International Environmental Modelling and Software Society (iEMSs)
8th International Congress on Environmental Modelling and Software
Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2016-proceedings

Water demand forecasting with AUTOFLOW© using State-Space approach

Khoi Anh Nguyen[1], Oz Sahin[1], Rodney Anthony Stewart[1], Hong Zhang[1]
Griffith University School of Engineering (k.nguyen@griffith.edu.au)

Abstract: The authors have recently developed an intelligent application, *Autoflow*©, which is a powerful tool to autonomously categorise residential water consumption data into a registry of single and combined events. This tool was developed using data collected in several cities in Australia, and when applied on standalone properties, the achieved accuracy ranged from 86% and 96% in terms of number of correctly classified events. Taking advantage of the analysis results from *Autoflow*©, the aim of this study is to propose a short-term water demand forecasting model that not only allows water utilities to predict the overall and peak water demand of up to 24-hour ahead, but also obtain the disaggregated forecast volume of each end-use category. Based on the periodic pattern of the water consumption data, a state-space approach has been adopted whose components including trend, seasonal and external effect from temperature were all modelled as stochastic processes. In this model, Dynamic Harmonic Regression, Kalman Filter and Fixed Interval Smooth algorithms were employed for the estimation of all above-mentioned components. The model has been tested against datasets collected at different time of the year to estimate its efficiency as well as the impact of temperature on water consumption. The verification process has showed that the achieved R^2 for all testings are above 0.9 when undertaking forecast of up to 24-hour ahead, whilst the obtained Mean Absolute Percentage Error (MAPE) of the disaggregated forecast volume of each category most of the time lie below 5%. Further model testing is planned to be applied on 1200 new apartment buildings in the Commonwealth Games Village located in Gold Coast Australia to validate its efficiency. Once finished, the overall *Autoflow*© application will have significant impact in practice that allows the water utility to have detailed plan in moderating water demand along the day, or assist in optimising the pumping activities at the least required energy. Most importantly, the ability to determine and approximate the peak demand will allow utilities to have flexible solutions to solve the over-capacity problem, especially at old metropolitans where the demand far exceeds the supply capacity of the old existing distribution network.

Keywords*:* water end-use, water demand forecasting, state-space model, water resource

1. Introduction

Reliable and accurate forecasts of urban water demand provide the basis for making operational, tactical, and strategic decisions for drinking water utilities (Billings & Jones 2008).Thus, accurate demand forecasting analyses can help water utilities understand spatial and temporal patterns of future water use to optimize system operations, plan for future water purchases or system expansion, or for future revenue and expenditures. There are several mathematical methods for estimating future demand, including; extrapolating historic trends, correlating demand with socio-economic variables, or more detailed simulation modelling (Donkor et al., 2014). Models vary in complexity in terms of the number of variables and the extent to which water users are disaggregated by sector, location, season, or other factors. Models also vary according to the forecast horizon. Long-term forecasting (i.e. decades) is typically more useful for infrastructure and capital planning whereas medium-term forecasting (years or months) is more useful for setting water rates. To perform these forecasting tasks, long historical water consumption data, in addition to socio-economic and weather information of the same period, should be available to ensure a reliable outcome.

In recent years, a new challenge troubling most water utilities in metropolitan cities is the water shortage or experience of low pressure during peak hours, which indicates the incapability of the existing network infrastructure in facilitating the growing water demand. There are two options to ease this issue. The first one is to upgrade the supply infrastructure network based on long-term demand forecasting; however, this choice is costly and sometimes not feasible for the outdated existing network in old cities as the upgrade can significantly affect other socio-demographic plans. Hence, the second option, and also the only economical and feasible solution is to: (i) predict when the water shortage could occur during the day based on collected data of previous few weeks, as this short period of time would effectively reflect the current water consumption trend of the supplied regions, and (ii) reschedule the pumping plan as well as provide instruction to water consumer on how to minimise all expected water supply problems. For example, based on the analysis, the customers can be advised to avoid irrigation from 6 pm to 9 pm, or limit their shower duration to 5 minutes if it occurs anytime in between 5 pm and 7 pm. In terms of environmental management, the achievement of this goal would result in significant reduction in urban water supply and energy consumption during peak hours (i.e. mainly from more effective pumping activities), which will help reduce the amount of green gas released to the environment.

There are existing models (Donkor et al., 2014) that allow very short-term forecasting at an hourly interval to be undertaken, however, their applications are still limited as the outcomes only help utilities in optimising the pumping activities to reduce the over-demand issues during peak periods, and will be ineffective if the current demand exceeds the supplier pumping capacity. In this case, actions are required from both water consumers and water utilities to overcome the issue. To achieve this goal, the authors developed the *Autoflow*©, a smart water monitoring and management system. This new system is capable to capture the water flowrate at a high resolution (i.e. every 5 or 10 seconds); autonomously categorise water consumption data into a registry of different water end-use (e.g. shower, tap, clothes washer, etc.); perform hourly prediction of water consumption; and provide suggestions on how to overcome the pressure loss issue if required. Thus, the *Autoflow*© enhances water security through capturing, analysing, providing real-time water consumption data and creating an interconnectivity between water utilities and customer. The *Autoflow*© (Fiugure 1) allows the disaggregation of collected data into different end-use categories (Nguyen et al., 2013a, 2013b, 2014, 2015), and generates a detailed end-use analysis report. In this context, the main focus of this paper is to develop a short-term hourly demand forecasting model, not only for the overall consumption, but also for each end-use categories, which could significantly enhance the ability of water utilities in making a better decision to solve the network over-capacity issue.

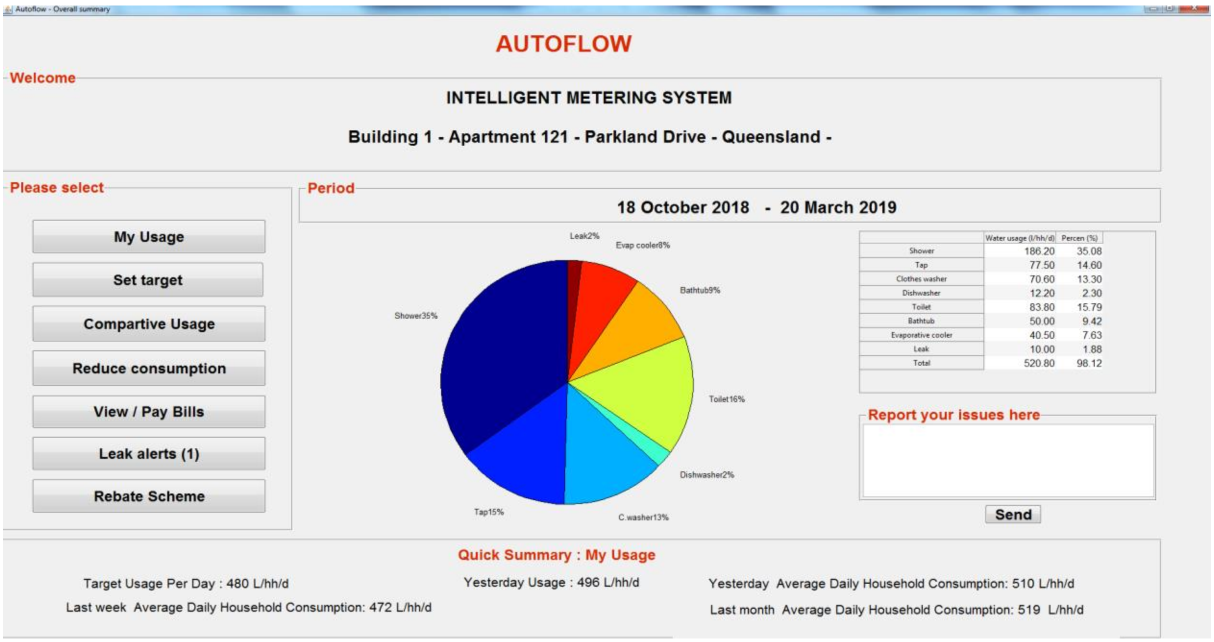

Figure 1 Smart water monitoring and management application (*Autoflow*©)

2. Background

2.1 Autoflow – A smart water monitoring and management system

With the advent of advanced water metering, logging and wireless communication technologies, developing a smart water management system has become one of the top priorities in major metropolitan areas globally as such system would help to enhance water security through capturing, analysing, and disseminating near real-time water consumption data and concise reports to both water utilities and customers. Specifically, once in place, the system enables customers and water utilities to actively monitor, through web-portal interfaces, real-time information about what, when, where and how water was consumed at their meter connection (e.g. 56 litres of water was used during a shower event occurred between 06:55-07:08 on Tuesday 25 May 2014). It also allows individual consumers to log into their user-defined water consumption web page to view their daily, weekly, and monthly consumption summaries, as well as more granular charts displaying their water end-use patterns across major end-use categories (e.g. leaks, clothes washer, dishwasher, tap, toilet, shower and irrigation). The analytical report generated by the new advanced integrated water management system would help utilities identify the water consumption patterns of their various consumer types and assist with a range of urban water planning and management functions (Stewart et al. 2010). The first version of such a system called Autoflow© was developed by Nguyen et al. (2013a, 2013b, 2014, 2015), which have successfully classified 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.

The main output from *Autoflow*© contains a list of descriptive statistical data on each end use category, which includes the average and mean values of volume, duration and flow rate extracted from the classified events. Other useful information from this system is the daily end-use diurnal demand graph. This graph can be automatically created from the repository of classified end-use events, and is highly beneficial to both consumers and water businesses who seek to better understand how residential water consumption is being used, at an end-use level, across various significant days of the year (i.e. average weekday, average weekend day, peak day, average day peak month). This data is particularly useful for water infrastructure planning (e.g. water pipe network augmentation planning) as it informs network modelling engineers of the peak demand flow rates as well as the key end-uses contributing to that peak demand (i.e. shower use combined with clothes washer contributes to morning peak). In this context, this diurnal pattern graph (Figure 2) plays a significant role in determining the volume percentage distribution of each category along the day, which is the key to forecasting the future demand of each component.

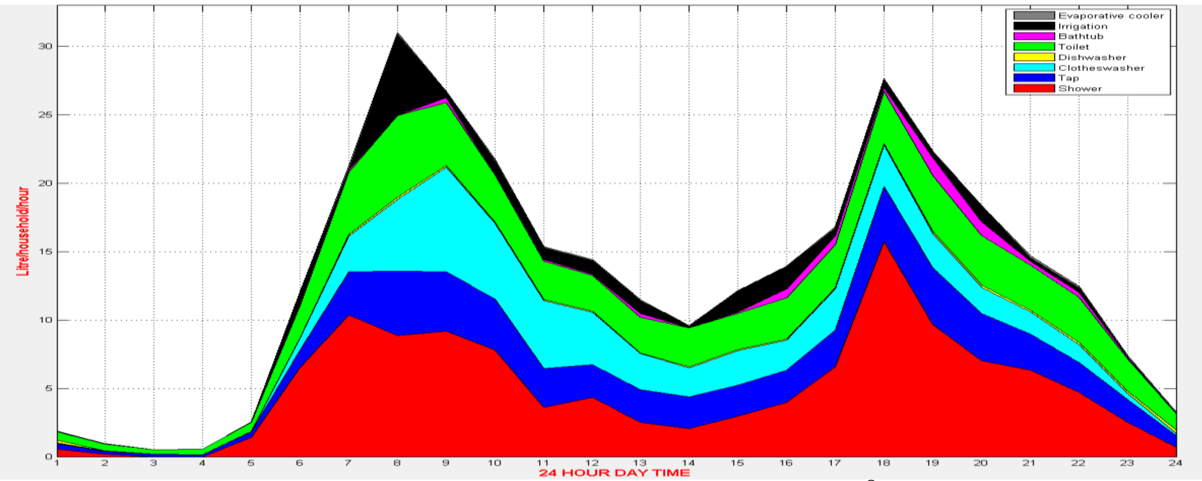

Figure 2 Diurnal pattern graph from *Autoflow*©

2.2 A new water demand forecasting module

The urban water demand forecasting task usually involves a range of variables measured at different periodicities, therefore, the question regarding which method to be employed for urban water demand forecasting cannot be adequately answered without specifying the forecast variable, its periodicity (i.e. daily, weekly, or monthly), and the horizon (i.e short term, medium term or long term forecast). A study conducted by (Donkor et al., 2014) has shown the priorities of water utilities (in decreasing order) as forecasting demand for peak day; daily total system demand; monthly total system demand; annual per capita demand; annual demand by customer class; and revenue , each of which required different input variable to perform the task. This study focuses on modelling the top two priorities, *daily peak demand* and *daily total system demand*, by using the collected hourly water consumption data and average environment temperature.

2.3 Applied techniques for water demand forecasting with Autoflow

Among various models that have been used for water demand forecasting including Univariate Time Series, Artificial Neural Network, Stochastic Process, Regression model or Composite model (Donkor et al., 2014), it was found that the stochastic state space approach was most effective as it is able to reflect the effects of different factors on the variation of water consumption (e.g. population growth over time, change in temperature, etc.) in the context of limited historical data for model development as in this study (Young & Pedregal, 1999a). The model is considered as the observation equation of a discrete time, stochastic state model and the associated state equations are used to model each of the components in Gauss- Markov (GM) terms.

3 Model development

3.1 Overall forecasting model establishment

The forecasting model in this study has the form of:

$$y(t) = S(t) + F(t) + e(t) \qquad e_t \sim N\{0, \sigma^2\} \tag{1}$$

where $y(t)$ is the forecasted hourly water consumption data (litre/hour) at time t; $S(t)$ is a seasonal component that directly reflects the periodic pattern of the data; $F(t)$ is the function that describes the influence of the external factor on the water consumption, which is the daily temperature in this case; and $e(t)$ is the noise component used to model the random changes of water consumption due different user behaviour. In order to allow for the nonstationary in the time series, all components in Eq. (1) can be characterised by stochastic, time variable parameters (Young, 1998). In this model, the most important component is the seasonal term $S(t)$ that determines the periodic pattern of the consumption data signal, and is defined as:

$$S(t) = \sum_{i=1}^{R} \{a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)\} \tag{2}$$

where $a_{i,t}$ and $b_{i,t}$ are stochastic parameters and $\omega_i, i = 1,2, \dots R$ are the fundamental and harmonic frequencies associated with the seasonality in the series.

The external signal component $F(t)$ allows for the possibility that the water consumption series is affected by temperature. An appropriate characterization of this component is the following deterministic transfer function (Young, 1998):

$$F(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t - \delta) \tag{3}$$

where u is the input temperature and δ is the time lag, which is set as 1 to imply that the effect of temperature on water consumption will come after one-time interval, i.e. one hour. $B(z^{-1})$ and $A(z^{-1})$ are the polynomials in the backward shift operator (z^{-1}) that define the transfer characteristics between the temperature and the water consumption, and have the following form of:

$$A(z^{-1}) = 1 + a_1(z^{-1}) + a_2(z^{-2}) + \dots + a_n(z^{-n}) \tag{4a}$$

$$B(z^{-1}) = 1 + b_1(z^{-1}) + b_2(z^{-2}) + \dots + b_n(z^{-n}) \tag{4b}$$

Given the fact that impact of weather on water consumption is relatively noticeable and straightforward, the first order of transfer function was recommended. In terms of state-space form, $F(t)$ can be presented as:

$$x_f(t) = \boldsymbol{F_f} x_f(t-1) + \boldsymbol{G_f} \varphi_f(t-1) \tag{5a}$$

$$F(t) = \boldsymbol{H_f} x_f(t) \tag{5b}$$

where $F_f = 1$, $G_f = 1$ and $H_f = 1$ when the first order transfer function was adopted (Harvey 1989).

With the availability of observation and state-space equations of all components, the aggregation of all subsystem matrices into a standard state space format can be undertaken as presented in Eq. (6).

State-space equation

$$\boldsymbol{X}(t) = \boldsymbol{F}x(t-1) + \boldsymbol{G\varphi}(t-1) \tag{6a}$$

Observation equation

$$y_t = \boldsymbol{HX}(t) + \boldsymbol{\varphi}(t) \tag{6b}$$

Where, the state vector $\boldsymbol{X}(t)$ is composed of all state variables from the seasonal and external temperature input model. The white noise vector $\boldsymbol{\varphi}(t)$ is defined by the white noise disturbance input of the constituent model. In order to have a good approximation of the collected water consumption data $y(t)$ from the state-space model, the key task is lying at an accurate estimation of the time variable parameters $\boldsymbol{X}(t)$ which can be done using Kalman Filtering accompanied by the optimal smoothing procedures as described in (Kalman, 1960 and Young & Ng, 1989).

3.2 Forecasting of future water consumption

With the availability of $\boldsymbol{X}(t)$ obtained from the previous step, forecasting of future water consumption can be performed straightforwardly by the applying the state-space filtering/smoothing algorithm. The f-step-ahead forecasts of the aggregate state vector $\boldsymbol{X}(t)$ in Eq. (6) are obtained at any point in the time series by using Eq. (7a):

$$\hat{X}(t+f|t) = F^f\hat{X}(t) \tag{7a}$$

Where (f) denotes the forecasting period. The associated forecast of $y(t)$ is provided by Eq. (7b):

$$\hat{y}(t+f|t) = H\hat{X}(t+f|t) \tag{7b}$$

3.3 Disaggregation of forecasted overall water consumption to end use level

The establishment of the forecast model developed in previous sections allows future water demand to be predicted at an hourly basis. Moreover, with the assistance of *Autoflow*©, the predicted demand at hourly basis can be further disaggregated to an end-use level using the following steps:

(i) Perform an end-use analysis using the previous two-week data, whose pattern would reflect the current consumption trend of the predicted day

(ii) With the disaggregated volume of all events obtained from step (i), determine the volume distribution for each category through Eq. (8), which is a $(m\ by\ 24)$ matrix P, where m is the number of end-use component and $volume_{i,j}$ is the total volume of category i collected at time j of the day.

$$P_{i,j} = \frac{volume_{i,j}}{\sum_{i=1}^{m} volume_j} \qquad i = 1,2,\ldots,m \text{ and } j = 1,2,\ldots,24 \tag{8}$$

(iii) Apply this volume distribution on the predicted 24-hour demand ahead. Eq. (8) can be interpreted as: if $i = 1$ corresponds to shower category, $P_{1,1}$ will represent the percentage of shower volume in comparison with the total volume collected, for instance, at 1 am during the last two weeks. For example, if $P_{1,1}$ = 25%, and the predicted water consumption at 1 am of the next day is 1000 litres, then the predicted shower at 1 am is 250 litres. By doing this, predicted volumes of all categories can be determined.

4. Model illustration

The developed forecasting model has been verified with three different datasets collected from a sample of 200 properties in Queensland, Australia at three different times of the year (i.e. in April, June and September) in order to evaluate the impact of temperature on water demand as well. Among a total of 100 tests carried out, one of them was randomly selected to explain the whole process. In this section, illustration was conducted on the April dataset (see Figure 3) where a series of 12-day water consumption data was randomly chosen, in which the first 11 days were used as historic data (blue line) and the last day was the reference for model validation (red line). The input temperature (in ^{o}C) for this model is:

$$temp = [26.2, 26, 26.7, 27.1, 26.6, 27, 25, 27.9, 25.1, 24.2, 24.4, 24.2]$$

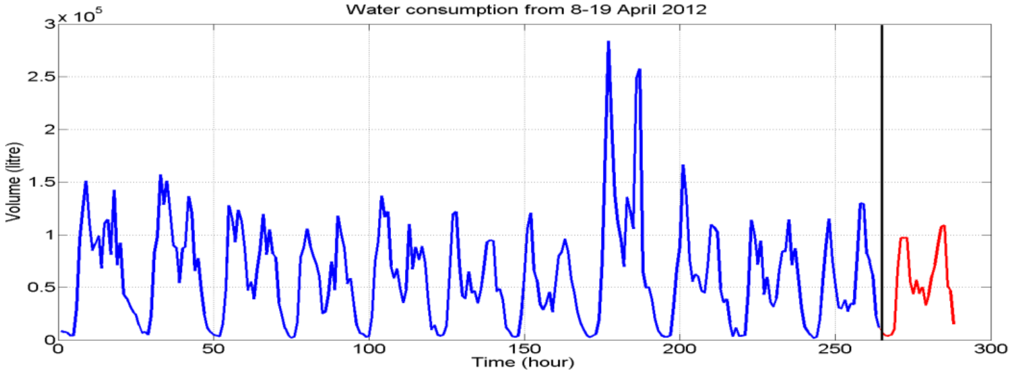

Figure 3 Collected water consumption for model validation

Step 1: Assess the correlation between the temperature and daily water consumption of the historic data. If this correlation coefficient is above 0.5, the temperature will be incorporated into the model to help improve the accuracy. In this illustration, there is a correlation between the temperature and the daily water consumption (R=0.65), then the temperature is considered as an exogenous input to the model.

Step 2: The second step is to reconstruct the model developed in Section 3 and use it to perform a 24-hour-ahead forecast. Figure 4 below shows the forecasted 24-hour-ahead consumptions in comparison with the actual data.

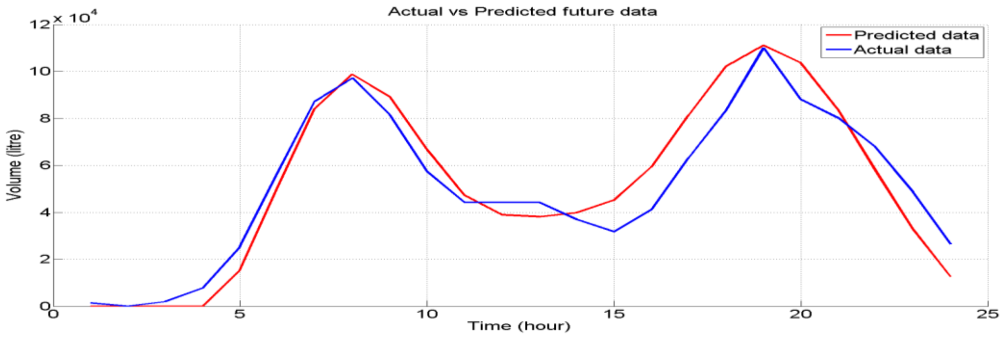

Figure 4 Actual vs predicted water consumption

Step 3: An end use disaggregation process is conducted on the first 11 days of historic data to find the volume distribution of each end use along the day (Figure 5)

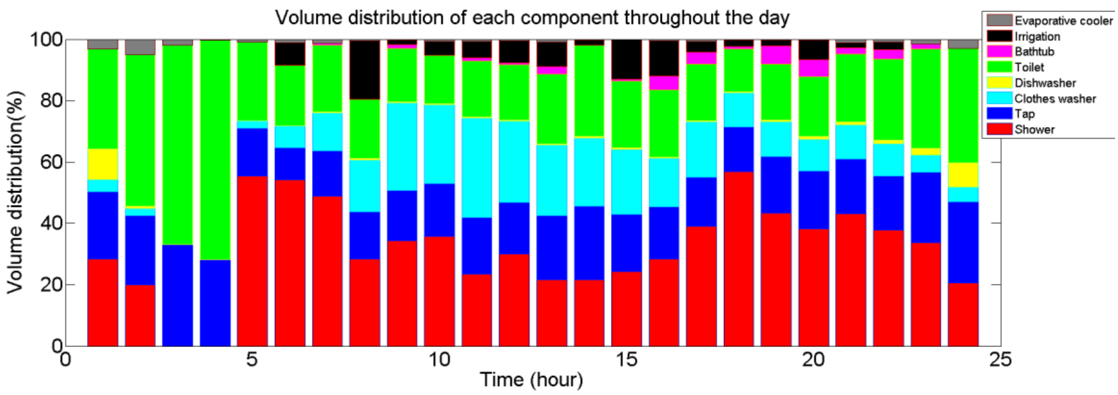

Figure 5 Determination of volume distribution of each end use category

Step 4: Apply that distribution on the forecasted consumption to find the volume of each end-use category. Figure 6 displays the outcome of *Autoflow*© on the forecasting of the next 24 hours consumption with a detailed volume of each component.

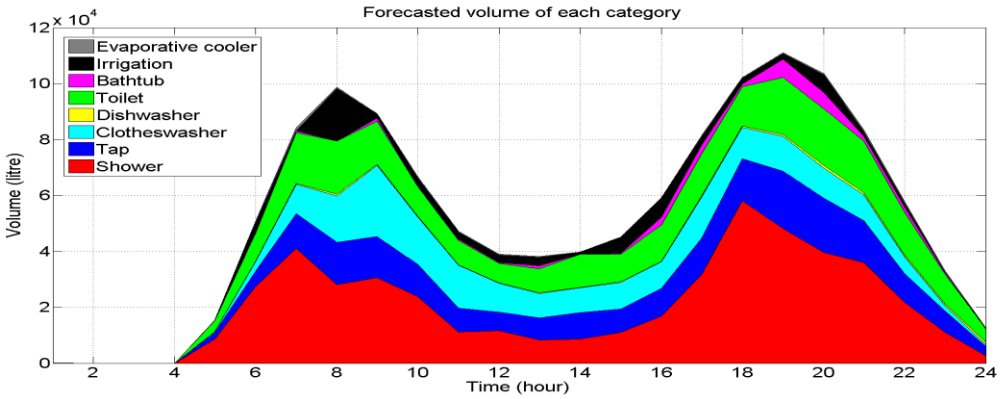

Figure 6 Disaggregation of predicted 24-hour water demand

5. Model verification

This section aims to estimate the overall model efficiency based on the obtained R^2 and the error of the forecasted daily water demand of the total 100 tests. As can be seen from Figure 7 that all R^2 values varied in between 0.83 to 0.95 (the average of 0.923), which shows a high degree in the goodness of fit of the forecasted consumption on the actual observed data.

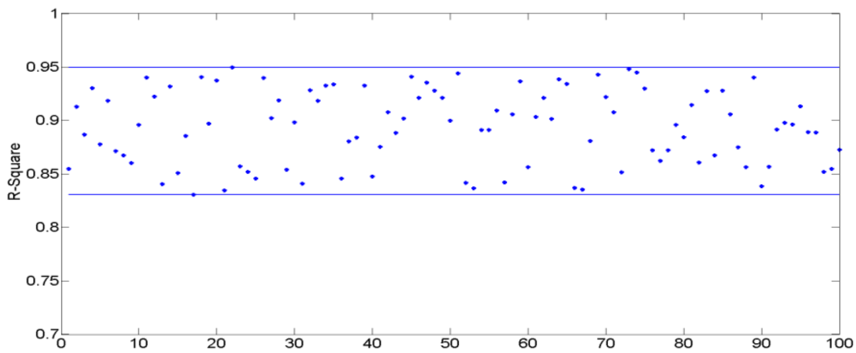

Figure 7 Obtained R^2 of 100 tests

In terms of estimating model efficiency on predicting water consumption at end-use level, it is important that the forecasted volume of each end-use should be as close to the actual volume as possible, hence, in this error analysis process, the difference (ε_t) between actual and forecasted volume at each hour along the day is estimated Eq. (9).

$$\varepsilon_t = \left|\frac{(y_t - \widehat{y_t})}{y_t}\right| \tag{9}$$

Where y_t and $\widehat{y_t}$ are the observed and forecasted water consumptions at t-step ahead from the forecast origin respectively.

In this verification process, the error was only estimated between 6 am to 10 pm of the day when major water consumption activities occur. The obtained error (%) of all 8 eight categories when measured during the tested period are less than 5%. In terms of the four main categories including shower, tap, toilet and clothes washer, which contribute most volume during the peak hours, the error

(%) always remains below 1.5%, which is a great indicator to help the water utility to moderate the supply. On the other hand, as the forecasting error for irrigation, evaporative cooler and bathtub are also low (less than 2%) during the peak period, the utility can rely on that to lower the peak demand by giving some instructions or incentive schemes to help reduce the water usages from these categories (e.g. higher rate applied on irrigation consumption during 5 pm to 7 pm).

6. Conclusion

This paper describes an accurate approach to the non-stationary water demand forecasting problem based on a state-space model that allows for the presence of trend, seasonal, irregular component and the external effect from temperature. This approach exploits the powerful properties of Kalman Filter to recursively estimate all the model components, where the hyper-parameters for trend and seasonal were optimised using a cost function defined in terms of the difference between the logarithm pseudo-spectrum of the model and the logarithmic autoregressive of the data. In terms of temperature effect, the Maximum Likelihood method was applied to estimate the Noise Variance Ratio, a critical hyper-parameter in the exogenous component. Once the main forecasting model has been developed, *Autoflow©* was then incorporated to help disaggregate the forecasted overall demand into each end use by taking advantage of knowledge obtained from the analysis of historic data. The forecasting model was validated against different dataset collected at different time of the year, including April, June and September in 2013 to estimate its efficiency. It was found that most of the future data (i.e. 24 hour ahead) can be explained by this model where the goodness of fit between actual and forecasted data is at least 0.9 for all testings. In terms of the accuracy obtained on the forecasted volume of each category, the percentage error of all eight categories is less than 5%, which proves that the developed technique is also effective in predicting demand for each category. However, to further verify the efficiency of this model, more testing on different dataset is required in future study. The developed forecasting model has great implication in practice as it can assist water utility in moderating the water supply throughout the day, optimising pumping schedule, or reducing peak demand by restricting the water consumption of the subordinate categories such as irrigation, dishwasher, or clothes washer to a certain limit. The next step of this study is to further extend *Autoflow©'s* capability to commercial buildings, which contribute a significant part to the overall consumption demand, so that more effective network distribution and supply planning can be undertaken.

Reference

Billings, B., and Jones, C. (2008). Forecasting urban water demand, 2nd Ed., *America Waterworks Association*, Denver, CO.

Donkor, E., Mazzuchi, T., Soyer, R., and Roberson, J. (2014). Urban water demand forecasting: A review of methods and models. *Journal of Water resource Planning and Management*, 140(2), 146–159

Harvey, A. C. (1989) Forecasting Structural Time Series Models and the Kalman Filter, *Cambridge: Cambridge University Press*

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, 108-127.

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.

Young, P. C. (1998) Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, 13, 105-122.

Young, P. C. and Pedregal, D. J. (1999a) Recursive and en-block approaches to signal extraction. *Journal of Applied Statistics*, 26, 103-128.

Young, P. C. and Ng, C. N. (1989)Variance Intervention. *Journal of Forecasting*, 8, 399 - 416.