniques were applied to this pattern recognition problem.

3.1. SOM for Water End-Use Pattern Recognition

The formulated six-step procedure detailed below applied the SOM technique for the initial clustering of residential water end use events into different water end-use categories.

Step 1-Input Layer Settings

As shown in Figure 2, the input layer are given vector E_l ($l = 1, 2, \dots, P$) of the water end-use events, which are then normalized as \hat{E}_l ($l = 1, 2, \dots, P$), where P is the number of unclassified events. In the present study, each water event has three features including volume, duration, and maximum flowrate. Therefore, an additional subscript x is used to present the features and input events are presented by \hat{E}_{xl} ($x = 1, 2, 3; l = 1, 2, \dots, P$).

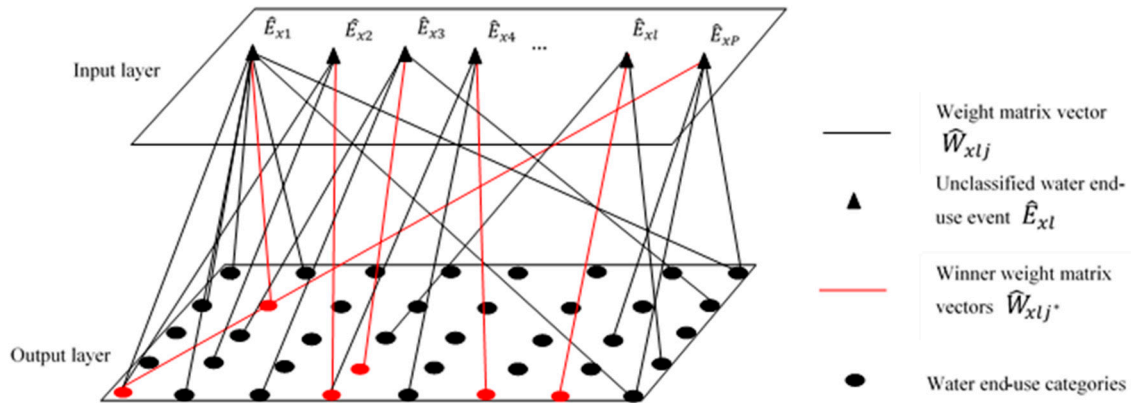


Figure 2. Water end-use network in SOM.

Step 2-Output Layer Settings

A 2D network of m rows and n columns of water end-use categories is set up in the output layer, which is shown in Figure 2. The total number of categories T is written as Equation (1).

$$(T = m \times n) \quad (1)$$

Step 3-Assign Matrix Vectors

Assign a random initial value to the weight matrix vector. Since the network is fully connected, each input node is connected to each category. Therefore, the weight matrix vector between the input event and categories is W_{lj} ($l = 1, 2, \dots, P; j = 1, 2, \dots, T$), which can be normalized as Equation (2).

$$\hat{W}_{lj} = \frac{W_{lj}}{\|W_{lj}\|} \quad (2)$$

In the output layer, the weight matrix vector is \hat{W}_{xlj} ($x = 1, 2, 3; l = 1, 2, \dots, P; j = 1, 2, \dots, T$).

Step 4-Determine Winner Category and Define the Winner Neighbourhood Area

In this step, the distance between input vectors and all categories in the output layer is calculated. The iterative method is used to determine the winner category j^* , according to the minimum distance from the input nodes to the output nodes. The weight matrix vector of the winner category is \hat{W}_{xlj^*} , which can be normalized as Equation (3).

$$\|\hat{E} - \hat{W}_{xlj^*}\| = \min_{j \in \{1, 2, \dots, T\}} \{\|\hat{E} - \hat{W}_{xlj}\|\} \quad (3)$$

Then, the initial neighbor area of the winner category is defined as $N_{j^*}(t = 0)$ and it is a larger zone, which decreases over the iteration steps. This is shown in Figure 3. The maximum iteration is I_{max} .

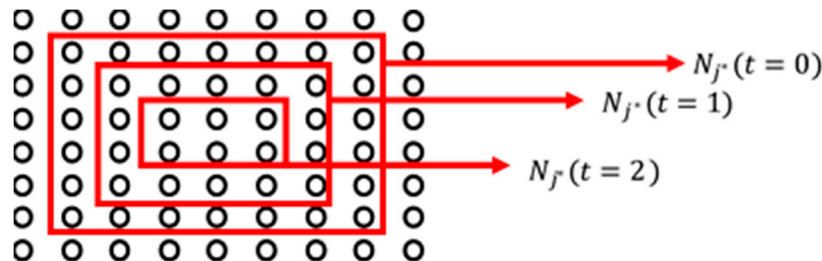


Figure 3. Change of the neighboring area with iterations.

In addition, the neighbor area will update per Equation (4).

$$N_{j^*}(t) = N_{j^*}(t = 0) \times e^{-\left(\frac{t \times \log N_{j^*}(t = 0)}{I_{max}}\right)} \quad (4)$$

Step 5-Adjust Winner Weight Matrix Vectors and Update Neighboring Area

According to the SOM network, the winner category has an activation value of ONE while other neurons have activations of ZERO, which is seen in Equation (5).

$$z_j(t + 1) = \begin{cases} 1 & j = j^* \\ 0 & j \neq j^* \end{cases} \quad (5)$$

However, only the winner neuron can adjust the weight matrix vector. The adjustment function is given by Equation (6).

$$\hat{W}_{xlj}(t + 1) = \hat{W}_{xlj}(t) + \eta(t) [\hat{E}_{xl} - \hat{W}_{xlj}] \quad x = 1, 2, 3; j = 1, 2, \dots, T; l = 1, 2, \dots, P \quad (6)$$

where $0 < \eta(t) \leq 1$ and it is a function, which decreases with the number of iterations to ensure the convergence of the algorithm. In this present paper, the learning rate is given by Equation (7).

$$\eta_{j^*}(t) = \eta_{j^*}(t = 0) \times e^{-\left(\frac{t}{I_{max}}\right)} \quad (7)$$

Step 6-End Judgement

When the learning rate $\eta(t) \leq \eta_{min}$ or $\eta(t) = 0$, the training process is finished. When the stopping condition is not satisfied, the algorithm will go back to Step 2 of the procedure until the maximum iteration number I_{max} is reached.

3.2. K-Means for End-Use Study

The specific process of the K-means algorithm is shown below.

Step 1-Initial Centroids Selection

Randomly select k samples as the initial cluster centroid C_k from all winner weight matrix vectors \hat{W}_{xlj^*} . In this paper, the winner weight matrix vectors are the input data in the K-means algorithm

in order to cluster the matrix first and get the final weight matrix as the initial centroids in the next grouping process. For input data with three features, each centroid can be represented as Equation (8).

$$C_k = [c_{k1}, c_{k2}, c_{k3}] \quad (8)$$

Step 2-Assign Matrix Vectors

In this step, each vector will be grouped according to the similarity (distance) d_k to the centroids. The distance to each centroid is calculated for each input vector and can be calculated by using Equation (9). Afterward, the object is assigned to the closest (minimum distance) centroid, which can be normalized as Equation (10).

$$d_k(w, c) = \sqrt{\sum_{i=1}^m (w_{ki} - c_{ki})^2} \quad m = 3 \quad (9)$$

$$w(j) \in \text{group}(\min\{d\}) \quad j = 1, 2, \dots, T \quad (10)$$

where w_{ki} is an element of the weight matrix vector.

Step 3-Update the Centroid

When all matrix vectors have been grouped, the centroids position is updated. The new position of it is based on the new means of each cluster. The position of group k is written as Equation (11).

$$P_k = \frac{\sum_{i=1}^{N_k} w_{ki}}{N_k} \quad (11)$$

where N_k is the number of elements belonging to group k .

Step 4-Iteration

Repeat Step 2 and Step 3 of this procedure until the centroid of the clusters no longer changes so far. At the end of the process, the distance between the clusters is maximal and the distance between the centroids and their data is minimal. After this step, the clustering result is a new set of matrix vectors \hat{W}_{xlj^*} $x = 1, 2, 3; j = 1, 2, \dots, k; l = 1, 2, \dots, P$.

Step 5-2nd K-Means Clustering for Unclassified Events

Input the previous result as the initial centroids. Then input the all unclassified events and group those using specific K-means. The process is the same as in Step 2. Then, the unclassified will be grouped into k groups.

3.3. Hybrid SOM-K-Means Model for Residential Water End-Use Pattern Recognition

Application of the algorithms for the data clustering process described above was crucial before the operation of the previously developed pattern matching algorithms and associated analysis processes. The structure of the SOM network is robust but simple and can handle outliers. In the situation where the mesh of neurons is larger, the number of clusters is larger. Reducing the number of clusters in the SOM algorithm can reduce the final number of water end-use groups. However, the quality or the accuracy of clustering will be decreased. In addition, the clustering result of the K-means algorithm is greatly affected by the selection of the initial cluster centroid. If the initial centers are not properly selected, the result may fall into a local optimal solution rather than a global optimal result. For these reasons, this section presents the model operationalization process for this specific problem including how these two techniques were combined and how they can improve the clustering process in *Autoflow v3.1*.

The existing water end-use model predominately applied only computing techniques [43]. The entire procedure for the hybrid technique includes two stages. First, each water event is presented by extracted features and grouped through SOM (reaching a certain number of cycles). In addition, a set of weight vectors are calculated. Second, these vectors that serve as the initial cluster centroids were applied in K-means clustering. Figure 1 presents the algorithms implemented in the previous

version of *Autoflow v2.1*. In this version, the initial pre-grouping process was undertaken by the Dynamic Time Warping algorithm before the decision-making stage. Moreover, this pre-grouping process is a major component of the whole pattern recognition analysis process and the accuracy of this stage can affect the final water category recognition accuracy directly. Hence, it is necessary to seek a method with high accuracy and efficiency in the pre-grouping process to improve the overall accuracy of *Autoflow*. In the developed version *Autoflow v3.1*, the hybrid technique combining SOM and K-means has replaced the DTW algorithm in the initial pre-grouping process, which is shown in Figure 4. The specific subset of the entire procedure for the newly included SOM-K-means clustering process is detailed in Figure 5. Readers are referred to Nguyen [5,7,16,17] for the detailed analytical procedures for the other parts of the procedure described in Figure 4.

The advantages of this hybrid technique include: (i) maintaining the characteristics of the self-organizing of the SOM algorithm, (ii) maintains the high efficiency of K-means clustering, (iii) reduces the convergence time in SOM, and (iv) reduces the impact of the initial centroid selection in the K-means algorithm.

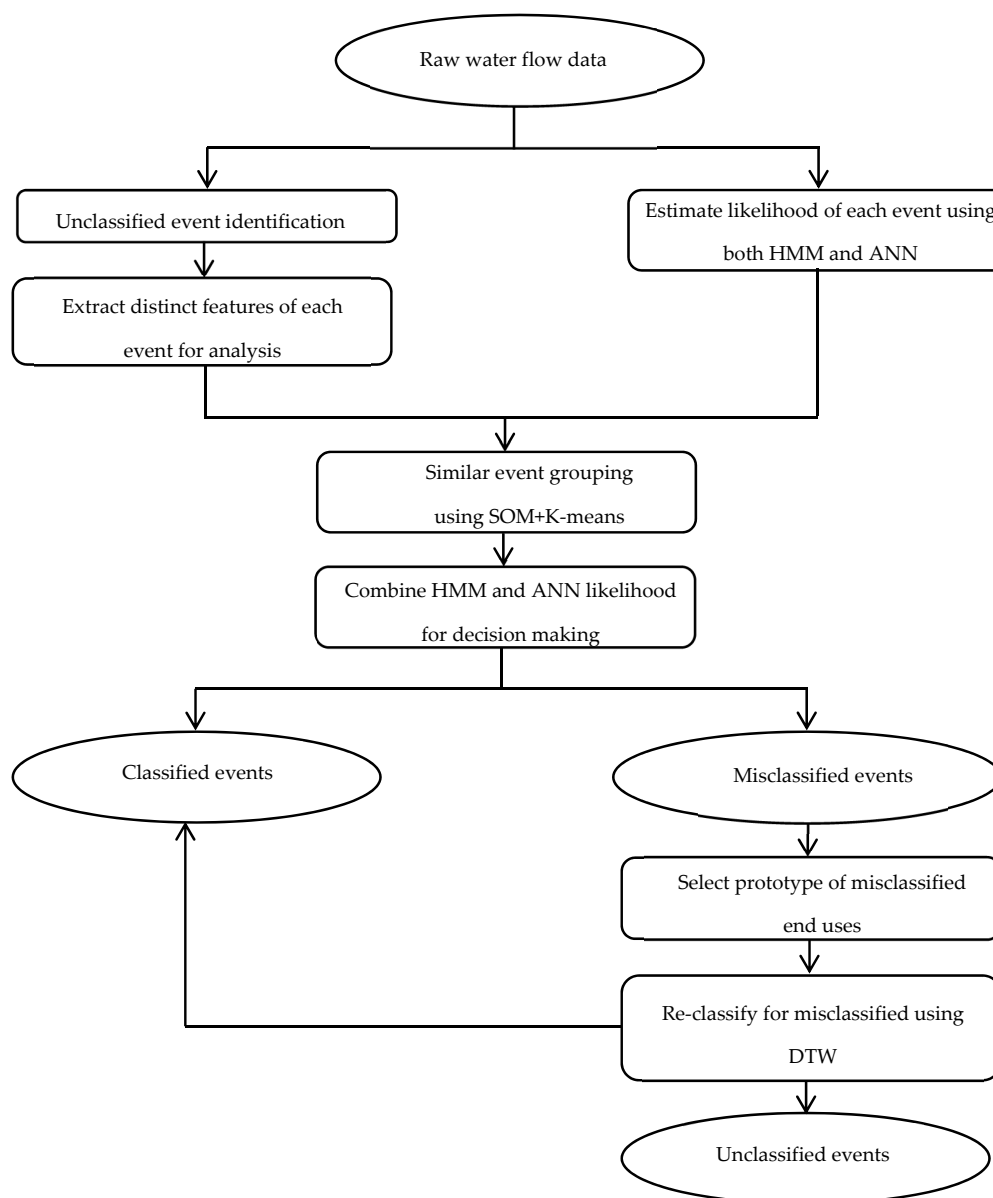


Figure 4. Overall water end-use event classification procedures with SOM + K-means.

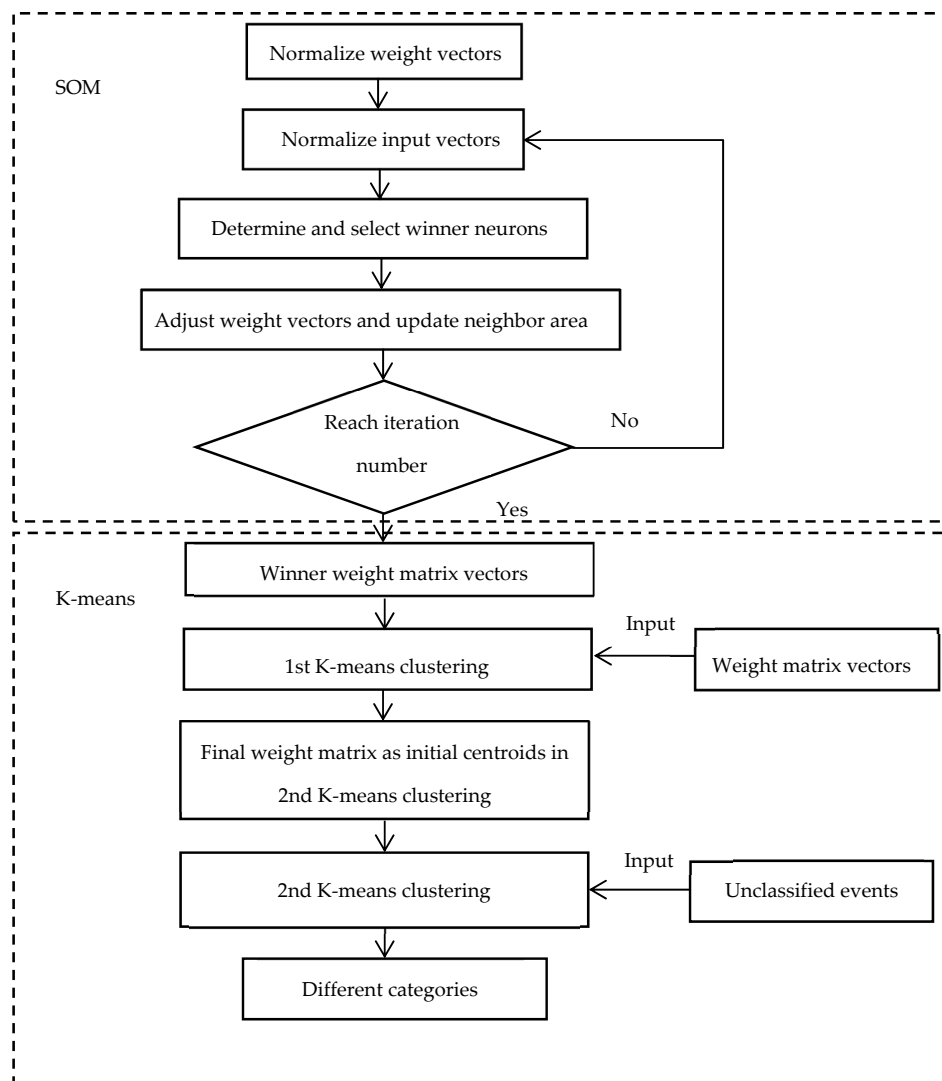


Figure 5. SOM-K-means clustering analysis procedure included in *Autoflow v3.1*.

4. Model Validation

As described below, two stages of model testing and verification were undertaken in this study.

- (i). Model testing utilized the herein recommended hybrid combination of SOM + K-means on a total of 9200 random event samples.
- (ii). Event model verification was conducted by applying the *Autoflow v3.1* software tool to 30 independent homes from Australia and comparing against using original *Autoflow v2.1*.

4.1. Independent Testing Process

A total of 9200 samples were randomly selected from 104,271 disaggregated samples of eight different categories for testing. These selected samples were mixed together and the SOM + K-means analysis module was adopted to partition them into eight different groups where each group was expected to consist of similar events. The correctly classified water end use events are shown in Figure 6. The model accuracy is determined by dividing the number of correctly grouped events for each category by the number of events for testing that category. For example, for the clothes washer event category, the number of testing samples was 1500 and the number of clothes washer events correctly grouped was 1398, which resulted in an accuracy of 93.2%. Table 1 details that the *Autoflow v3.1* water end use event recognition accuracy ranged from 86% for the bathtub to

94.2% for the dishwasher. It is interesting to note that in Table 1, the hybrid model improved the accuracy of both the more predictable mechanized water end-use event categories (i.e., clothes washer, dishwasher, evaporative cooler, and toilet) as well as those more variable categories heavily dependent on human behaviors (i.e., tap, bathtub, irrigation, and shower). Overall, the newly developed hybrid method that has been augmented in *Autoflow v3.1* has achieved a higher overall pattern recognition accuracy when compared to the prior utilized DTW technique applied in *Autoflow v2.1* for this similar clustering function.

Classified	Clothes washer	1,398	49	21	15	21	16	3	0
	Dishwasher	60	1,413	15	47	16	15	4	2
	Toilet	20	9	1,392	31	23	70	0	0
	Evaporative cooler	4	3	35	1329	70	52	0	0
	Shower	11	8	21	20	1,341	13	1	1
	Tap	4	6	5	26	17	1,326	4	8
	Bathtub	1	7	2	30	4	1	86	2
	Irrigation	2	5	9	2	8	7	2	87
		Actual							

Figure 6. Correctly classified water end use events using the technique.

Table 1. Water end-use event classification accuracy comparison (*v3.1* versus *v2.1*).

Category	No. of Total Samples	No. of Samples for Testing	DTW Accuracy (v2.1) (%)	Hybrid Technique Accuracy (v3.1) (%)	Improvement (%)
Clothes washer ^a	11,168	1500	86.2	93.2	7
Dishwasher ^a	3458	1500	91.8	94.2	2.4
Toilet ^a	13,852	1500	89.8	92.8	3
Evaporative cooler ^a	23,542	1500	86.9	88.6	1.7
Shower	5743	1500	80.4	89.4	9
Tap	45,959	1500	86.2	88.4	2.2
Bathtub	217	100	80	86	6
Irrigation	332	100	80	87	7
All categories	104,271	9200			

Notes: ^a Mechanical water end-use event categories.

4.2. Autoflow Recognition Accuracy Using Independent Homes

To determine the true accuracy of the *Autoflow v3.1* model, 30 independent home datasets were collected. These 30 homes based in Australia were independent of the original training and testing datasets and each water end-use event was certified by home owners.

Table 2 compares the average accuracy achieved for each category between the *Autoflow v2.1* developed using DTW, HMM, and ANN and the *Autoflow v3.1* (i.e., SOM + K-means + ANN + HMM) for the independent sample of 30 homes. Water end use event pattern recognition accuracy for *v3.1* ranged from 75.1% for irrigation to 90.1% for the clothes washer. Irrigation event recognition accuracy should be treated with caution and is not considered reliable for small samples (i.e., N = 30 homes) since this usage category has a smaller number of samples and is sporadic. As can be seen from this table, an average accuracy of more than 85% has been achieved for mechanical categories with the maximum of 90.1% for the clothes washer and the minimum of 85.5% for the toilet. In comparison with the

Autoflow v2.1, the new *v3.1* model has provided greater overall accuracy especially for the mechanized water end use categories. Notable version comparison improvements include 87.6% compared to 82.1% for the dishwasher, 86.8% compared to 83.4% for the evaporative air cooler, and 85.5% compared to 83.1% for the toilet. It is noticeable that the newly included SOM-K-means clustering algorithm has been more effective for mechanized water end use event categorization.

Table 2. Independent data verification comparison (*v3.1* versus *v2.1*).

Category	Applied Model	Number of Home			Average Accuracy (%)
		Accuracy > 90%	80–90%	70–80%	
Clothes washer ^a	1 ^b	6	20	4	90.1
	2 ^c	4	10	16	83.5
Dishwasher ^a	1	13	11	6	87.6
	2	9	12	9	82.1
Toilet ^a	1	5	20	5	85.5
	2	6	13	11	83.1
Evaporative cooler ^a	1	10	16	4	86.8
	2	8	14	8	83.4
Shower	1	15	8	7	86.2
	2	14	5	11	86.6
Tap	1	20	3	7	88.5
	2	18	6	6	88.3
Bathtub	1	0	12	18	86.8
	2	0	10	20	83.4
Irrigation	1	0	0	30	75.1 ^d
	2	0	0	30	75.6 ^d

Notes: ^a Mechanical water end-use events, ^b 1-*Autoflow v3.1* and ^c 2-*Autoflow v2.1*, ^d Irrigation water end use events are fewer and sporadic. Therefore, it is difficult to quantify accuracy in a small sample.

Additionally, regarding water end use pattern recognition accuracy improvements, the processing time to analyze the 30 independent residential homes water end-use events was determined, which is shown in Figure 7. Figure 7 illustrates the processing time comparison between *Autoflow v3.1* and previous version *Autoflow v2.1*. *Autoflow v3.1* has reduced the flow data end-use disaggregation time compared with the previous version. As evident in Figure 7, processing time per event reduces further with *Autoflow v3.1* when there is a greater number of unclassified events. In terms of the number of events, although *Autoflow v2.1* had a good efficiency, it still cannot achieve a processing speed of 100 events per second. In the developed *Autoflow v3.1*, a pattern recognition speed greater than 100 events per second could be achieved. Overall, a 20% improvement in processing time was achieved for *Autoflow v3.1* compared to *Autoflow v2.1*.

Testing results have indicated that the replacement of DTW by the combination of SOM with K-means has improved recognition accuracy in *Autoflow v3.1* especially for the mechanized water end use categories. The hybrid technique was less effective for behaviorally influenced water end use categories (e.g., shower) since there is a much greater variety in the patterns of these events. Thereby, this makes clustering more challenging.

However, *Autoflow* still can be improved in future research. The developed analytical approach was reliable for categorizing most end-use categories with the exception of irrigation, tap, and shower. In order to significantly improve recognition accuracy of these end-use event categories, which heavily depend on human behavior, further research on how to unobtrusively incorporate local contextual data (i.e., residential household descriptive and behavioral information) into the *Autoflow* analysis process is required. To achieve this, the development of self-learning algorithms for the incorporation into the final analytical phase needs to be completed. Collecting and analyzing each event feature of a much larger sample of events (e.g., different regions, dwelling type, etc.) will be the focus of future research. The future developed version of *Autoflow* that includes self-learning functionality should

complete a more accurate disaggregation of water end-use data for new residential households having some different water using fixtures.

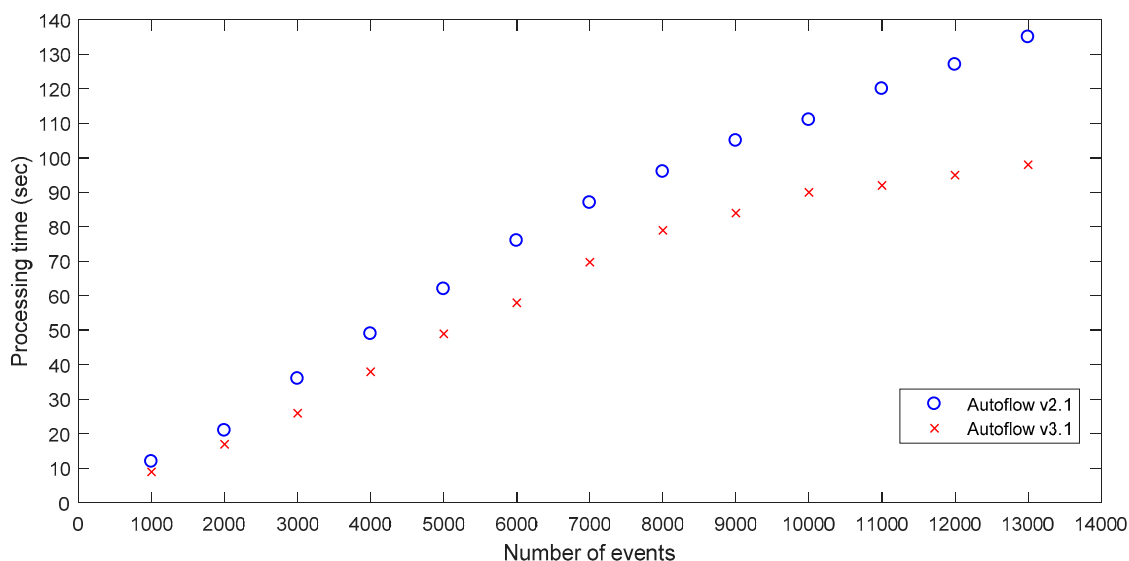


Figure 7. Computational processing time of *Autoflow* (v3.1 versus v2.1).

5. Conclusions

The *Autoflow* software, which is an integrated water management system, employs smart water metering and a series of intelligent algorithms to automate the disaggregation of high-resolution residential water flow data into discrete water end-use events. *Autoflow v3.1* presented in this paper includes a new analysis module to the prior version that groups similar events using a hybrid combination of customized SOM and K-means algorithm. The verification process demonstrated that the enhanced version increased the recognition accuracy. Moreover, the study showed that the inclusion of the SOM+K-means method provides better clustering of mechanized water end-use events such as clothes washer, dishwasher, evaporative cooler, and toilet. *Autoflow v3.1* has significant implications for the urban water industry.

Autoflow is a commercial software, which is currently being used by various water utilities in Australia (e.g., Yarra Valley Water, City West Water, and South-east Water). This application was developed using MATLAB and it can be used both as a desktop and as a web application. At this stage, the main features of the software include, but are not limited to, collecting and transferring water consumption data through a smart water metering system as well as, analyzing the data and producing a wide range of reports that can be accessed by various users (i.e., consumers, water utilities, and government organizations). However, work is underway to embed this novel software as firmware within smart meters and transfer processed water end-use data to consumers via an application on their phone or computer.

Autoflow has significant implications for consumers and water businesses, government departments, the metering and software industry, and the wider urban water industry. For consumers, *Autoflow* provides them with an easily accessible interface, which allows consumers to monitor where and when they are consuming water in their household. By consumers having water end use data readily available to them on a daily basis, they will be have better understanding on their water consumption and be able to target efficiency efforts (i.e., understand that their shower consumption is high). For the water utility, compiled anonymized city-level water end use data can assist them with a range of operational functions including water demand management, distribution network infrastructure planning, post-meter leakage management, and customer engagement. Government departments and agencies charged with urban water policy can use water end use data for designing effective water efficiency policies (i.e., rebates, restrictions, etc.). The water metering industry are

currently developing advanced digital meters that can record and transmit high resolution water usage data (i.e., 0.01 L/1 s). Such metering technology facilitates widespread autonomous water end use analysis. *Autoflow* software could be embedded into future smart meters firmware or alternatively housed in a server for cloud-based processing. Based on these arguments, there is a strong need for further research on developing intelligent water consumption analysis algorithms and associated software such as *Autoflow*. The creation of such software poses significant benefits for infrastructure planners, water demand managers, architects and developers, and policy advisors who seek to better understand water consumption patterns.

Author Contributions: Conceptualization, H.Z. and R.S. Methodology, A.Y. and K.N. Software, A.Y. Validation, A.Y. and K.N. Formal Analysis, A.Y. Writing—Original Draft Preparation, A.Y. Writing—Review & Editing, H.Z. and R.S.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Gerland, P.; Raftery, A.E.; Ševčíková, H.; Li, N.; Gu, D.; Spoorenberg, T.; Bay, G. World population stabilization unlikely this century. *Science* **2014**, *346*, 234–237. [[PubMed](#)]
- Curmi, E.; Richards, K.; Fenner, R.; Allwood, J.M.; Kopec, G.M.; Bajželj, B. An integrated representation of the services provided by global water resources. *J. Environ. Manag.* **2013**, *129*, 456–462. [[CrossRef](#)] [[PubMed](#)]
- UNDESA. *World Urbanization Prospects: The 2014 Revision-Highlights*; United Nations Department of Economic and Social Affairs Population Division: New York, NY, USA, 2014.
- Cosgrove, C.E.; Cosgrove, W.J. *The United Nations World Water Development Report—N° 4—The Dynamics of Global Water Futures: Driving Forces 2011–2050*; Nations Educational, Scientific and Cultural Organization: Paris, France, 2012.
- Nguyen, K.A.; Stewart, R.A.; Zhang, H.; Jones, C. Intelligent autonomous system for residential water end use classification: Autoflow. *Appl. Soft Comput.* **2015**, *31*, 118–131. [[CrossRef](#)]
- Sahin, O.; Stewart, R.A.; Porter, M.G. Water security through scarcity pricing and reverse osmosis: A system dynamics approach. *J. Clean. Prod.* **2015**, *88*, 160–171.
- Nguyen, K.A.; Zhang, H.; Stewart, R.A. Development of an intelligent model to categorise residential water end use events. *J. Hydro-Environ. Res.* **2013**, *7*, 182–201.
- Stewart, R.A.; Willis, R.; Giurco, D.; Panuwatwanich, K.; Capati, G. Web-based knowledge management system: Linking smart metering to the future of urban water planning. *Aust. Plan.* **2010**, *47*, 66–74. [[CrossRef](#)]
- Willis, R.M.; Stewart, R.A.; Panuwatwanich, K.; Williams, P.R.; Hollingsworth, A.L. Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *J. Environ. Manag.* **2011**, *92*, 1996–2009.
- Cominola, A.; Giuliani, M.; Piga, D.; Castelletti, A.; Rizzoli, A.E. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environ. Model. Softw.* **2015**, *72*, 198–214. [[CrossRef](#)]
- DeOreo, W.B.; Heaney, J.P.; Mayer, P.W. Flow trace analysis to access water use. *J. Am. Water Works Assoc.* **1996**, *88*, 79–90. [[CrossRef](#)]
- Kowalski, M.; Marshallsay, D. A system for improved assessment of domestic water use components. In Proceedings of the II International Conference Efficient Use and Management of Urban Water Supply, Canary Islands, Spain, 2–4 April 2003.
- Froehlich, J.; Larson, E.; Saba, E.; Campbell, T.; Atlas, L.; Fogarty, J.; Patel, S. A longitudinal study of pressure sensing to infer real-world water usage events in the home. In Proceedings of the International Conference on Pervasive Computing, San Francisco, CA, USA, 12–15 June 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 50–69.
- Arregui, F. New software tool for water end-uses studies. In Proceedings of the 8th IWA International Conference on Water Efficiency and Performance Assessment of Water Services, Cincinnati, OH, USA, 21–24 April 2015; pp. 20–24.

15. Vitter, J.S.; Webber, M. Water Event Categorization Using Sub-Metered Water and Coincident Electricity Data. *Water* **2018**, *10*, 714. [[CrossRef](#)]
16. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environ. Model. Softw.* **2013**, *47*, 108–127. [[CrossRef](#)]
17. Nguyen, K.A.; Stewart, R.A.; Zhang, H. An autonomous and intelligent expert system for residential water end-use classification. *Expert Syst. Appl.* **2014**, *41*, 342–356. [[CrossRef](#)]
18. Mayer, P.W.; DeOreo, W.B.; Opitz, E.M.; Kiefer, J.C.; Davis, W.Y.; Dziegielewski, B.; Nelson, J.Q. *Residential End Uses of Water*; American Water Works Association: Denver, CO, USA, 1999.
19. Kalteh, A.M.; Hjorth, P.; Berndtsson, R. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environ. Model. Softw.* **2008**, *23*, 835–845. [[CrossRef](#)]
20. Blokker, E.J.; Furnass, W.R.; Machell, J.; Mounce, S.R.; Schaap, P.G.; Boxall, J.B. Relating water quality and age in drinking water distribution systems using self-organising maps. *Environments* **2016**, *3*, 10. [[CrossRef](#)]
21. Laspidou, C.; Papageorgiou, E.; Kokkinos, K.; Sahu, S.; Gupta, A.; Tassioulas, L. Exploring patterns in water consumption by clustering. *Procedia Eng.* **2015**, *119*, 1439–1446. [[CrossRef](#)]
22. Kuo, R.J.; An, Y.L.; Wang, H.S.; Chung, W.J. Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation. *Expert Syst. Appl.* **2006**, *30*, 313–324. [[CrossRef](#)]
23. Brentan, B.; Meirelles, G.; Luvizotto Jr, E.; Izquierdo, J. Hybrid SOM+ k-Means clustering to improve planning, operation and management in water distribution systems. *Environ. Model. Softw.* **2018**, *106*, 77–88. [[CrossRef](#)]
24. Stewart, R.A.; Willis, R.M.; Panuwatwanich, K.; Sahin, O. Showering behavioural response to alarming visual display monitors: Longitudinal mixed method study. *Behav. Inf. Technol.* **2013**, *32*, 695–711. [[CrossRef](#)]
25. Gurung, T.R.; Stewart, R.A.; Beal, C.D.; Sharma, A.K. Smart meter enabled water end-use demand data: Platform for the enhanced infrastructure planning of contemporary urban water supply networks. *J. Clean. Prod.* **2015**, *87*, 642–654. [[CrossRef](#)]
26. Beal, C.D.; Stewart, R.A.; Fielding, K. A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption. *J. Clean. Prod.* **2013**, *60*, 116–128. [[CrossRef](#)]
27. Savic, D.A.; Morley, M.S.; Khoury, M. Serious gaming for water systems planning and management. *Water* **2016**, *8*, 456. [[CrossRef](#)]
28. Willis, R.M.; Stewart, R.A.; Giurco, D.P.; Talebpour, M.R.; Mousavinejad, A. End use water consumption in households: Impact of socio-demographic factors and efficient devices. *J. Clean. Prod.* **2013**, *60*, 107–115. [[CrossRef](#)]
29. Beal, C.D.; Stewart, R.A. Identifying residential water end uses underpinning peak day and peak hour demand. *J. Water Res. Plan Man.* **2013**, *140*. [[CrossRef](#)]
30. Siems, R.; Sahin, O.; Stewart, R.A. Modelling the impact of energy intensity on the economic and environmental costs of internally plumbed rainwater tanks systems. In Proceedings of the 20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013.
31. Stewart, R.A.; Nguyen, K.A.; Beal, C.; Zhang, H.; Sahin, O.; Bertone, E.; Giurco, D. Integrated intelligent water-energy metering systems and informatics: Visioning a digital multi-utility service provider. *Environ. Model. Softw.* **2018**, *105*, 94–117. [[CrossRef](#)]
32. Loh, M.; Coghlan, P. *Domestic Water Use Study in Perth, Western Australia, 1998–2001*; Water Corporation: Perth, Australia, 2003.
33. SDU. *Urban Water Management Plan 2010*; City of Sacramento: Sacramento, CA, USA, 2011.
34. Polebitski, A.S.; Palmer, R.N. Seasonal residential water demand forecasting for census tracts. *J. Water Res. Plan. Manag.* **2009**, *136*, 27–36. [[CrossRef](#)]
35. Makki, A.A.; Stewart, R.A.; Panuwatwanich, K.; Beal, C. Revealing the determinants of shower water end use consumption: Enabling better targeted urban water conservation strategies. *J. Clean. Prod.* **2013**, *60*, 129–146. [[CrossRef](#)]
36. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [[CrossRef](#)]
37. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 7 January 1966; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.

38. Moon, J.W.; Jung, S.K.; Kim, J.J. Application of ANN (artificial-neural-network) in residential thermal control. In Proceedings of the Eleventh International IBPSA Conference, Glasgow, UK, 27–30 July 2009; pp. 27–30.
39. Sannasiraj, S.A.; Zhang, H.; Babovic, V.; Chan, E.S. Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Adv. Water Resour.* **2004**, *27*, 761–772. [[CrossRef](#)]
40. Zhang, H.; Jeng, D.S.; Cha, D.; Blumenstein, M. Parametric study on the prediction of wave-induced liquefaction using an artificial neural network model. *J. Coast. Res.* **2007**, *50*, 374–378.
41. Dorffner, G. *Limitations of the SOM and the GTM*; Department of Medical Cybernetics and Artificial Intelligence: Vienna, Austria, 2001.
42. Rauber, A.; Merkl, D.; Dittenbach, M. The growing hierarchical self-organizing map: Exploratory analysis of high-dimensional data. *IEEE Trans. Neural Netw.* **2002**, *13*, 1331–1341. [[CrossRef](#)] [[PubMed](#)]
43. Pastor-Jabaloyes, L.; Arregui, F.J.; Cobacho, R. Water end use disaggregation based on soft computing techniques. *Water*. **2018**, *10*, 46. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).