Short-Term Forecasting for Urban Water Consumption

Alaa H. Aly, M.ASCE,[1] and Nisai Wanakule, M.ASCE[2]

Abstract: An approach is presented for short-term (i.e., daily and monthly) forecasting of municipal water use that utilizes a deterministic smoothing algorithm to predict monthly water use. The smoothing algorithm considers level, trend, and seasonality components of the time series. Daily deviations from the monthly average are then forecasted for up to six days using autocorrelation and weather dependence. While providing accurate operational forecasts, the approach required about six years of daily data to develop and validate the models. The approach is applied and evaluated for a number of municipalities near Tampa, Fla. Results show that the approach provides accurate daily forecasts as measured using a validation period of about three years.

DOI: 10.1061/(ASCE)0733-9496(2004)130:5(405)

CE Database subject headings: Water demand; Urban areas; Forecasting; Time series analysis; Regression models.

Introduction

Water use varies widely over short periods of time; such use provides operational challenges for public water suppliers. Several factors combine to explain this variability including, seasonal/climatic factors (e.g., winter versus summer, tourist population, etc.), daily weather factors (e.g., precipitation, temperature, and relative humidity), and day-to-day activities (e.g., weekday versus weekend) with each factor affecting different scales of time. Most of the approaches in the literature on forecasting urban water use deal with longer time periods (monthly to annual). However, daily forecasts are often required for municipal systems with little or no storage. Furthermore, daily forecasts can be useful for operational planning purposes (e.g., maintenance scheduling, treatment chemicals inventory management, etc.).

Several approaches have been used to forecast urban water use. Saleba (1985) classified these into three major categories: end use forecasting, econometric forecasting, and time series forecasting. End use forecasting utilizes specific water use data for the population of interest. While requiring large amounts of data, this approach has shown little success in application. This lack of success has led many researchers to econometric and time series forecasting approaches. Econometric forecasting is based on the identification of historical relationships (econometric models) between water use (response variable) and socioeconomic and environmental factors (explanatory variables). Once these relationships are established, they are assumed to hold into the future and are used to forecast water use based on forecasts of the explanatory variables. End use forecasting and econometric forecasting are particularly unreliable for short-term forecasting due to the effects of daily weather on water use. Finally, the most-widely used time series forecasting depends on the direct identification of patterns existing in the water use data.

Presented here is a framework for short-term forecasting of urban water use. A two-step approach is employed whereby monthly forecasts are first developed using an adaptive exponential smoothing algorithm. This is followed by development of daily forecasts using linear regression for up to six days lead time. Details of the proposed two-step approach and a description of the algorithms are presented, along with application results and discussion of some of the implementation details. Recommendations for implementation and for further improvement opportunities are also provided.

Short-Term Water Use Forecasting

Few studies can be found in the literature for short-term forecasting of urban water use. For example, Maidment and Miaou (1986) utilized a time series approach to study daily water use at nine cities in the United States. Their approach defined base use as the weather-insensitive part of water use which they estimated by averaging winter use. Weather-sensitive seasonal use was defined as the deviation of actual use from base use, and the relationships between seasonal use and weather factors were then analyzed. The most important weather factor was found to be the number of days since the last precipitation event, but other relevant weather factors included precipitation amount (spatially averaged) and air temperature. Application results displayed a wide range of performance, with correlation between forecasted and recorded daily water use values ranging from 0.99 for Dallas to 0.67 for Allentown, Pa.

Shvarster et al. (1993) presented a model for forecasting hourly water use based on time-series analyses and pattern recognition. They fitted low-order auto-regressive integrated moving-average (ARIMA) models to each system state (i.e., rising, oscillating, and falling), and then modeled the transition from each state to the next as a Markov process. These models produced reasonable results in application but the models required continuous (hourly) monitoring of water use. Unfortunately, these data are not available for most municipalities.

Homwongs et al. (1994) developed an adaptive smoothing-filtering approach to forecast hourly municipal water use. Their

[1]Senior Engineer, INTERA Incorporated, 137 2nd. Ave., Suite 200, Niwot, CO 80544. E-mail: aaly@intera.com

[2]Senior Engineer, Tampa Bay Water, 2535 Landmark Dr., Suite 211, Clearwater, FL 33761. E-mail: nwanakule@tampabaywater.org

Note. Discussion open until February 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on November 18, 2002; approved on August 4, 2003. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 130, No. 5, September 1, 2004. ©ASCE, ISSN 0733-9496/2004/5-405–410/$18.00.

J. Water Resour. Plann. Manage. 2004.130:405-410.

methodology decomposed the hourly time-series signal into three components: (1) Cyclical (unknown frequency) or seasonal (known frequency), (2) trend (a monotonic function of time), and (3) irregular (random). Similar to the model presented by Shvarster et al. (1993), the results indicated that reasonably accurate forecasts could be generated for 24 hours. The authors did not investigate the extension of their models to forecast beyond the next 24 hours.

Kenward and Howard (1998) developed a two-step approach to forecasting daily water use. In the first step, monthly average water use is forecasted using an exponential smoothing algorithm. In the second step, the monthly average forecast is perturbed to generate a daily forecast. The perturbations are based on meteorological factors and the monthly average value. The work presented in this paper is an extension of this methodology with the main difference being in considering the short-term memory of the daily water use time-series. The addition of these memory terms significantly improved the forecasting accuracy for daily water use. Furthermore, the gained accuracy allowed accurate forecasting for several days into the future (up to six).

Zhou et al. (2000) applied models similar to those proposed by Maidment and Miaou (1986) to forecast daily water use for Melbourne, Australia. Furthermore, these models predicted water use per capita, which are not useful for areas with a high tourist population.

Forecasting Algorithm

Let $u_{t,k}$ be daily water use during day t in month k and m_k be average monthly water use during month k. The developed approach employs two steps. In the first step, a smoothing algorithm is used to forecast m_k based on seasonality and trend. Second, the deviation of $u_{t,k}$ from m_k is forecasted using a time-series regression model based on autocorrelation and weather dependence. That is:

$$u_{t,k}=m_k+(u_{t,k}-m_k)=m_k+d_{t,k} \tag{1}$$

In the first step, an exponential smoothing algorithm is used to forecast m_k. The idea is to use the monthly average value to filter out the level (usually depends on the specific application location), trend (usually depends on population dynamics), and seasonality (usually depends on weather and population dynamics). Exponential smoothing algorithms are based on three smoothing steps; overall smoothing, trend smoothing, and seasonal smoothing. The trend could be modeled using linear, exponential or saturated growth models. Seasonality could either be additive or multiplicative. Fig. 1 shows the potential combinations of trend and seasonality. For example, the Winters' method assumes multiplicative seasonality and a linear trend. This method could be written as (Makridakis et al. 1998):

Level smoothing

$$L_k=\alpha\,\frac{m_k}{S_{k-s}}+(1-\alpha)(L_{k-1}+b_{k-1}) \tag{2}$$

Linear trend smoothing

$$b_k=\beta(L_k-L_{k-1})+(1-\beta)b_{k-1} \tag{3}$$

Multiplicative seasonal smoothing

$$S_k=\gamma\,\frac{m_k}{L_k}+(1-\gamma)S_{k-s} \tag{4}$$

Monthly forecast

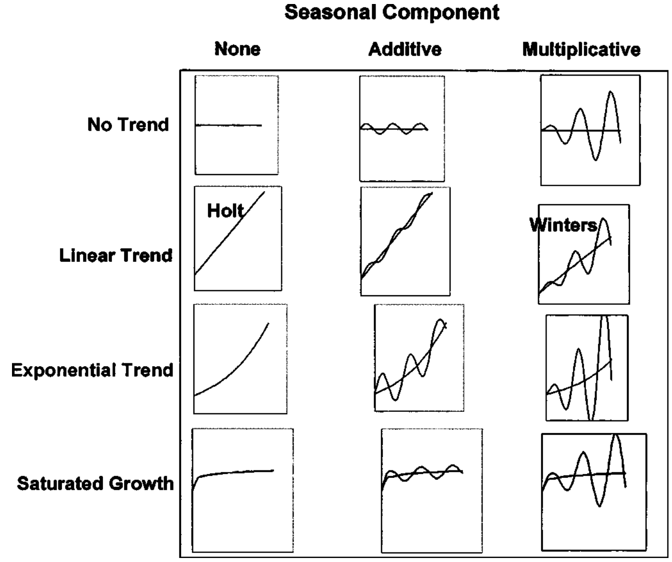

Fig. 1. Classification of exponential smoothing techniques

$$F_{k+n}=(L_k+b_kn)S_{k-s+n} \tag{5}$$

where s=length of seasonality ($s=12$ months); L_k=deseasonalized level value (local mean) of the time series; b_k=trend (average difference between water use value in the current month and that of the previous month); S_k=seasonal factor (average ratio between water use value in the current month and that of the previous month); and F_{k+n}=forecasted water use value for n months ahead. Note that the level, trend, and seasonal factor values are updated whenever a new value is recorded for the time series. Their value will be an average value of the most recent observation and the previous average value. The factors (α, β, and γ) determine how much memory is retained during the updates. When any of these factors has a value of one, the respective component will have no memory and its value will simply be the most recent observation. For positive values smaller than one, the smaller the value of the factor, the more the memory retained in the system. The values of α, β, and γ are determined by fitting the models to the observed monthly average values. Criteria such as average error and root mean square error are usually employed. We found that a better statistic to minimize during the fitting is the Theil's U which is expressed as (Theil 1966):

$$U=\sqrt{\sum_{k=1}^{N-1}\left(\frac{F_{k+1}-m_{k+1}}{m_k}\right)^2\Big/\sum_{k=1}^{N-1}\left(\frac{m_k-m_{k+1}}{m_k}\right)^2} \tag{6}$$

Theil's U statistic measures relative performance of the forecasting algorithm as compared to a naive forecast (i.e., next month forecast equals current month average value). A value of one indicates a forecast that is as good as the naive forecast and a value of zero indicates a perfect forecasting algorithm (the forecasted value is always equal to the observed value). Obviously, a forecasting algorithm that achieves a minimum value for U is desirable. The only disadvantage for U is that zero values are not allowed for the monthly average time series. This is usually not a problem for water use values. However, for cases where the time series can assume zero values, either an exponential transformation or other evaluation criteria (e.g., root mean square) could be used.

To illustrate the effectiveness of the Theil's U statistic, consider a time series with a strong first-order autocorrelation. A

J. Water Resour. Plann. Manage. 2004.130:405-410.

forecast generated using a naive estimator will have a high correlation coefficient (equal to the first-order autocorrelation coefficient) but a Theil's U value of one. In practice, it is common for the first-order correlation coefficient to be close to 0.90. An algorithm that generates a forecast with more accuracy might not provide a significant improvement over the naive model as measured by the correlation coefficient. However, an algorithm that can reduce the Theil's U to, say, 0.5 presents quite an improvement over the naive model.

The most significant advantage of exponential smoothing algorithms is their adaptive nature. As long as the underlying models (i.e., linear trend and multiplicative seasonality) do not change, the exponential smoothing algorithm will continue to provide forecasts with the same accuracy over time. In practice, as with any empirical model, these models should be evaluated frequently (every two to three years) to ensure their applicability has not changed.

The exponential smoothing algorithm is not necessarily unbiased. That is, model forecasts can display a linear bias when compared with the observed monthly average values. This could be easily eliminated by fitting a simple linear regression line between observed monthly average values and the forecasts calculated using the exponential smoothing algorithm. In other words, the final forecast is obtained by applying the linear correction to the results of the smoothing algorithm as shown in

$$M_k = a + bF_k \tag{7}$$

where a and b are obtained by fitting the smoothing algorithm forecasts (F_k) to observed monthly average values.

The second step in our forecasting algorithm is to forecast the daily deviation from the forecasted monthly average [defined in Eq. (1)] using two sources of information, the recent memory of the time-series and the dependence of water use on weather conditions. The general model equation is:

$$d_{t,k} = f(d_{t-1,k}, d_{t-7,k}, P_{t,k}, T_{t,k}, RH_{t,k}) + \varepsilon \tag{8}$$

where $d_{t,K}$=tth daily deviation from average kth monthly value; $d_{t-1,K}$=previous similar one-day lagged daily deviation; $d_{t-7,k}$=observed daily deviation one week ago; $f(\cdot)$=general linear function; $P_{t,k}$, $T_{t,k}$, and $RH_{t,k}$=variables representing recent history (prior to day t) for precipitation, temperature, and relative humidity, respectively; and ε=random component. To fit the linear regression models, we used a typical least-squares loss function. Weather and water usage data were used to build the models. Each of the independent variables was also transformed to test for increased predictive power. The variable transformations were mostly of the Box-Cox forms including log functions, squares and other power transformation (with exponents greater than and less than one), and absolute values. The transformations that performed well across a number of daily water use time-series are listed in Table 1.

Table 1. Summary of Variables and Transformations Used in Daily Forecasts

Variable	Transformation	Regression coefficient	Standard error of regression coefficient
Intercept	none	0.232	0.020
Last similar-day lagged daily variance	none	0.132	0.033
Last similar-day lagged daily variance	X abs(X)	0.072	0.011
One-day lagged daily variance	none	0.726	0.033
One-day lagged daily variance	X abs(X)	−0.123	0.011
Three-day average precipitation	log(1+X)	−0.216	0.023
Number of rainless days	log(1+X)	0.096	0.036
Three-day average daily temperature	log(X)	1.153	0.174
Daily average relative humidity	log(X)	−1.164	0.202

Application

Tampa Bay Water is the largest wholesale water supplier in Florida, serving more than 2 million people with an annual average use of 10.82 m^3/s or 247 million gallons per day. Its Member Governments consist of Hillsborough, Pasco and Pinellas Counties and the Cities of New Port Richey, Tampa, and St. Petersburg. The mission of the Agency is to provide its members with reliable supplies of high-quality water to meet present and future needs in an environmentally and economically sound manner.

Tampa Bay Water owns 12 groundwater supply systems, which are operated in accordance with permits issued by the Southwest Florida Water Management District (SWFWMD). The 11 wellfields under the Consolidated Water Use Permit are operated as an integrated system utilizing a set of simulation-optimization models, which prioritize minimization of environmental impacts while reliably meeting demands. Wellfield production is scheduled by the Optimized Regional Operations Plan (OROP) (Tampa Bay Water 2002). The OROP has been implemented since January 1999 to generate a weekly pumping schedule for 175 wells within eleven wellfields.

The objective of the OROP is to maximize water levels in a set of monitor wells (currently consisting of 34 surifical aquifer wells and two Upper Floridan wells); these wells are used as surrogate health indicators for wetlands in the region. The OROP is a classic resource allocation optimization model where one of the primary constraints is to meet weekly projected demands at ten Member Government's Points of Connection, which are delivery points for wholesale water. Fig. 2 is a schematic diagram of the interconnected "Regional System" between water supply sources and Points of Connection (POCs).

Tampa Bay Water has an increasing need to develop a defensible short-term water use forecasting model to replace the current approach where projected use at each POC has been estimated based on the previous week's use and adjustments based on near weather forecast and experience. This estimate cannot be reproduced, and no measurable variability in prediction is possible. In addition, a more rigorous and quantifiable approach to short term use forecasting is needed to support the Agency's future plans to implement stochastic optimization, and increase the forecasting frequency from one week to one day. Results of the short term forecasting model for the Moon Lake Water Treatment Plant point of connection are described in the next section.

The first step of the applied methodology is to forecast monthly average use. Winters' method coefficients were calibrated to minimize Theil's U using the first six years of data. The calibrated coefficients are 0.81, 0.01, and 0.09 for α, β, and γ, respectively. Large values for any of these coefficients (close to one) indicate a strong memory component while small values (close to zero) indicate less memory and more dependence on

J. Water Resour. Plann. Manage. 2004.130:405-410.

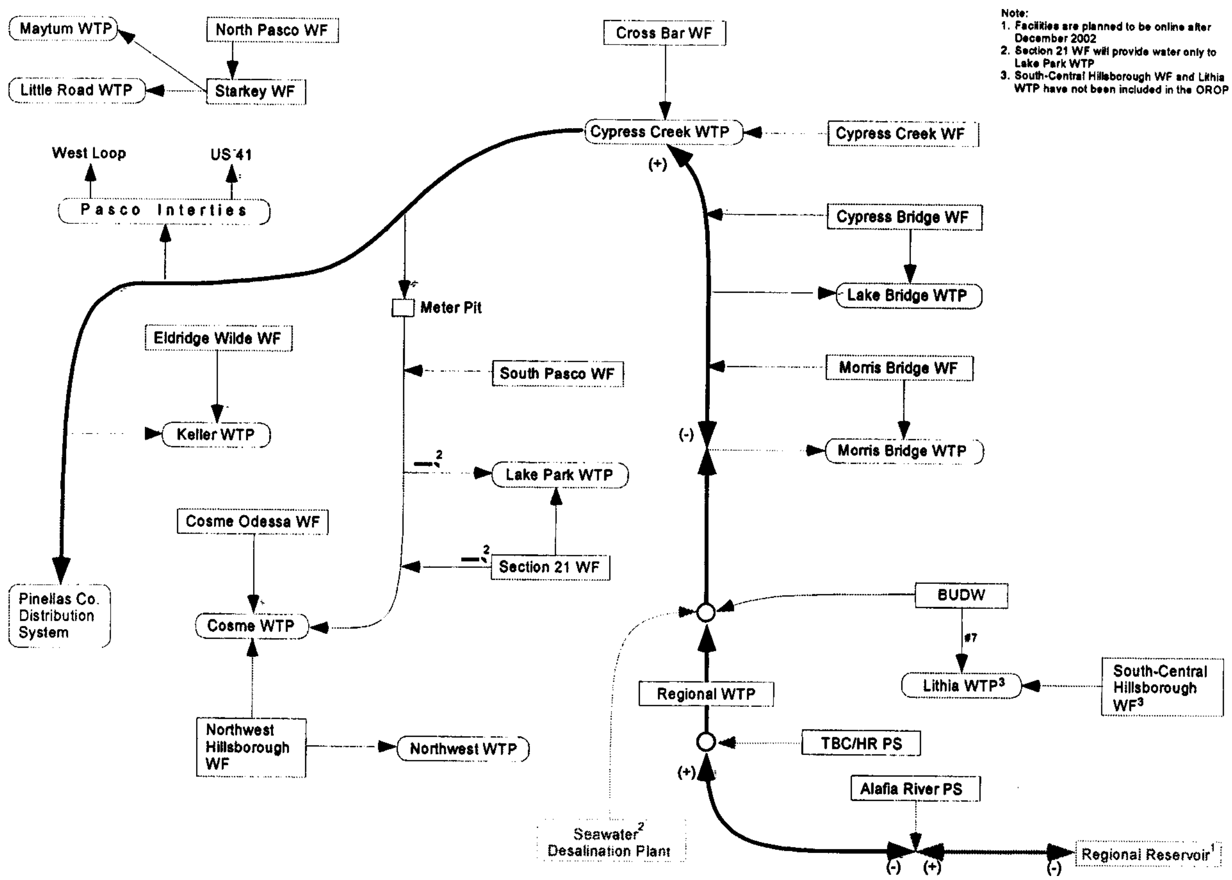

Fig. 2. Schematic diagram of the interconnected regional system

recent history. Figs. 3 and 4 show a comparison between observed and forecasted monthly values of water use. With a simple correlation coefficient of 0.83 and Theil's U of 0.54, the model captures the seasonal and trend components reasonably. Due to the limited data base, only the first six years of data were used for model initialization (estimating seasonal factors and model parameters). The next two years were used to verify the model predictions. The model predictions between January, 1997 and December, 1999 are indicative of the model's predictive power.

After an accurate model was established for average monthly values, daily deviations from monthly average water use were evaluated using regression analysis based on autocorrelation and correlation with meteorological variables. A multiple regression model was formulated using explanatory variables derived from recent daily water use values and recent weather. All the considered variables were checked for correlation with daily deviations using stepwise regression and a series of Box-Cox transformations. In general, daily water use deviations were much more correlated with short-term moving averages (three-day) of use than with seven-day moving averages or daily values for meteorological variables. Similarly, daily water use deviations' autocorrelation was much more pronounced for short lags (one and seven days) than for long lags (one year). Using the first six years of data for model fitting, the final regression model had a correlation coefficient of 0.86 with a Theil's U value of about 0.69. Table 1 shows the fitted regression coefficients and their standard errors.

Model prediction accuracy during the validation period (January 1997 through December 1999), was very similar to that of the

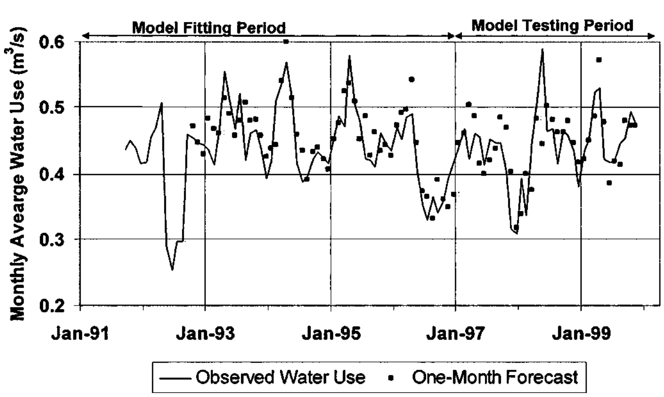

Fig. 3. Observed and forecasted monthly average water use at Moon Lake WTP

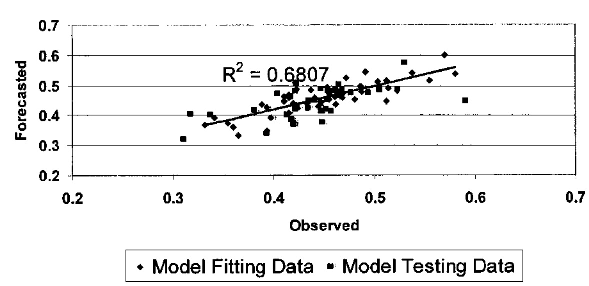

Fig. 4. Scatter plot of observed and forecasted monthly average water use values

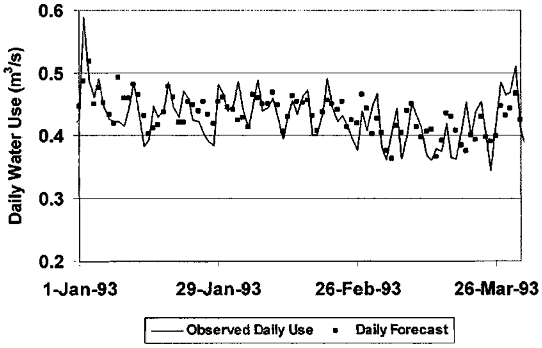

Fig. 5. Observed and forecasted daily water use (first quarter of 1993)

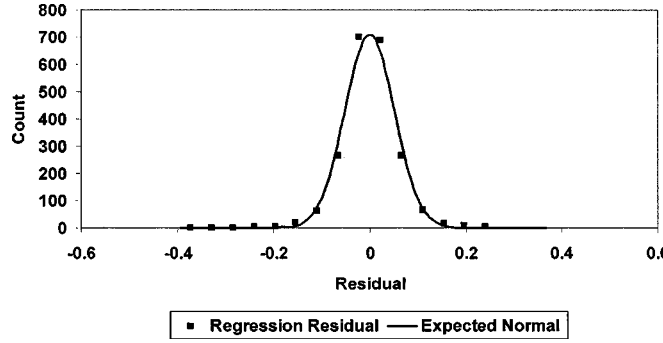

Fig. 7. Residual distribution and expected normal density function for regression analysis of daily water use values

fitting period. The daily forecasting model explains about 70% of the variability of daily use values. The unexplained 30% is assumed to be random and could only be forecasted in a stochastic sense. Clearly, a random component is present for daily use values and no highly accurate predictions should be expected from a deterministic model. The motivation for the above methodology is that this random component should be eliminated when monthly average values were calculated. Furthermore, using meteorological variables for the daily model allows a stochastic evaluation of forecasted water use using an ensemble of possible meteorological conditions. Figs. 5 and 6 show observed and forecasted daily water use values for the entire record. Consistent with the underlying statistical model, Fig. 7 shows that the residuals from the regression model follow a normal distribution.

To forecast stochastic use values, several ensembles of possible future meteorological values could be utilized. For each ensemble, one set of forecasted water use is generated. The empirical distribution function of future daily water use is then constructed by sorting the forecasted use values for each day. The methodology does not depend on the specific forecasting technique used to generate possible future meteorological values. To illustrate the approach, daily meteorological values recorded between 1984 and 1998 are used to predict daily values for 1999 (relative humidity data was not available before 1984 at the utilized weather station). Fig. 8 shows minimum, maximum, and average forecasted daily use for 1999. Longer weather records or synthetically generated records could be used to evaluate the entire distribution function or a specific percentile.

Conclusions

The outlined approach provides a framework for accurate forecasting of municipal water use for both monthly and daily values.

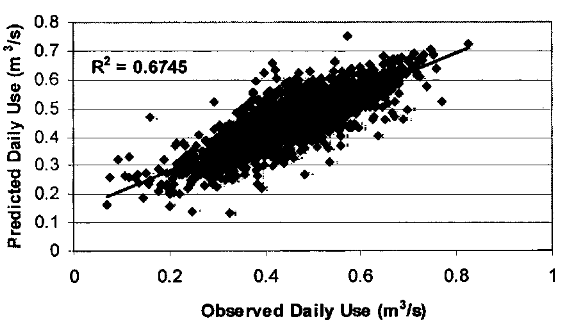

Fig. 6. Scatter plot of observed and forecasted daily water use values

The two-step approach provides an estimate for average monthly water use which is used as a datum for forecasting daily deviations from the monthly average. A highly accurate monthly forecast is developed using an adaptive exponential smoothing algorithm. The accuracy of such algorithm is much higher for shorter time periods (one or two months) than for longer forecasts. Daily values forecasted as deviations from the monthly average are obtained using linear regression models for up to six days into the future. Two sets of variables are used in the linear regression models representing the memory properties of the time series and water use dependence of weather factors. Results show that the time-series memory is much more pronounced than the weather dependence. However, in terms of weather dependence the results were similar to what others have concluded, i.e., that the number of days since the last precipitation event is the most significant weather-related factor. As could be expected, the forecasting accuracy decreases as the number of future days increases, but results show that the deterioration is not significant.

Extensive application results show that the developed forecasting algorithms seem to provide highly accurate monthly average forecasts as well as reasonably accurate daily forecast for six days into the future. However, the developed models are empirical in nature. These models, in general, assume that recent historic patterns are valid in the near future. Regular evaluation of this assumption will be necessary. The preferred approach is to use the developed models and evaluate their performance. If their performance becomes questionable, the first question will be whether parameter adjustments can help alleviate the problem. If param-

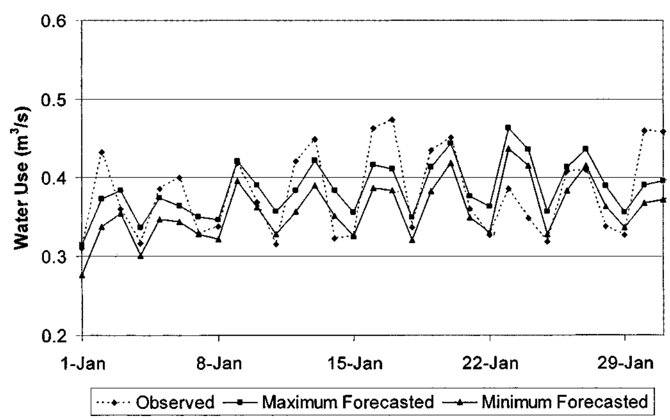

Fig. 8. Maximum and minimum forecasted daily water use in January 1999

eter adjustments are not sufficient, the next question will be whether the underlying phenomena are changing. In this case, either different models will be needed or new algorithms must be developed and evaluated. At least two or three extra years of data should be available before a meaningful evaluation of the monthly forecasting models could be performed. However, daily models could be reasonably evaluated after a relatively short time period (e.g., 6–9 months).

Notation

The following symbols are used in this paper:

b_k = trend value of the monthly average time series;
$d_{t,K}$ = tth daily deviation from average kth monthly value;
F_{k+n} = forecasted water consumption value for n months ahead;
L_k = deseasonalized level value of the monthly average time series;
m_k = average monthly water consumption during month k;
$P_{t,k}$ = precipitation depth during day t in month k;
$RH_{t,k}$ = relative humidity during day t in month k;
S_k = seasonal factor of the monthly average time series;
s = length of seasonality;
$T_{t,k}$ = average temperature during day t in month k;
U = Theil's U statistic;
$u_{t,K}$ = daily water consumption during day t in month k;
α = smoothing factor for level value;
β = smoothing factor for trend value;
γ = smoothing factor for seasonality value; and
ε = a random component.

References

Homwongs, C., Sastri, T., and Foster, III, J. W. (1994). "Adaptive forecasting of hourly municipal water consumption." *J. Water Resour. Plan. Manage.*, 120(6), 888–905.

Kenward, T. C., and Howard, C. D. D. (1998). "Forecasting for urban water demand management." *Proc., 26th Annual ASCE Conference on Water Resources Planning and Management*, Reston, Va., Chap. 9A188.

Maidment, D. R., and Miaou, S. P. (1986). "Daily water use in nine cities." *Water Resour. Res.*, 22(6), 845–851.

Makridakis, S., Wheelright, S. C., and Hyndman, R. J. (1998). *Forecasting: Methods and applications*, Wiley, New York.

Saleba, G. S. (1985). "Water demand forecasting." *Proc., AWWA Seminar on Demand Forecasting and Financial Risk Assessment*, Denver.

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

Tampa Bay Water. (2002). *Optimized Regional Operations Plan Annual Rep.*, prepared for Southwest Florida Water Management District, Brooksville, Fla.

Theil, H. (1966). *Applied economic forecasting*, North-Holland, Amsterdam, The Netherlands.

Zhou, S. L., McMahon, T. A., Walton, A., and Lewis, J. (2000). "Forecasting daily urban water demand: A case study of Melbourne." *J. Hydrol.*, 236, 153–164.

J. Water Resour. Plann. Manage. 2004.130:405-410.