Hourly water demand forecasting for micro water grids

Juneseok Lee and Soo-Kwon Chae

ABSTRACT

Many countries have problems related to water scarcity and are thus seeking to promote greater water efficiency. A micro water grid (MWG) is a high-efficiency water management system that integrates information and communication technologies (ICT) for the water distribution systems in individual buildings. More accurate forecasting of hourly water demand is necessary if these systems are to function correctly and thus is the focus of this paper. Autoregressive integrated moving average (ARIMA) variant models were developed to create 24-hour lead-time forecasts. The forecast and observed values for both variant and traditional models were compared and developed models seem to perform well. It is therefore recommended to use linear stochastic models when developing MWGs for forecasting water demand to ensure sustainable water resource planning and management in MWG projects.

Key words | autoregressive integrated moving average (ARIMA), hourly water demand, micro water grid (MWG)

Juneseok Lee
Soo-Kwon Chae (corresponding author)
Department of Civil and Environmental Engineering,
San José State University,
San José,
CA 95112,
USA
and
Department of Environmental Health and Safety,
Eulji University,
212, Yangji-dong, Sujeong-gu,
Seongnam-si,
Gyeonggi-do 461-713,
South Korea
E-mail: *cskwen@eulji.ac.kr*

INTRODUCTION

The US National Academy of Engineering (2008) announced 14 major challenges that must be addressed in the 21st century, many of which are related to the sustainability of water resources. The Leadership in Energy and Environmental Design (LEED) program also encourages the sustainable use of water (UGBC 2006). According to Dziegielewski *et al.* (2000), water used in office buildings accounts for approximately 9% of the total water use in commercial and institutional facilities in the United States. According to the Office of the Federal Environmental Executive, green building design should include practices that improve the efficient use of water in buildings by implementing better design, operation, and maintenance throughout the building's entire lifecycle (Cassidy 2003). Adopting and promoting water efficiency strategies in buildings could therefore have a significant impact and are thus increasingly being applied in green building designs (UGBC 2006).

Micro water grid

A smart water grid (SWG) is defined as a high-efficiency water management system that integrates information and communication technologies (ICT) and can make a major contribution towards sustainable water resource management. In this paper, a micro water grid (MWG) is defined as *an SWG implemented at the individual building scale*. Customers within a building will have access to their total water use on a real-time basis and thus be able to make informed decisions related to their water use. At the same time, central water operators will receive real-time water demand information that will make future forecasts more accurate.

Demand forecasting

Arandia-Perez *et al.* (2009) presented time series modeling of hourly urban water use. They used autoregressive moving average (ARMA) models to identify the most suitable model. Mishra & Desai (2005) used autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) to forecast severity of water shortage. Their models could be used to forecast up to 2 months of lead-time (for monthly data).

doi: 10.2166/aqua.2015.144

Shvartser *et al.* (1993) used ARIMA model to forecast hourly demands for a period of one to several days ahead considering demand patterns. They noted that with increase of computational capacities, forecast can be performed in real time when the system state deviates from the planned or when new data become available. Now, this concept precisely applies to the MWG settings in which bi-directional and intelligent flow of water information between customer and MWG-based central operating and information service programs need to be updated on a real-time basis. In this paper, variants of ARIMA models are developed and applied to MWG. The authors believe that no studies have been done for developing MWG's hourly forecasting models.

Water conflicts and shortages arising from increasing population pressure, competing users, environmental concerns, and climate change impacts will all continue to put pressure on global water resources. Faced with rising costs, urban areas around the world are taking significant steps towards setting water conservation goals, especially in buildings' water use. To analyze and forecast water demand in an MWG enabled building, the authors therefore opted to utilize hourly water demand data from existing buildings that are of a similar scale to the possible MWG implementation. In this paper, hourly water demand is analyzed and forecast for a 13-floor office building located in San Francisco, CA. The results will help in the design, planning, and optimization of an MWG platform for green buildings.

METHOD

For this study, an ARIMA time series model was utilized for the time series component of the hourly water demand data to forecast future water demand. This approach assumes that all values of the variable represent measurements at equally spaced time intervals and that closer observations have a stronger dependency. Since the data are interdependent, autocovariance and autocorrelation can be used to measure the dependency among the variables in the time series (Box & Jenkins 1976) and the time series analysis applied to identify patterns and forecast future values from the historical time series data. Below are the mathematical formulations of ARIMA:

$$Y_t = \mu + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t \tag{1}$$

$$Y_t = C + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \tag{2}$$

$$Y_t = \mu + \sum_{k=1}^{p} \phi_k Y_{t-k} + \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k} \tag{3}$$

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)\epsilon_t \tag{4}$$

where ϕ = autoregressive or damping parameter; θ = moving average parameter; μ = mean value of the process; ϵ_t = forecast error at time t, in which ϵ_t is assumed to follow a normal $(0,\sigma)$ distribution and σ = standard deviation of the process. The models that are assumed to follow an autoregressive process of order p, expressed AR(p), a moving-average process of order q, symbolized by MA(q), and an autoregressive moving-average process of order (p,q), as ARMA(p,q). Autoregressive (AR) term predicts values from previous values and a moving average (MA) term accounts for previous random trends. When the time series has to be differenced by order d to make it stationary, the stationary AR backshift operator $\phi_p(B) = (1-\phi_1 B- \ldots -\phi_p B^p)$ and the invertible MA backshift operator $\theta_q(B) = (1-\theta_1 B- \ldots -\theta_p B^q)$, can be used to represent an ARIMA (p,d,q) process.

A seasonal component was added to the ARIMA models following the same structures as the non-seasonal part, which consists of periodic fluctuations defined by their repetitive movement around a trend line within time intervals. In this paper, the seasonal period is the 24 hours in each day which is known to have an on- and off-peak patterns. Time series data for selected days are displayed in Figure 1 (Wednesdays and Sundays from March 1st to June 30th). Figure 2 shows characteristics of hourly water use for selected days. Based upon observed data sets, it is assumed that each day's water demand is independent and ARIMA models are applied to each day instead of being applied to consecutive days, which is a typical practice for traditional ARIMA models.

The development of the ARIMA model involved three stages: identification, estimation, and diagnostic checking (Box & Jenkins 1976). In the identification stage the data

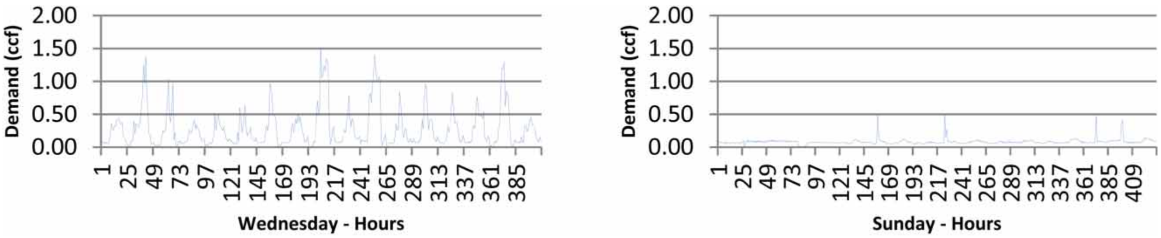

Figure 1 | Time series for selected days (all units in CCF: 1 CCF = 2.83 m^3).

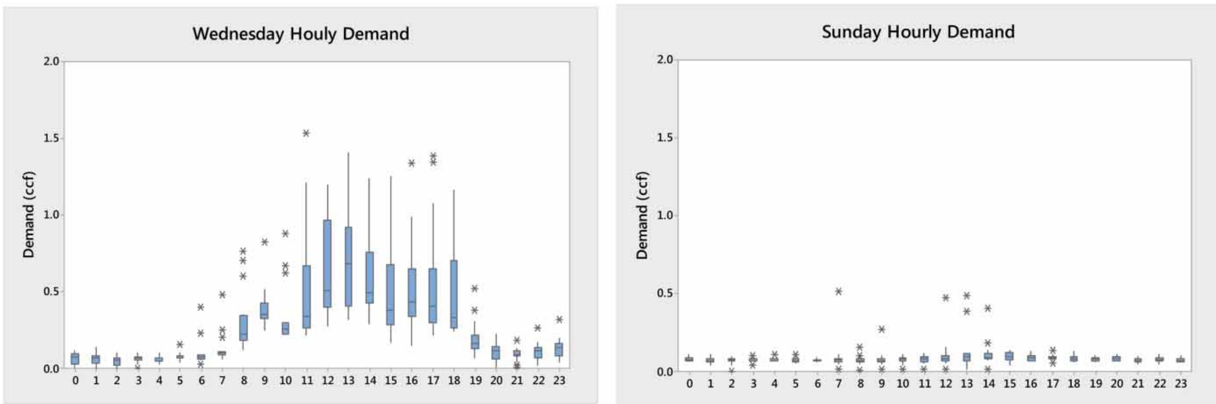

Figure 2 | Hourly water use for each day (March–June 2014; all units in CCF: 1 CCF = 2.83 m^3).

were transformed (if necessary) to improve the normality and the stationarity of the time series and determine the general form of the model to be estimated. Also, the temporal correlation structures of the transformed data were identified by examining the auto-correlation function (ACF) and the partial auto-correlation function (PACF). During the estimation stage the model parameters for the autoregression and moving average were calculated using the method of moments, least square methods, or maximum likelihood methods. Finally, diagnostic checks were applied to the fitted model to verify the model's adequacy. Based upon observed data sets (Figures 1 and 2), it is assumed that each day's water demand is independent and various ARIMA models were fitted for hourly demand instead of being applied to consecutive days, which is a typical practice for traditional ARIMA models. The seven best were then used to forecast demand for a 24-hour lead-time and the results compared with the observed data sets. To check the variant models' performance, forecast values were compared with those of traditional ARIMA models whose forecast was done for all days (without assuming independence for each day).

Case study sites

The hourly water demand data for a 13-floor office building in San Francisco, including its basement and roof top mechanical level, was entered in a spreadsheet file. The building's total square footage is approximately 277,500 ft^2 (=25,780 m^2) and each floor is about 20,000 ft^2 (=1,858 m^2). There are 34 bathrooms in the building; most floors have two bathrooms and there are eight showers in the basement bathrooms. The meter for the whole building is a compound meter, so the original data have tabs for both the low and high flow registers. Both registers were combined to give the exact hourly consumption. The building's regular business hours are Monday to Friday 7 am to 7 pm and there are about 950 employees working in the building. Lunch time is normally 12 noon to 12:30 pm.

The building uses reclaimed water to flush its toilets. This reclaimed water is generated in house via a 'Living Machine' sewage treatment system that processes sewage to a reclaimed water standard. All waste lines return to the Living Machine, where it is processed as raw sewage. This water is then treated and sent to a reuse tank that pumps the reclaimed water to all the toilet fixtures in the building. The domestic water uses in the building include faucets and showers, breakroom sinks, custodial sinks and closets, and the cooling tower. These represent the total hourly water demand for the current analysis.

The hourly water demand data were provided for the 4 months from March 1st to June 30th 2014. The meter readings are given in units of 100 cubic feet (CCF: 1 CCF = 2.83 m^3). The average monthly water demand for the 4 months was 153.71 CCF (±5.01%). As this is an office building it will have distinctly different water use patterns for weekdays and weekends (Figure 2). It was found that each workday (Monday to Friday) has a relatively large water use during the day (10 am–6 pm), but on Saturday and Sunday very little water is used. This is not unexpected as it represents a typical office-type water usage pattern.

RESULTS AND DISCUSSION

Following Box & Jenkins (1976), identification and estimation were performed and the residuals checked to reveal any patterns that were unaccounted for; the residuals left over after fitting the model should be white noise. All validation tests were carried out on the residual series. First, the residual ACF and PACF were obtained to confirm that the residuals were indeed white noise. A correlogram was drawn by plotting r_k against lag k, where r_k represents the residual ACF or PACF function. If some of the residual ACFs were significantly different from zero, this could indicate an inadequate model. The ACF, PACF, and plots of residuals for a selected day are shown in Figure 3. Most of the values lie within confidence limits except for a very few individual correlations that are close to the confidence limits. This indicates that

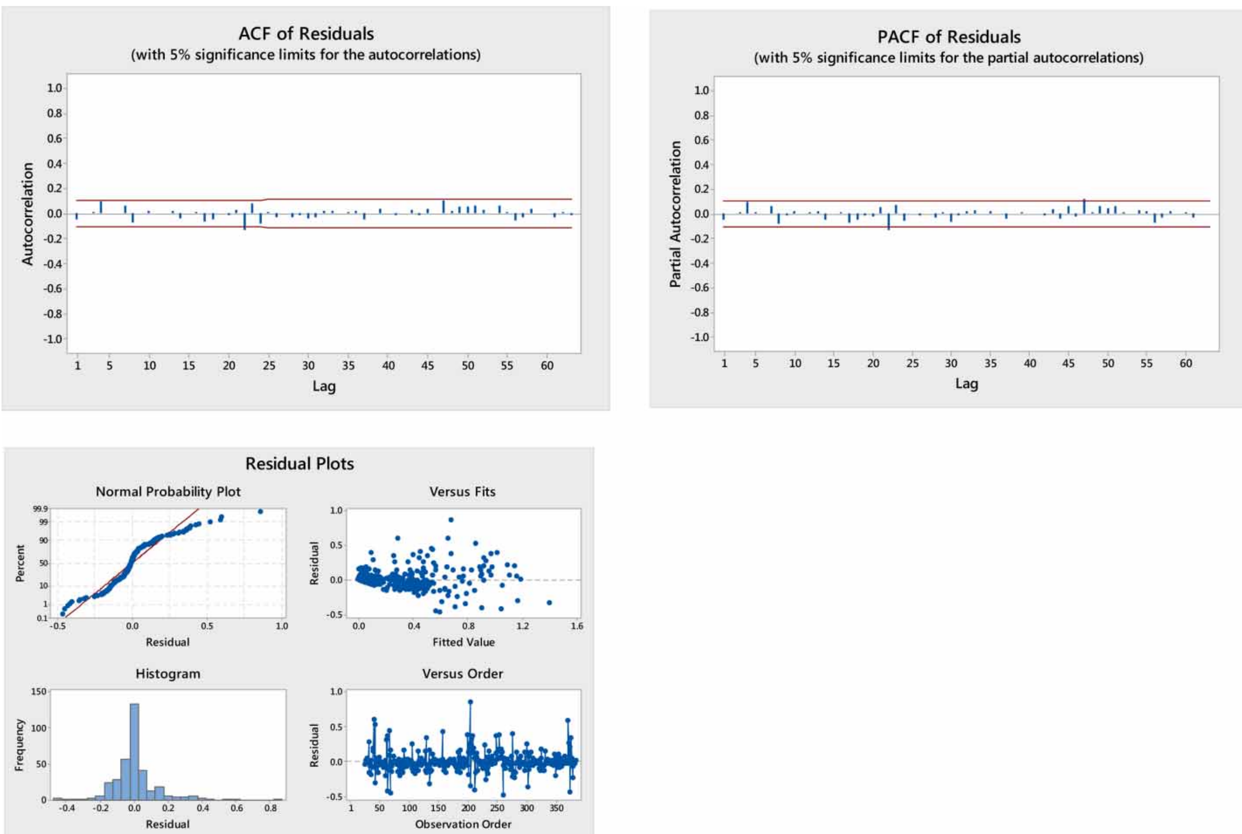

Figure 3 | ACF, PACF of residuals and residual plots for Wednesday.

Table 1 | Summary of ARIMA model coefficients

	Model	AR1 coeff	SE coeff	*p*	SMA24 coeff	SE coeff	*p*
Monday	$(1,0,0)(0,1,1)_{24}$	0.622	0.040	0.000	0.917	0.034	0.000
Tuesday	$(1,0,0)(0,1,1)_{24}$	0.644	0.041	0.000	0.885	0.037	0.000
Wednesday	$(1,0,0)(0,1,1)_{24}$	0.739	0.036	0.000	0.934	0.035	0.000
Thursday	$(1,0,0)(0,1,1)_{24}$	0.519	0.045	0.000	0.844	0.039	0.000
Friday	$(1,0,0)(0,1,1)_{24}$	0.519	0.045	0.000	0.834	0.040	0.000
Saturday	$(1,0,0)(0,1,1)_{24}$	0.462	0.045	0.000	0.898	0.029	0.000
Sunday	$(1,0,0)(0,1,1)_{24}$	0.280	0.049	0.000	0.876	0.041	0.000

there is no significant correlation between the residuals. Second, the normal probability of the residuals, which is a graph of the cumulative distribution for the residual data, normally appears as a straight line when plotted on normal probability paper. Figure 3 shows a normal probability plot where the residuals look fairly linear so the normality assumptions of the residuals hold (Box & Jenkins 1976). Third, the histograms of the residuals also show that the residuals for the 2 days are normally distributed, which again indicates they consist of white noise. Lastly, plotting the residuals against the forecast values shows them to be quite evenly distributed around the mean, which also confirms that the selected models are adequate (Govindaswamy 1991).

The modified Ljung-Box-Pierce statistic (p) was checked and confirmed that the present models are adequate for the available data at a 5% significance level. The values of the

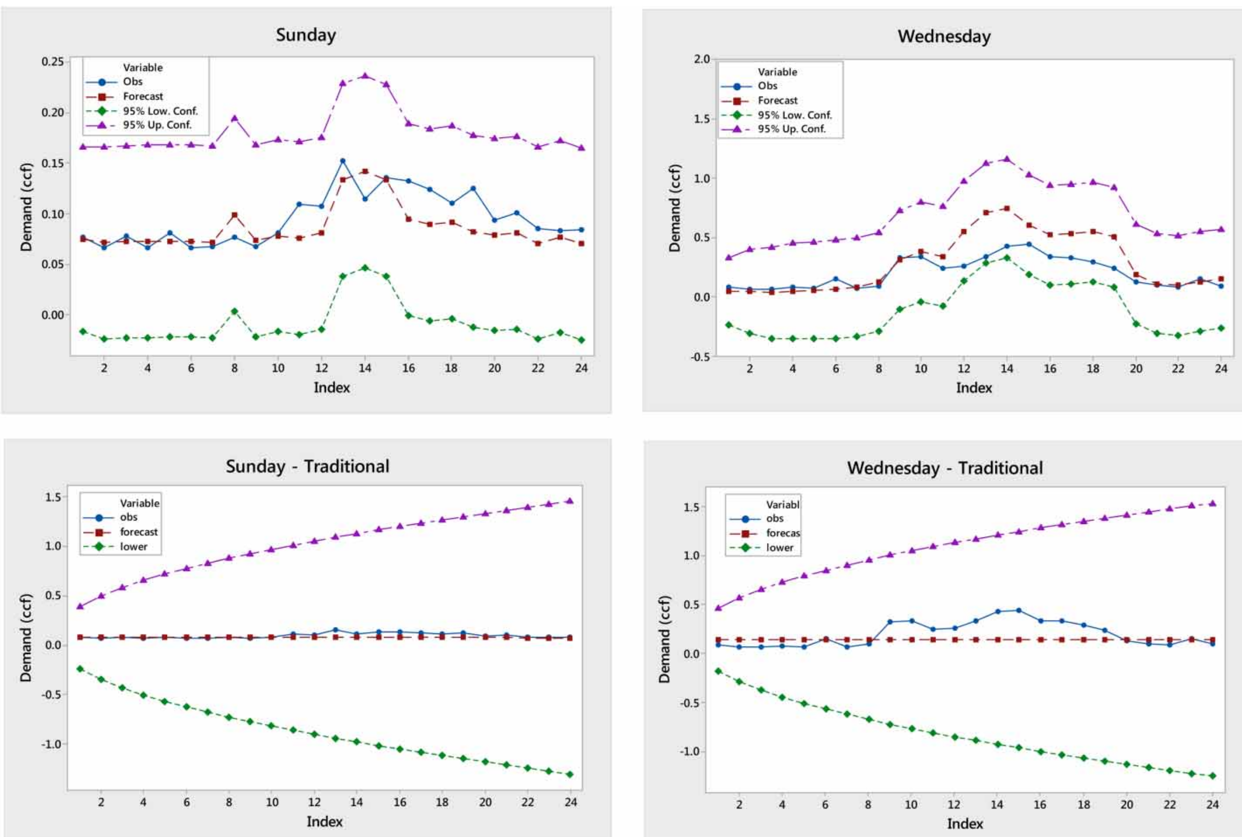

Figure 4 | Comparison of observed and forecast values for variant and traditional ARIMA.

Table 2 | Root mean square for errors for ARIMA variant and traditional

	ARIMA variant	Traditional ARIMA
Sunday	0.020	0.031
Monday	0.212	0.382
Tuesday	0.089	0.194
Wednesday	0.157	0.138
Thursday	0.184	0.336
Friday	0.094	0.267
Saturday	0.036	0.041

parameters, associated standard errors, and *P*-values are listed in Table 1. The standard errors calculated for the model parameters are generally small compared to the parameter values so most of the parameter estimates are statistically significant and these parameters should be used in the models.

The forecast was performed for a 24-hour lead-time; the data set from the first week of March to the third week of June 2014 was used for validating the model and the test data were for the last week of June 2014. The plot between observed data and predicted data, with a 95% confidence interval, is shown in Figure 4. As the figure shows, the predicted data follow the observed data quite closely. To check the performance of the ARIMA variant which was used in this paper, root mean square for errors (RMSE) were computed for both variant and traditional ARIMA. For traditional ARIMA, whole data sets are utilized to forecast the following 24-hour lead-time. For example, Wednesday June 25th, 2014 data were forecast using March 5th to June 24th, 2014. RMSE results are shown in Table 2 and forecasting graphs are shown in Figure 4. It is found that magnitudes of RMSE of the variant version are relatively smaller or similar to the traditional version.

CONCLUSIONS

Forecasting hourly water demand is of utmost importance for an MWG to ensure optimal planning and operation of the water systems. To do so, the authors developed ARIMA models for weekday-based hourly water demand and confirmed that the new models provided a reasonably good fit to the observed data. It is therefore recommended that linear stochastic models be used for this purpose for MWGs at an hourly time scale to identify future water demand patterns. These new stochastic models can be used for the development of an MWG pilot project to ensure effective sustainable water resource planning for green buildings.

ACKNOWLEDGEMENT

The authors greatly appreciate the grant (12-TI-C01) received from the Advanced Water Management Research Program of the Ministry of Land, Infrastructure and Transport in the South Korean Government.

REFERENCES

Arandia-Perez, E., Uber, J., Shang, F., Boccelli, D., Janke, R., Hartman, D. & Lee, Y. 2009 Preliminary spatial-temporal statistical analysis of hourly water demand at household level. In: *World Environmental and Water Resources Congress 2009, 17–21 May*, Kansas City, MO, USA, pp. 1–15.

Box, G. E. P. & Jenkins, G. M. 1976 *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco, CA, USA.

Cassidy, R. (ed.) 2003 *White Paper on Sustainability: A Report on the Green Building Movement*. A supplement to Building Design & Construction November. http://www.usgbc.org/DisplayPage.aspx?CMSPageID=78& (accessed 24 February 2014).

Dziegielewski, B., Kiefer, J., Opitz, E., Porter, G., Lantz, G., DeOreo, W., Mayer, P. & Nelson, J. 2000 *Commercial and Institutional End Uses of Water*. American Water Works Association Research Foundation, Denver, Co, USA.

Govindaswamy, R. 1991 Univariate Box-Jenkins forecasts of water discharge in Missouri River. *Water Resour. Dev.* **7** (3), 168–177.

Mishra, A. K. & Desai, V. R. 2005 Drought forecasting using stochastic models. *Stoch. Environ. Res. Risk Assess.* **19**, 326–339.

Shvartser, L., Shamir, U. & Feldman, M. 1993 Forecasting hourly water demands by pattern recognition approach. *J. Water Resour. Plann. Manage.* **119** (6), 611–627.

UGBC 2006 *LEED Policy Manual. Foundations of the Leadership in Energy and Environmental Design Environmental Rating System: A Tool for Market Transformation*. US Green Building Council, Washington, DC, USA. http://www.usgbc.org/ShowFile.aspx?DocumentID=2039 (accessed 8 February 2014).

US National Academy of Engineering 2008 *Grand Challenges for Engineering*. National Academy of Engineering, Washington, DC, USA. http://www.engineeringchallenges.org/ (accessed 8 February 2014).

First received 28 October 2014; accepted in revised form 2 July 2015. Available online 3 August 2015