21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015
www.mssanz.org.au/modsim2015

Intelligent model to categorise mechanised water end uses

K.A. Nguyen [a], **R.A. Stewart** [a], **H. Zhang** [a] **and S.H. Chowdhury**

[a] *Griffith School of Engineering, Griffith University*
Email: k.nguyen@griffith.edu.au

Abstract: The current population boom has pushed demand on residential land in most world metropolitan cities to its extreme limit, and also imposed significant pressure on water security. To ease this problem, construction of residential apartment building has been the top priority of the governments. With the rapid advancement in technology, especially in water metering area, the next generation of these buildings is expected to not only offer their residents a comfortable living area but also make the water supply and management become much more effectively controllable. To contribute to this goal, the aim of this study was to develop an autonomous and intelligent system for residential water end-use classification that could interface with customers and water business managers via a user-friendly smartphone and web-based applications. The authors recently developed an intelligent application called *Autoflow*© which served as a prototype tool to solve the complex problem of autonomously categorising residential water consumption data into a registry of single and combined events. This model was developed using data collected in several cities in Australia, and when applied on single residential property, the achieved accuracy ranged from 86% - 96%. The only limitation of *Autoflow*© is that when being deployed overseas, a new water end use dataset must be collected to retrain the model to adapt to the new regions. The ultimate goal of this study is to develop the next *Autoflow*© generation, called *Autoflow*^{AB}, which can be applied in all future apartment buildings to help disaggregate water consumption autonomously into the six main categories including: shower, toilet, tap, clothes washer, dishwasher and evaporative air cooler without relying on any previously collected data for training. The key employed techniques are Dynamic Time Warping (DTW) and Decision Tree. The model has been tested on 150 residential properties in Australia where the accuracy ranges from 87 – 94 %, and also planned to be applied on 30 new 7-story buildings in the Commonwealth Games Village located in Gold Coast Australia.

Keywords: *Water end use event, residential water flow trace disaggregation, water micro-component, Hidden Markov Model, dynamic time warping algorithm*

1. INTRODUCTION

With the advent of advanced water metering, logging and wireless communication technologies, one of the top priorities in major Australia metropolitans is to develop a smart water management system that presents new approach to enhance water security through capturing, analysing, providing real time water consumption data and creating an inter-connection between water utilities and customer (Stewart et al., 2013, Beal et al., 2011). The first version of such a system called *Autoflow©* was developed by Nguyen et al. (2013a, 2013b, 2014, 2015), which have successfully classified the eight main end use categories including shower, faucet, clothes washer, dishwasher, evaporative air cooler, toilet, irrigation and bathtub with a relatively high accuracy of 93%. The overall classification process was undertaken using a series of pattern recognition models developed from a database of nearly 100,000 end use samples collected from different regions in Australia. Several verification processes have indicated that effective classification has been achieved on most of the tested areas whose samples were included in the database; however, a significant accuracy drop was also found when applying the models on properties whose end use patterns were not introduced for model training before. This problem is expected to be aggravated when using the model internationally; therefore, innovation techniques are demanded to overcome the dependence on the existing prototype resources. In that context, this study proposed the development of the next *Autoflow©* generation which can be applied to all residential apartment buildings in the world where the collection of new water end use patterns is not required. The solution is based on the fact that for apartment buildings, the available end-uses are limited to shower, tap, clothes washer, dishwasher, toilet and evaporative air cooler, in which the last four are mechanical dependent categories, whose patterns are usually consistent over the time. Among the remaining two categories, shower and tap (i.e. in this study, bathtub and shower are put in the same category, and named as shower in general), although they are user-dependent which can induce an infinite number of patterns, their volumes are significant different to each other which allows the disaggregation to be undertaken easily using the decision tree method.

2. BACKGROUND

2.1. Existing autonomous water end use classification models

Alongside with the recent rapid advent of communication and smart metering technology, several models have also been developed to automatically classify the collected real-time or near real-time water consumption data into different categories. There are currently three approaches to the water end use classification problem: (1) simple decision tree method based on three physical features of each event, namely volume, duration and flow-rate (e.g. *Trace Wizard* and *Identiflow*); (2) sensor networks on water end use appliances supported by data mining techniques (e.g. *Hydro Sense*); and (3) a hybrid combination of pattern recognition algorithms and data mining techniques to learn distinct flow signature patterns for each end use category to perform the classification process (e.g. *Autoflow©).*

The first approach is resource intensive requiring significant analysis to disaggregate water end use patterns into discrete events accurately (Stewart et al. 2010). The second approach achieves high accuracy and does not require human interaction once the system is operating but requires sensors to be attached to many water use devices in the home, which makes this technique cost-intensive, intrusive (Froehlich et al. 2009, Nguyen et al. 2013b) and can artificially influence water use behaviour. The third approach (see Nguyen et al. 2013a, b; 2014) overcomes the deficiencies of the first two, by only requiring a smart meter installed at the property boundary. This approach uses pattern recognition (i.e. HMM and ANN) coupled with other data mining techniques (i.e. event probability analysis) to automate the end use analysis process. A software tool (i.e. *Autoflow©*) was developed to provide a user-friendly platform to aid this process.

2.2. The next generation of apartment building and water management system

With the ultimate aim of turning *Autoflow©* into a commercialised product that can be used nationally, and further internationally, this study aims to develop a new version, call *Autoflow^{AB}*, to help assign all major end uses available in an apartment building, such as clothes washer, dishwasher evaporative cooler, toilet, tap and shower into appropriate categories without any reliance on the pre-trained model. The basic difference between two versions is that the classic *Autoflow©* software is recommended for Australia market only and can be used on both free-standing properties and apartment buildings as it was developed using data collected all over Australia. In terms of *Autoflow^{AB}*, this software package is especially developed for residential apartment building only which can be deployed anywhere in the world due to its independence on collected data. The working mechanism in *Autoflow^{AB}* is presented in Figure 1, which starts with the disaggregation of all mechanised end uses with the order as clothes washer, dishwasher, evaporative air cooler and toilet using DTW. When events of these categories have been correctly classified, the last step is to assign the remaining end uses such as shower and tap using DTW and Decision Tree method. In the context of this paper, only the classification of the four main mechanised categories is brought into attention, which is the key to the development of *Autoflow*^{AB}

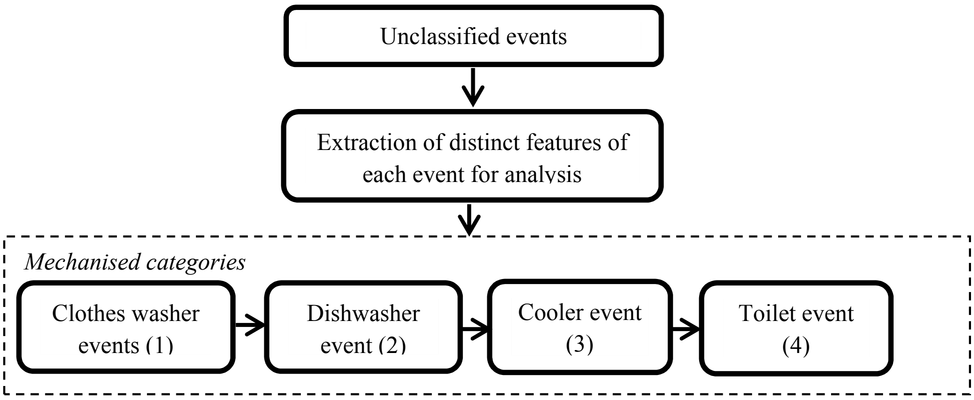


Figure 1. Classification process using new methodology for mechanised end-use categories

3. Clothes washer event classification

It can be seen from Figure 1 that clothes washer event classification is the first step of the overall process. To enable the application of the new method all clothes washer events have to satisfy the features presented Table 1, which have been expanded to make sure that all possible characteristics of clothes washer events in the world will be covered.

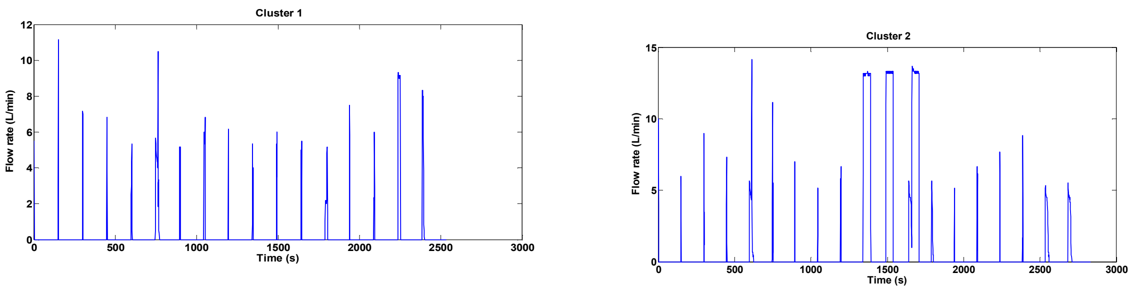
Table 1. Expanded features for clothes washer event classification

Feature No.	Clothes washer features	Minimum value (a)	Maximum Value (b)
1	Volume per wash (litre)	20	250
2	Maximum flow rate (litre/min)	4	35
3	Duration of 1 cycle	10 seconds	10 minutes
4	Volume of 1 cycle (litre)	0.5	120
5	Time interval between 2 cycles	10 seconds	2 hours

From the collected data, the first task is to calculate the basic characteristics of each event, including volume (litre), duration (second), maximum flow rate (litre/min) and most frequent flow rate (litre/min). From these calculate features, the classification process starts with:

Step 1: Select all samples that concurrently satisfy *Feature 2* (i.e. maximum flow rate in between 4 and 35 litres/min), *Feature 3* (i.e. duration in between 10 seconds and 10 minutes) and *Feature 4* (volume in between 0.5 and 120 litres). This step aims to remove all events that are unlikely belonging to clothes washer.

Step 2: With the time associated with each recorded sample, the main task in *Step 2* is to disaggregate all samples in *Step 1* into different clusters in which the time interval of any two consecutive samples does not exceed 2 hours (i.e. based on *Feature 5*, the maximum time interval between 2 consecutive cycles should be less than 2 hours). At the end of this step, several clusters were established, two of which are presented in Figure 2.



a. Cluster 1 b. Cluster 2

Step 3: In *step 3*, all samples within a cluster that have approximate flow rate will be grouped together.

Figure 2. Examples of two random clusters

However, only groups that satisfy *Feature 1a* (i.e. the overall volume of that group must be at least 20 litres) will be retained. Figure 3 display one of the groups extracted from Cluster 2 presented in Figure 2. All others samples that cannot be grouped together will be removed. It should be noted that the utilisation of *Feature 1a* can also help eliminate most of the typical dishwasher events, whose total volumes are usually less than 20 litres. In case there are groups of dishwasher samples that satisfy all requirements in *Step 3*, the identification of these groups will be presented in Step 5.

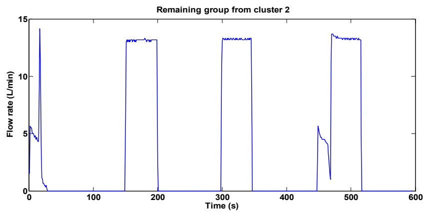
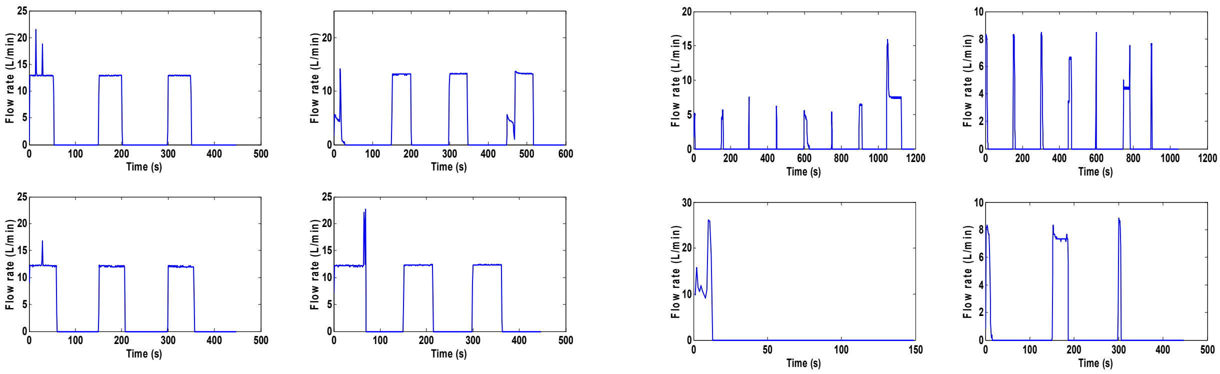


Figure 3. One selected group from Cluster 2

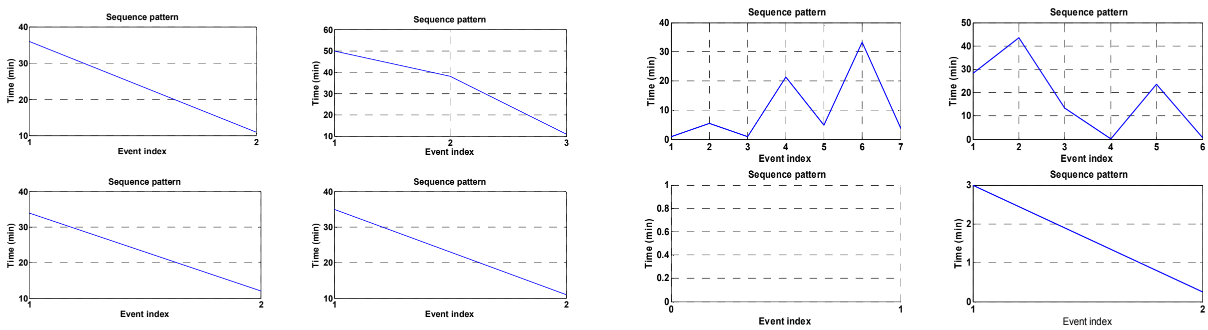
Step 4: At step 4, all groups that contain samples with approximate flow rate from all clusters will be put together in one set



a. Set 1 b. Set 2

Figure 4. Combing groups of similar flow rate

***Step 5**:* As clothes washer events are machine-dependent whose time interval between each cycle follows a certain time pattern, the identification of this type of end use can be done by searching for sets obtained in *Step 4* whose time sequence of events within each group follows a particular trend (i.e. time sequence shows the time interval between consecutive events in a group). This task will be undertaken using DTW to estimate the similarity of time sequence of each group. Figure 5 below shows the time pattern of all groups contained in Set 1 (Figure 5a) and Set 2 (Figure 5b).



a. Time sequences of all groups in Set 1 b. Time sequences of all groups in Set 2

Figure 5. Time interval between any 2 consecutive samples

Step 6: In this step all clothes washer will be identified by adding all samples in the selected set together.

4. DISHWASHER EVENT CLASSIFICATION

Table 2. Basic features of dishwasher event

No.	Dishwasher features	Minimum value	Maximum value
1	Volume per wash (litre)	5	20
2	Flow rate (litre/min)	2.5	10
3	Duration of 1 cycle	10 seconds	5 minutes
4	Volume of 1 cycle (litre)	1	15
5	Maximum time interval between 2 consecutive cycles (minutes)	2	20

The classification of dishwasher will be undertaken after all clothes washer events have been classified. Apart from the uncommon dishwasher events that were isolated in *Step 5* of the previous section, almost all dishwasher events in the world follows the features as presented in Table 2. To deal with this end use category, the same techniques as in clothes washer event classification was applied using the features presented in Table 2. However, it should be noted that as dishwasher is not present in all households, it is important to identify the existence of this end use in the tested property. This task can be achieved at *Step 5* of the classification procedure. At this step, if there is no set that contain groups of similar samples which follow a certain time pattern can be identified, it can be confirmed that there is no dishwasher in the currently tested home.

5. EVAPORATIVE AIR COOLER CLASSIFICATION

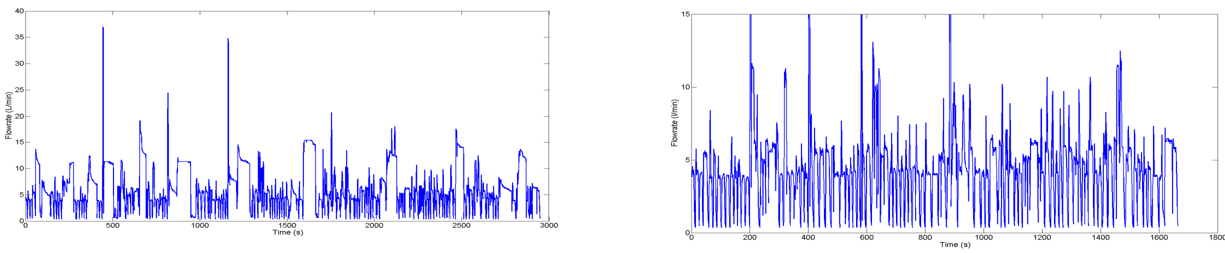
Evaporative air conditioner is an uncommon category that is just present in a few regions in Australia (e.g. Melbourne, Adelaide, etc.). The same method as in clothes washer event classification is applied however, modification has to be made at some points.

The analysis starts with the disaggregation of the remaining samples into clusters where time interval between two consecutive samples in each cluster should not exceed 30 minutes *(Step 1).* This selected threshold is a conservative value as the maximum time interval between two consecutive cycles is about 10 minutes as found in the existing database collected for Melbourne water end use study. In *Step 2*, all events that have similar flow rate, volume and duration will be grouped together as almost 95% of Type 2 evaporative cooler events possess these characteristics. *Step 3* aims to gather all extracted groups in *Step 2* that have similar flow rate together into different sets. At this step, most of evaporative cooler events will be visually identified; however, there is still a possibility that a set of toilet events also exists as this category also have similar flow rate, duration and volume. The disaggregation of these two end uses in case they appear together is undertaken in *step 4* by finding the time sequence pattern of samples in each group. Set containing more groups whose samples follow a certain time pattern will be assigned to evaporative cooler and the other one will be classified as toilet.

6. Toilet event classification

Toilet event classification is the next step in the overall process when clothes washer, dishwasher and evaporative air cooler events have been identified. As toilet is a mechanised category whose volume and pattern are quite determinate, the categorisation of this end use can be undertaken in these four steps and illustrated through the following examples.

Step 1: From the remaining unclassified samples as presented in Figure 6a, search for all events whose volumes are in between 2 and 20 litres (Figure 6b)



a. Remaining samples b. Events that are likely belonging to toilet

Figure 6. Remaining samples for toilet event classification

Step 2: Apply the sample grouping technique based on DTW distance to groups all events that have similar pattern together. At the end of this step, there could be many groups created. Figure 7 below shows that events having similar pattern have been put together in two different groups.

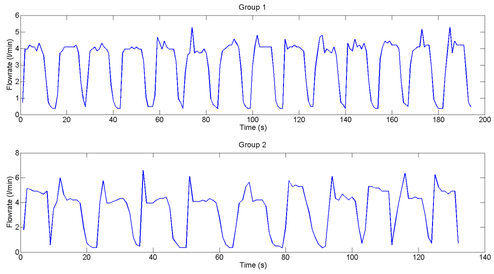


Figure 7. Similar events grouping process

Step 3: Select groups that contain events having similar volumes. In this example, volumes of events in both groups are approximate, which indicates that events in these groups all belong to toilet category.

Step 4: The last step of toilet event classification is to remove any event from other category that has been misplaced into toilet group. This task can be performed by determining the most frequent volume of each group (i.e. the typical volume of toilet event), and any sample whose volume is different from this typical volume by 2 litres will be removed. In this example, no such an event exists in both groups and all events in both groups are eventually classified as toilet.

7. MODEL VERIFICATION

In this study, a verification process has been conducted using data collected from 150 homes from different regions in Australia. To estimate the efficiency of the proposed model on different data patterns, the testing was carried out by using *Autoflow*© and *Autoflow^{AB}* to provide an evaluation on the efficiency of the new model in comparison with the former one. It should be noted that only accuracies obtained from the mechanised categories are compared in this verification process herein.

Table 3. Model testing on 150 homes

Category	Applied model	Number of home			Average accuracy (%)
		Accuracy >90%	85–90%	80-85%	
Clothes washer	**1***	**116 – (79.3%)**	**29– (19.3%)**	**5 – (3.4%)**	**92.3**
	2*	125 – (83.3%)	21 – (14%)	4 – (2.7%)	93.5
Dishwasher	**1**	**42 – (60%)**	**18– (25.7%)**	**10 – (14.3%)**	**89.3**
	2	50 – (71.4%)	13– (18.6%)	7 – (10%)	92.1
Evaporative cooler	**1**	**5– (13.3%)**	**4– (20%)**	**5 – (66.7%)**	**85.1**
	2	8– (53.3%)	3– (20%)	4– (26.7%)	82.6
Toilet	**1**	**80– (53.3%)**	**32– (21.3%)**	**38 – (25.3%)**	**88.3**
	2	90– (53.3%)	15– (20%)	45– (26.7%)	90.6

1- *Autoflow^{AB}* 2- *Autoflow*©

Table 3 compares the average accuracy achieved for each category between the classic *Autoflow*© developed using HMM and ANN and the proposed *Autoflow^{AB}* in the testing on 150 homes. As can be seen from this table, an average accuracy of above 85% has been achieved for all categories when using *Autoflow^{AB}*, with the maximum of 92.3% for clothes washer and minimum of 85.1% for evaporative air conditioner. In comparison with the classic *Autoflow*©, the proposed model has resulted in lower accuracies with 92.3% compared to 93.5% for clothes washer, 89.3% compared to 92.1% for dishwasher, and 88.3% compared to 90.6% for toilet. Most of the misclassified clothes washer events were found at the homes where the flow rate of this category showed a highly fluctuated pattern. The second problem that caused the misclassification of clothes washer is when *Autoflow^{AB}* was applied on homes where toilet events have approximate flow rate and pattern to clothes washer (4 homes in the testing data).

In terms of dishwasher, apart from the similar issues as faced in clothes washer classification, another problem occurred when analysing homes that only used dishwasher once or twice during the testing period of 2 weeks. When inspecting these homes, the proposed model was not able to find the common time sequence from the limited number of dishwasher events, which eventually led to the conclusion that there was no dishwasher event in these tested homes. This problem has resulted in a significant accuracy drop as presented in Table 3; however, as the volume of each dishwasher operation is just in between 20-30 litres, which is less than 0.1% of the overall water consumption in two week period, the impact of this misclassification is minute and can be ignored in these homes.

Toilet is the category whose classification accuracy from *Autoflow^{AB}* are very close to that from the classic *Autoflow©*. There are two main reasons leading to the misclassification of these end uses are: (i) About 12 out of 150 homes whose toilet exposed leak problem which make the pattern of every single toilet event in that home different from each other, and (ii) About 20 homes whose toilet pattern does not have typical pattern, which was the result of using toilet cistern model where the volume of each flush is not fixed. In these homes, toilet events were misclassified as tap by both *Autoflow©* and *Autoflow^{AB}*.

However, the most significant advantage of this new *Autoflow^{AB}* package is at classifying Evaporative Cooler pattern as the achieved accuracy is higher than that using HMM and ANN (85.1% compared to 82.6%). The verification process has pointed out that due to the variant patterns of this complicated end use, which can be similar to tap, toilet, dishwasher and clothes washer, the efficiency of the advanced HMM and ANN model in *Autoflow©* dropped considerably when analysing any home whose evaporative cooler patterns possess similar features to the above mentioned end uses (i.e. this technique performs the classification based on assessing the flow rate pattern and physical characteristics of each event to make decision without inspecting the time pattern).

8. CONCLUSION

The establishment of an integrated water management system, which employs smart water metering, in conjunction with a series of intelligent algorithms to automate the flow trace analysis process, is becoming feasible thanks to the development of *Autoflow©*. The first version of this software tool offers a robust pattern recognition procedure through the hybrid combination of customised HMM and ANN algorithms, which have successfully assigned most of the unclassified samples into appropriate categories with the average accuracies ranging from 80%-94% when testing on over 200 homes in Australia. However, it is expected that when applying the model in international context, the unlimited number of untrained patterns will be the main cause to the significant efficiency reduction of the existing combined HMM and ANN model in *Autoflow©*; therefore it is crucial to develop a dynamic tool that can work with all different data patterns without any reliance on the pre-trained model. This study has been conducted to achieve that goal when developing *Autoflow^{AB}*, the next version especially designed for residential building, that can classify common categories autonomously based on their physical pattern features and working mechanisms. The achieved accuracy of 85-93% when testing the model on a large number of compliscated untrained samples from machine-dependent categories has shown its outperformance in comparison with the existing HMM-ANN combined model and its promising efficiency in dealing with these types of end-uses in the worldwide scenario.

REFERENCES

Beal, C., Stewart, R.A., Huang, T.T., Rey, E. (2011). SEQ residential end use study. *Journal of the Australian Water Association,* 38 (1), 80-84.

Froehlich, J.E., Larson, E., et al. (2009). HydroSense: Infrastructure-mediated single-point sensing of whole-home water activity. *Prococeedings of UbiComp 2009*, Orlando, Florida, USA, 235–244.

Nguyen, K.A., Stewart, R.A. and Zhang H. (2013a). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-Environment Research*, 7(3), 182-201.

Nguyen, K.A., Zhang, H., and Stewart, R.A. (2013b). Intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Journal of Environment Modelling and Software*, 47, 108-12

Nguyen, K.A., Stewart, R.A. and Zhang H. (2014). An autonomous and intelligent expert system for residential water end-use classification. *Journal of Expert Systems with Application*, 41(2), 342-356.

Nguyen, K.A. Stewart, R.A. Zhang, H. Jones, C. (2015) Intelligent autonomous system for residential water end use classification: Autoflow. *Applied Soft Computing*, 31, 118-131

Stewart, R.A., Willis, R.M., Giurco, D., Panuwatwanich, K., and Capati, B. (2010). Web-based knowledge management system: linking smart metering to the future of urban water planning. *Australian Planner*, 47(2), 66-74.

Stewart, R.A., Willis, R.M., Panuwatwanich, K. and Sahin, O. (2013). Showering behavioural response to alarming visual display monitors: longitudinal mixed method study. *Behaviour & Information Technology*, 32(7), 695-711.