

## Modeling water use in schools: a comparative study of quarterly and monthly models

**S. Barua<sup>a</sup>, A.W.M. NG<sup>a</sup>, S. Muthukumar<sup>a</sup>, F. Huang<sup>a</sup>, P. Roberts<sup>b</sup> and B.J.C. Perera<sup>a</sup>**

<sup>a</sup>Victoria University, P.O. Box 14428, Melbourne, Victoria 8001, Australia

<sup>b</sup>Yarra Valley Water, Lucknow street, Mitcham, Victoria 3132, Australia

Email: suchana.barua@live.vu.edu.au

**Abstract:** Urban water supply systems provide water for a range of water uses from human consumption to fire control, and from garden irrigation to industrial processes. The amount of water consumed in an urban area depends on the activities. Therefore, urban water use is categorised into different sectors such as residential, non-residential and unmetered water use.

Modeling urban water use has significant importance in water resources planning and management. From the literature, a vast amount of research was found on modeling residential water use. However, a significant portion (e.g. around 30% of total water use in Melbourne, Australia) of urban water use belongs to non-residential, and limited attention has been given to modelling non-residential urban water use, possibly due to the difficulty of data collection at end-use level. However, billing data are now available in monthly and quarterly time steps with the latter being widely available. Such data can be utilized to build models for predicting non-residential water use.

Therefore, this study aims to develop non-residential water demand models by identifying unique groups based on their homogeneous nature of water use and then further disaggregated into different sub-groups based on their volumes of water uses. Quarterly billing data from 2000 to 2011 were obtained within the Yarra Valley Water (YVW) service area for 46,000 non-residential customers. Of them, monthly water use data available only for 100 customers. Therefore, comparative study of quarterly and monthly time step modeling was carried out by developing a regression model to predict non-residential urban water use in two schools within YVW service area and is presented in this paper.

Monthly water use data from 2005 to 2011 were used for this study as water usage patterns are significantly different in pre and post 2005 due in part to permanent water saving rules, which was introduced in 2005. Monthly data were accumulated into quarterly data for developing the quarterly model. Past water use, dummy variables for fixed quarterly effects, incidence of water restrictions, total rainfall and mean monthly maximum daily temperature were considered as influential variables. Although there were five levels of restrictions during this period, due to insufficient data points for some of the restriction levels, all types of restrictions were considered as a single level of restriction. Model performance was measured using Nash-Sutcliffe efficiency ( $E$ ) and relative error values.

Model calibration and validation were performed using two independent data sets by splitting the total record into two periods. Two sets of calibration periods: 2006-2009 and 2006-2010 and validation periods: 2010-2011 and 2011 were also tested to find the best period for model development and validating the models. It was found that the models developed with 2006-2010 calibration period have given better results for both of the schools.

Although, both of the monthly and the quarterly time step models were found to predict well, quarterly model performed slightly better than the monthly model with less relative error and good  $E$  value.

**Keywords:** *Non-residential urban water demand, Modeling, Regression, Monthly, Quarterly, School.*

## 1. INTRODUCTION

Urban water supply systems provide water for a range of water uses from human consumption to fire control, and from garden irrigation to industrial processes. The amount of water consumed in an urban area depends on the activities occurring within its boundaries; each urban area has its own economic base, creating its own pattern of water use. The type of activities creating this pattern are categorized into different sectors such as residential (e.g. single house, multi-unit apartment, etc.), non-residential (e.g. industrial, schools, etc.) and unmetered (non-revenue) (Institute for Sustainable Futures, 2002).

There are a large number of studies conducted on modeling urban water use around the world. These studies used approaches such as trend analysis, climate correction approach, analysis of base and seasonal use, and modeling behavioral response on water use. Billings and Jones (2008) forecasted urban water use for several cities in USA by identifying long-term trend in water demand to provide an indication of long-term changes. Maheepala and Roberts (2006) and Perera et al. (2009) used the climate correction approach to estimate water use under normal climate conditions in the absence of restrictions using multiple regression. Danielson (1979) used base and seasonal water use analysis in urban water demand modeling for the City of Raleigh, in USA. A similar approach was also used by Maidment et al. (1985), Miaou (1990), Zhou et al. (2000) and Gato et al. (2007a; 2007b). To address the human behavior on urban water use, an end-use demand model was used by Blokker et al. (2010) in the Netherlands considering behavioral parameters (e.g. shower length and frequency). Billari et al. (2006) developed an agent-based model to simulate consumer behavior and responses to urban water planning policy and initiatives.

The aforementioned studies focused either on the total urban water use or in most cases on modeling the residential water use component. However, a significant portion of urban water usage is non-residential. For example, around 30% of total water use in Melbourne is non-residential (Melbourne Water, 2006). There is not much attention given to modeling of the non-residential urban water use. This is an important omission, but it has enormous importance to mitigate the emerging water-related challenges including the need for a reliable water supply, rising water prices and seasonal water scarcity (Worthington, 2010). One possible reason for this omission is that the appropriate data required for estimating non-residential water demand is difficult to collect and also the heterogeneous nature of this sector. However, disaggregated data are now available in electronic formats (Polebitski and Palmer, 2010). These data can be used to model non-residential water demand.

Currently, an end-use model is used by water authorities in Melbourne as the primary tool for its demand forecasts, in which the residential component is given more importance (Institute for Sustainable Futures, 2005). Total household water usage in a single-family home is broken down into the end-use level (e.g. toilets, showers, washing machines, etc.) for forecasting the water demand in the residential sector. On the other hand, historical trend-based annual water demand is considered for the non-residential sector, as a whole. There are no temporal (i.e. quarterly or monthly) and spatial disaggregated water demand model for forecasting non-residential water demand for Melbourne.

As mentioned earlier, disaggregated billing data in electronic format for each customer are now available in monthly and quarterly time steps with the latter being widely available. Therefore, this study aims to develop non-residential water demand models by identifying unique groups based on their homogeneous nature of water use and then further disaggregated into different sub-groups based on their volumes of water use. In particular, investigation of quarterly and monthly time step modeling was carried out and presented in this paper. Quarterly and monthly time step models were developed using Multiple Linear Regression (MLR) technique for non-residential urban water uses in two schools located in Eastern part of Melbourne, Australia. The rest of the paper begins with a brief description of the study area followed by data sources and processing. Then the methodology is presented followed by results and discussions and finally summary and conclusion is presented.

## 2. STUDY AREA

Yarra Valley Water (YVW) is the largest Melbourne's water retailer providing water and sewerage services to about 660,000 residential homes (more than 1.7 million people) and 52,000 non-residential customers in an area covering over 4,000 square kilometers of Melbourne's northern and eastern suburbs. Within this region, YVW maintains and operates 9,490 kilometers of water mains and 9,004 kilometers of sewer mains (Yarra Valley Water, 2011). YVW purchases treated bulk water from Melbourne Water Corporation which manages the city's water supply catchments, large reservoirs and transfer system. The total water usage recorded for YVW in 2009/2010 was 132,000 ML. Of this total water usage, 19% was consumed by the non-residential sector. The non-residential customers of YVW consist of different groups such as top 42 (highest

42 water users), schools, sport grounds, restaurants, laundries, hospitals, hotel and motels, and council properties.

### 3. DATA SOURCES AND PROCESSING

Quarterly billing data from 2000 to 2011 were obtained for 46,000 non-residential customers located within YVW's service area. Furthermore, monthly water use data for 100 customers for the same period with different stages of water restrictions was obtained from YVW. However, it was found that water usage patterns are significantly different in pre and post 2005 due in part to permanent water saving rules which were introduced in 2005 and also possibly because of alternate sources of water use (e.g., stormwater, recycle water). Therefore, it is assumed that the current reduced water use pattern will continue into the future and monthly water use data for two schools from 2005 to 2011 were used in this study. For quarterly time step model, monthly water use were aggregated into four quarters (i.e. January to March, April to June, July to September, and October to December). This monthly time-frame used for defining quarters was selected based on the quarterly billing's time-frame in YVW.

Mean monthly maximum daily temperature data ( $^{\circ}\text{C}$ ) and total monthly rainfall data (mm) were also collected from Bureau of Meteorology. Total quarterly rainfall and mean quarterly daily maximum temperature were considered for climate effect in quarterly time step modeling.

### 4. METHODOLOGY

Water consumption varies over time due to the effects of population growth, socioeconomic factors, climate and water use restrictions (Gato *et al.*, 2007a). There are several approaches used for modelling water use, as discussed earlier in Section 1. However, due to the lack of data availability and limited length of data period, a simple and commonly used multiple linear regression (MLR) approach was employed in this study.

House-Peters and Chang (2011) revealed that there was strong influence of past water use on current water use. Therefore, the past water use in terms of the (t-1) and (t-4) time-lagged (i.e. the previous quarter and the same quarter in the preceding year) water use was considered as influential variable for modelling quarterly water use in this study to consider the lagged effect and adaptive seasonal variations respectively. In monthly time step modelling, these variables considered as (t-1) and (t-12) time-lagged water use. The quarterly model also accounted for four quarters mentioned in Section 3 to consider the seasonal effect, and water restriction information (i.e. no restrictions and imposed water restrictions) through dummy independent variables. Tables 1 and 2 show how these dummy variables were considered in MLR to account for restrictions and seasons respectively. It should be noted that there were five levels of water restrictions such as permanent water saving rules, stage 1, stage 2, stage 3 and stage 3a during the period 2005-2011. However, there were insufficient number of stage 1, stage 2 and stage 3 restrictions data points in the calibration data set used in this study. Therefore, all types of restrictions were considered as a single restriction level for quarterly and monthly time step modelling in this study.

Table 1. Dummy variables used for water restrictions

Water restriction information	R
No restriction	0
Imposed water restriction	1

Table 2. Dummy variables used for four quarters

Quarter	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>
Quarter 1 (January – March)	0	0	0
Quarter 2 (April – June)	1	0	0
Quarter 3 (July – September)	0	1	0
Quarter 4 (October – December)	0	0	1

Both schools have gardens, sport grounds and pools where water use depends on climatic factors. Therefore, total rainfall (Rn) and mean maximum daily temperature (T) were also considered as likely independent climatic variables for modeling water use in schools in this study.

In order to make the variables normally distributed, water use data, rainfall and temperature were transformed using a log function (which is a requirement for multiple linear regression), before developing the regression models. Mathematical details of the regression model used for quarterly model in this study are as follows:

$$\text{Log}_{10}(\text{WU}_t) = b_0 + b_1 \log_{10}(\text{WU}_{t-1}) + b_2 \log_{10}(\text{WU}_{t-4}) + b_3 \log_{10}(T) + b_4 \log_{10}(\text{Rn}) + b_5 R + b_6 D_2 + b_7 D_3 + b_8 D_4 \quad (1)$$

$$\text{WU}_t = 10^{(\log_{10}(\text{WU}_t))} \quad (2)$$

where,

$\text{WU}_t, \text{WU}_{t-1}$ and $\text{WU}_{t-4}$	= quarterly water use at time $t, (t-1)$ and $(t-4)$ respectively (KL/q)
$T$	= Quarterly mean maximum daily temperature ( $^{\circ}\text{C}$ )
$\text{Rn}$	= Total quarterly rainfall (mm)
$R$	= dummy variables to represent water restriction as shown in Table 1
$D_2, D_3$ and $D_4$	= dummy variables to represent four quarters as shown in Table 2
$b_0, b_1 \dots b_8$	= regression constant and coefficients

Mathematical details for the monthly model are similar to the quarterly model but there was no dummy variables considered for months.

Model calibration and validation were performed using two independent data sets by splitting the total data period into two sets. Furthermore, two different calibration data sets were used in this study, to see whether model predict better for both of the calibration and the validation period with increased calibration data set. Therefore, two separate length of data set were used for calibration: 2006-2009 and 2006-2010 and for validation: 2010-2011 and 2011. It should be noted that data of year 2005 was not used in modeling water use, since  $\text{WU}_{t-4}$  data were not available for 2004 as water use data pre-2005 was not considered in this study as mentioned in Section 3.

The model performance was measured using the Nash-Sutcliffe efficiency ( $E$ ) which is commonly used for such performance evaluation. This measure is as follows:

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (3)$$

Where,  $O$  and  $P$  are the observed and predicted water use values respectively;  $\bar{O}$  is the average observed water use value; and  $n$  is the number of data points.

$E$  is a normalized statistic that determines the relative magnitude of the residual variance compared to the observed data variance (Nash and Sutcliffe, 1970).  $E$  indicates how well the plot of observed vs. modeled data fits the 1:1 line.  $E$  varies from 1 (perfect fit) to  $-\infty$ . An efficiency of lower than zero indicates that the mean value of the observed time series would have been a better predictor than the model. A model with a large  $E$  value (e.g. close to 1) was considered in selecting a good model.

To check the validity of the model, relative error was also employed in this study. The relative error determines the ratio of the absolute error to the measured values in percent. This statistic eliminates the bias of high errors that mask those of low value by sheer magnitude and is a standard measure of error in data analysis (Gato et al. 2007a).

All parameters considered in this study are tested of its statistical level of significance by its  $p$  value. The  $p$  value is the smallest level of significance at which the parameter is significant. Conventionally (and arbitrarily) a  $p$  value of 0.05 (5%) is generally regarded as sufficiently small. The 5% value is called the significance level of the test. In this study, a  $p$  value lower than 0.05 is considered statistically significant.

The quarterly MLR model was first considered using all six selected variables (i.e.  $\text{WU}_{t-1}, \text{WU}_{t-4}, R$ , three dummy variables for four quarters,  $\text{Rn}$  and  $T$ ) and variables were removed one at a time until the remaining variables shows statistically significant. Thereafter,  $E$  values, and the time series plot was obtained for calibration and validation period using the best model. However, if one of these criteria showing the sign of poor result, then one or two variables were added from the earlier removed variables those showed closely statistical significant, and thereafter re-performed the above process to obtain the best model. The reason for including these in-significant variables is that it may provide good contribution in the model if number of

data points are limited (Lewis-Beck, 1980), which was the case in this study. For monthly model development, a similar approach to quarterly model development was followed.

## 5. RESULTS AND DISCUSSIONS

The  $E$  values of the best possible monthly and quarterly time step models for two calibration periods (i.e. 2006-2009 and 2010) and validation periods (i.e. 2010-2011 and 2011) are presented in Table 3.

From Table 3, it can be seen that for school 1  $E$  value is slightly decreased and relative error modestly increased in calibration period for including one year more data. However, it can be seen from Table 3 that during validation period of school 1 quarterly model,  $E$  value is considerably high and relative error also low for model developed with increased calibration period. Moreover, it is also identifiable from Table 3 that  $E$  value for monthly model in school 1 is almost same for both validation periods but relative error decreased for model developed with 2006-2010 calibration period.

In case of school 2, it is identifiable from Table 3 that, although  $E$  value increased for calibration for both of monthly and quarterly model but decreased during validation with the model developed with increased calibration period. However, relative error for monthly and quarterly model decreased with the model developed with 2006-2010 calibration period for both cases of calibration and validation.

From the above discussion, it can be concluded that Model developed with 2006-2010 calibration period providing an overall better result than the model developed with 2006-2009 calibration period. Therefore, coefficient obtained for the best models using this calibration period are presented in Table 4 and comparison of observed and modelled water use time series for both of the schools are shown in Figure 1.

Table 3.  $E$  values for the best models

School	Modelling time step	Calibration : 2006-09			Calibration : 2006-10		
		Calibration (06-09)		Validation (10-11)	Calibration (06-10)		Validation (11)
		$E$	Variable used	$E$	$E$	Variable used	$E$
School (1)	Monthly	0.71 (17.92)	WU <sub>t-1</sub> , T, Rn, R	0.61 (19.68)	0.67 (21.90)	WU <sub>t-1</sub> , T, Rn, R	0.55 (6.28)
	Quarterly	0.88 (12.79)	WU <sub>t-1</sub> , T, Rn, R	0.38 (80.82)	0.83 (17.92)	WU <sub>t-1</sub> , T, Rn, R	0.47 (8.20)
School (2)	Monthly	0.71 (10.46)	T, Rn, R	0.71 (15.60)	0.74 (9.77)	T, Rn, R	0.59 (10.20)
	Quarterly	0.87 (7.49)	T, Rn, R	0.75 (10.64)	0.89 (5.92)	T, Rn, R	0.13 (1.20)

Note: Relative error is presented in the parenthesis

Table 4. Estimated regression coefficients for both schools' water use models (calibration period: 2006-10, validation period: 2011)

School	Modelling time step	Constant	Log (WU <sub>t-1</sub> )	Log (T)	Log (Rn)	R
		b <sub>0</sub>	b <sub>1</sub>	b <sub>3</sub>	b <sub>4</sub>	b <sub>5</sub>
School (1)	Monthly	-0.78	0.23 (0.05)	2.48 (0.00)	-0.14 (0.09)	-0.22 (0.04)
	Quarterly	1.62	-0.23 (0.14)	3.23 (0.00)	-0.75 (0.01)	-0.4 (0.00)
School (2)	Monthly	0.88	–	2.13 (0.00)	-0.11 (0.02)	-0.09 (0.09)
	Quarterly	1.48	–	2.38 (0.00)	-0.28 (0.01)	-0.14 (0.02)

Note:  $p$  values are presented in the parenthesis

### 5.1 School 1

From Table 4, it can be seen that all included variables are nearly significant with expected sign for temperature, rainfall and restriction in both quarterly and monthly time step model. From Table 3 and Figure 1 it can be seen that monthly model and quarterly model, both are giving good prediction. However, quarterly time step model gives slightly better result with higher  $E$  value and lower relative error during calibration period. Although, during validation period monthly model shows better results, reason could be the less data point for quarterly model.

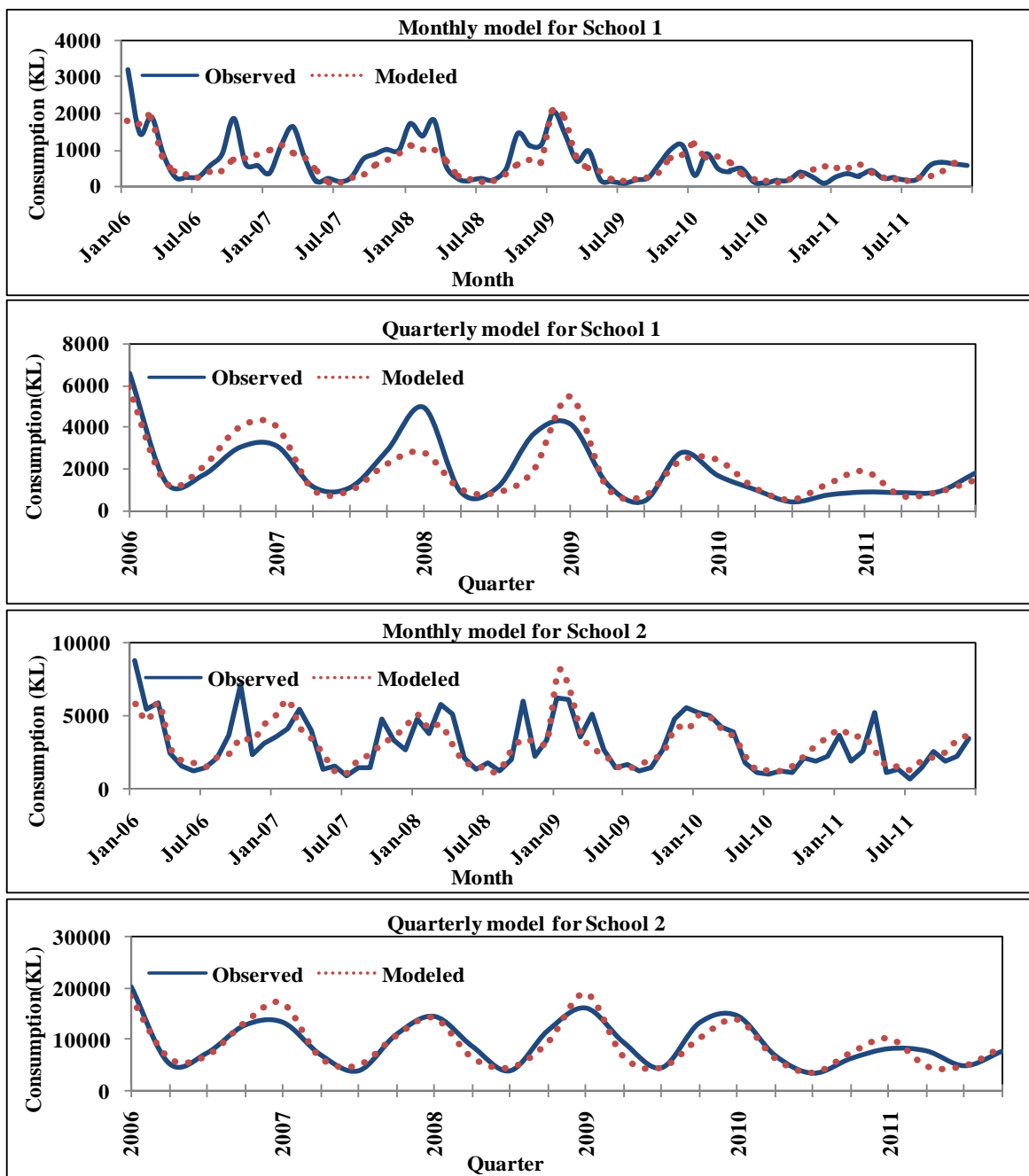


Figure 1. Comparison of observed and modeled time series for two schools

### 5.2 School 2

Similar to school (1), quarterly model for school (2) gives slightly better prediction than monthly model both in terms of *E* value and relative error during calibration period. Moreover, quarterly model gives lower relative error during validation period than the monthly model, although, *E* value is higher for monthly model. As mentioned for school 1 this might be because of less data points and sudden changes in water use pattern in 2011.

## 6. SUMMARY AND CONCLUSIONS

Urban water supply systems provide water for a range of water use. The amount of water use in an urban area depends on the activities. Therefore, urban water use is categorised into different sectors such as residential, non-residential and unmetered water use. This study focused on non-residential urban water demand modeling.

Quarterly billing data were obtained for 46,000 non-residential customers located within Yarra Valley Water's (YVW) service area along with monthly billing data for 100 customers. Therefore, an investigation of quarterly and monthly time step modeling was carried out in this particular study by developing regression model to predict non-residential urban water use in two schools.

Monthly billing data for six years of 2006–2011 were used in this study. Monthly data were aggregated to quarterly data for modeling purposes. Two different calibration periods: 2006-2009 and 2006-2010 were selected to find the best calibration period and rest of the data: 2010-2011 and 2011 were used for validation. The Multiple Linear Regression (MLR) models were developed for both schools using monthly and quarterly time step. The independent variables used for model development included time-lagged (past) water use, quarters for quarterly time step models, water restrictions, total rainfall (mm) and mean maximum daily temperature (°C), to predict the water usage for the next year on a monthly or quarterly basis.

It was found that modelled time series of water use data closely matched with the observed data with good Nash-Sutcliffe efficiency ( $E$ ) values and less relative error using 2006-2010 calibration periods for model development. Moreover, results indicate that, quarterly model gives slightly better prediction than the monthly model with less relative error and at the same time considerably good  $E$  value.

## REFERENCES

- Billari, F.C., T. Fent, A. Prskawetz, and J. Scheffran (2006). *Agent-based computational modelling: An introduction*, in Billari, F.C., T. Fent, A. Prskawetz, and J. Scheffran, "Agent-Based Computational Modelling - Applications in Demography, Social, Economic and Environmental Sciences", Physica-Verlag - A Springer Company, Heidelberg, New York, USA.
- Billings, R.B. and C.V. Jones (2008). *Forecasting Urban Water Demand*, American Water Works Association, Denver, Colorado, USA.
- Blokker, E.J.M., J.H.G. Vreeburg, and J.C. van Dijk (2010). Simulating Residential Water Demand with a Stochastic End-Use Model. *Journal of Water Resources Planning and Management*, 136(1), 19-26.
- Danielson, L.E. (1979). An analysis of residential demand for water using micro time-series data. *Water Resources Research*, 15(4), 763-767.
- Gato, S., N. Jayasuriya, and P. Roberts (2007a). Forecasting residential water demand: case study. *Journal of Water Resources Planning and Management*, 133(4), 309-319.
- Gato, S., N. Jayasuriya, and P. Roberts (2007b). Temperature and rainfall thresholds for base use urban water demand modelling. *Journal of Hydrology*, 337(3-4), 364-376.
- House-Peters, L.A., and H. Chang (2011). Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, 47(5), W05401.
- Institute for Sustainable Futures (2002). Melbourne end use and water consumption influences study. *Report to Water Resources Strategy Committee*, University of Technology, Sydney, Australia.
- Institute for Sustainable Futures (2005). Melbourne Options Model (MOM). *Manual*, University of Technology, Sydney, Australia.
- Lewis-Beck, and S. Michael (1980). *Applied Regression: An introduction*, Sage Publications, Beverly Hills, California, USA.
- Maheepala, S. and P. Roberts (2006). Towards development of an improved method for estimation of climate-corrected urban water use. Paper presented at 30<sup>th</sup> Hydrology and Water Resources Symposium, Launceston, Tasmania, Australia, December 4-7.
- Maidment, D.R., S.P. Miaou and M.M. Crawford (1985). Transfer function models of daily urban water use. *Water Resources Research*, 21(4), 425-432.
- Melbourne Water (2006). *Water Supply-Demand Strategy for Melbourne 2006–2055*, Melbourne Water, Melbourne, Australia.
- Miaou, S.P. (1990). A class of time series urban water demand models with nonlinear climatic effects. *Water Resources Research*, 26(2), 169-178.
- Nash, J.E., and J.V. Sutcliffe (1970). River Flow Forecasting through Conceptual Models: Part 1. A Discussion of Principles. *Journal of Hydrology*, 10(3), 282-290.
- Perera, B.J.C., N. Muttill and M. Hasofer (2009). Climate corrected urban water use and estimation of water savings. Paper presented at 18th MODSIM Congress, Cairns, Australia, July 13-17.
- Polebitski, A.S. and R.N. Palmer (2010). Seasonal residential water demand forecasting for census tracts. *Journal of Water Resources Planning and Management*, 136(1), 27-36.
- Worthington, A.C. (2010). Commercial and industrial water demand estimation, *Estudios de economía aplicada*, 28(2), 237-258.
- Yarra Valley Water (2011). Sustainability report, Yarra Valley Water, Melbourne, Australia.
- Zhou, S.L., T.A. McMahon, A. Walton, and J. Lewis (2000). Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*, 236(3-4), 153-164.