WATER RESOURCES BULLETIN
VOL. 22, NO. 4 AMERICAN WATER RESOURCES ASSOCIATION AUGUST 1986

AN EVALUATION OF WEEKLY AND MONTHLY TIME SERIES FORECASTS OF MUNICIPAL WATER USE[1]

Sheryl L. Franklin and David R. Maidment[2]

ABSTRACT: A cascade model for forecasting municipal water use one week or one month ahead, conditioned on rainfall estimates, is presented and evaluated. The model comprises four components: long term trend, seasonal cycle, autocorrelation and correlation with rainfall. The increased forecast accuracy obtained by the addition of each component is evaluated. The City of Deerfield Beach, Florida, is used as the application example with the calibration period from 1976-1980 and the forecast period the drought year of 1981. Forecast accuracy is measured by the average absolute relative error (AARE, the average absolute value of the difference between actual and forecasted use, divided by the actual use). A benchmark forecast is calculated by assuming that water use for a given week or month in 1981 is the same as the average for the corresponding period from 1976 to 1980. This method produces an AARE of 14.6 percent for one step ahead forecasts of monthly data and 15.8 percent for weekly data. A cascade model using trend, seasonality and autocorrelation produces forecasts with AARE of about 12 percent for both monthly and weekly data while adding a linear relationship of water use and rainfall reduces the AARE to 8 percent in both cases if it is assumed that rainfall is known during the forecast period. Simple rainfall predictions do not increase the forecast accuracy for water use so the major utility of relating water use and rainfall lies in forecasting various possible water use sequences conditioned on sequences of historical rainfall data.
(KEY TERMS: statistics; time series analysis, municipal water use; forecasting.)

INTRODUCTION

As a city's water demand comes close to or exceeds its reliable supply, city water managers are faced with the need for forecasts of short term demand. Some relevant issues include (1) what is the minimum error of the forecasts, (2) is the additional forecast accuracy of increasingly complex methodologies worth the increased modeling effort, and (3) how can the variability of future water use be determined? These issues are examined using a cascade model to forecast water use on both a monthly and a weekly time step, conditioned on various assumptions regarding future rainfall sequences. The model parameters are determined by statistical analysis of past records of city water use, population and rainfall.

The model contains four components, two *long-term* components reflecting the slowly changing characteristics of the city's water use pattern that result from population growth and seasonal variation, and two *short-term* components that adjust the forecast according to the water use observed immediately prior to the forecast period and an estimate of the rainfall expected during this period. Thus, the long-term components project the basic structure of the water use pattern over time, while the short-term components allow for the "tracking" of the current pattern of water use, a pattern which may reflect unusual circumstances not observed in past years. The methodology follows the format presented by Maidment and Parzen (1984a, 1984b). Although both monthly and weekly time steps were analyzed, only monthly results are reported in detail here. Details of the weekly analyses may be found in Franklin and Maidment (1983).

The City of Deerfield Beach, Florida, was chosen as a test city. Deerfield Beach had a 1981 population of approximately 43,000 and is located on the rapidly urbanizing lower east coast of Florida approximately 40 miles north of Miami. The variable modeled is the total volume of water treated by the city, i.e., "finished water," in each month. These data are reproduced, along with monthly rainfall, in Table 1. A total of 72 months of water use data were available; these were divided into two periods, a *modeling period* from January 1976 through December 1980 and a *forecasting period* from January 1981 through December 1981. The forecasting period encompassed a sharp, severe drought in South Florida that necessitated a mandatory 10 percent reduction in city water supply. The drought period is used as the forecasting period to test the robustness of the model. The model is able to forecast water use reasonably well, even during extreme weather conditions not present during the modeling period. Forecast error is measured by the average absolute relative error (AARE), defined as the average absolute value of the differences between each actual and forecasted use, divided by the actual

[1]Paper No. 86038 of the *Water Resources Bulletin*. Discussions are open until April 1, 1987.
[2]Respectively, Water Operations Manager, Brazos River Authority, P.O. Box 7555, Waco, Texas 76714-7555; and Associate Professor, University of Texas, Department of Civil Engineering, Austin, Texas 78712.

TABLE 1. Monthly Water Use and Rainfall, Deerfield Beach, Florida, January 1976 Through December 1981.

Year	January	February	March	April	May	June	July	August	September	October	November	December
					WATER USE (MILLION GALLONS PER MONTH)							
1976	183	190	214	235	160	143	241	216	201	200	202	170
1977	188	179	251	231	199	207	258	217	191	243	216	201
1978	196	189	229	237	220	204	228	200	200	180	173	192
1979	173	214	283	251	188	285	321	266	181	208	199	299
1980	259	240	312	240	295	256	259	292	265	287	245	267
1981	285	236	321	344	251	225	288	223	211	239	253	289
					RAINFALL (INCHES PER MONTH)							
1976	2.3	6.3	0.6	0.9	10.6	7.0	3.1	6.6	3.8	7.1	3.1	4.9
1977	4.0	1.3	0.1	3.3	9.4	6.2	3.1	9.1	8.0	5.1	6.6	3.6
1978	2.5	2.9	3.3	2.7	5.5	5.0	8.1	7.2	2.6	9.8	3.4	4.3
1979	2.1	0.9	0.6	14.6	5.1	1.2	0.6	5.2	8.6	6.6	8.9	2.3
1980	2.9	2.3	2.5	5.5	5.1	5.7	5.1	4.1	2.9	1.5	6.2	1.9
1981	0.6	2.9	1.0	0.3	5.0	4.4	4.7	11.2	8.4	5.6	3.2	0.1

use (see Equation 22). When using the actual month's rainfall to provide a measure of the forecasting accuracy inherent in the model, there is an 8.0 percent AARE between the observed water use and the full cascade model's one-month-ahead forecasts. Although actual rainfall is not available for forecast purposes, this condition is of value when using historical rainfall sequences to obtain a measure of expected water use variability for risk analysis.

Various schemes are employed to predict rainfall to test the model's forecasting ability in situations where future rainfall is unknown. When predicting rainfall to be the monthly average (assuming a return to normal conditions), the AARE is 11.7 percent; when predicting next month's rainfall excess or deficiency to be the previous month's excess or deficiency (assuming continuing relative weather conditions), the AARE is 13.6 percent.

The analysis is limited by the available data and the nature of the test city. The dependent variable is "water use," which is based on the city's water production – i.e., "use" from the regional water management district's viewpoint, not the actual citizen's consumption. The difference between the amount of water produced by the utility and the consumer's use is the system losses, which may result from unmetered usage such as fire fighting, or exfiltration from distribution lines. Since the system losses were fairly consistent over the study period, ranging from 13 percent to 19 percent, with an average of 16 percent (personal communication, Dale Holinbeck, City of Deerfield Beach), it is felt that there is no major inconsistency introduced by employing production rather than consumption data. Water use was not divided into categories of consumers such as residential, commercial and industrial. Deerfield Beach is relatively small (43,000 people) highly residential (approximately 85 percent of land area) and has no significant industrial or agricultural water use included in the data.

LITERATURE REVIEW

The earliest comprehensive study on modeling monthly water use as a time series is reported by Salas-LaCruz and Yevjevich (1972). They investigate municipal, agricultural and hydropower water uses and develop mathematical models for trends and periodicities in monthly means and standard deviations. They also investigate the coherence and cross-correlation functions of water use, rainfall and air temperature to determine the interrelationships among these three variables. The cities used to fit their model were all fairly large and are concentrated in the Western United States. The relative value of each component in explaining the water use is determined by the reduction in the data's variance attributable to that component. Depending on the city, trends (long term growth) accounted for 10-69 percent (average of 37 percent) of the variance in water use, periodicity (seasonal cycles) accounted for 24-79 percent (average of 54 percent) of the variance and residuals accounted for 3-12 percent (average of 7 percent) of the variance.

Oh and Yamauchi (1974) examine monthly water use in Honolulu, Hawaii. They show seasonal fluctuation in the demand from single family residences and in commercial districts and canning industries. Multi-family dwellings showed a constant water demand.

Maidment and Parzen (1984a) investigate monthly water use and its relationship to climatic variables in six cities in Texas, three in the semi-arid High Plains and three in humid East Texas. The water use data are detrended using regression against the explanatory variables of population, number of water connections, real average buying income per household and real marginal water price. Population is found the most significant variable in all cases where there is substantial population growth. The periodicities are removed using Fourier-fitted monthly means. A 12-month cycle is present in all cities, with various harmonics of six, four, three and two months present in various cities. The autoregressive nature of the series displays dependence on from one to three past values for cities with a smoothly varying pattern of water use and up to 13 past values in two cities with a water use pattern with erratic changes in usage. Climatic data are similarly modeled and a multiple regression analysis performed to relate water use to monthly average maximum temperature, monthly average minimum temperature, monthly pan evaporation and monthly precipitation. Precipitation is significant in all cases, with pan evaporation also significant for cities in the semi-arid areas. The residuals of this regression are examined and all water use residuals are determined to be normally distributed.

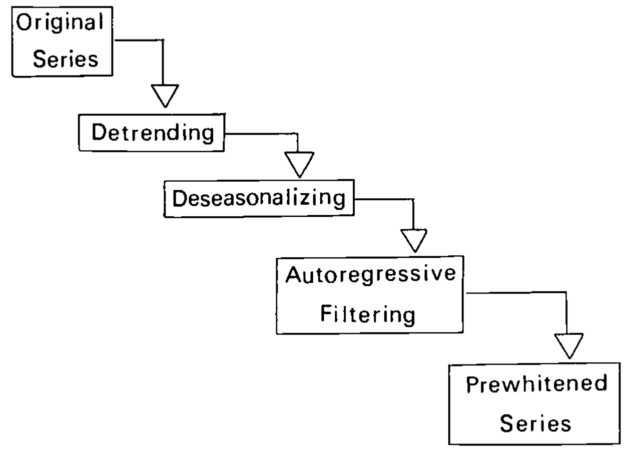

Figure 1. Cascade Model.

METHODOLOGY FOR DETERMINING MODEL PARAMETERS

Model

In the following discussion, W represents total water use; the subscripts a, b, c and d indicate the growth, seasonal, autoregressive and weather related use components, respectively. A carat, ^ , indicates estimated values.

The cascade model may be summarized as:

$$\hat{W}(t) = \hat{W}_a(t) + \hat{W}_b(t) + \hat{W}_c(t) + \hat{W}_d(t) \tag{1}$$

where t is a monthly time index.

Procedure for Analysis

The analysis stage of the model examines the data for the presence of identifiable characteristics in a series of analytical steps (Figure 1). Each step checks for the presence of a specific characteristic and eliminates it if present. If the data do not exhibit the trait, they are passed on to the next step unaltered. Because each step eliminates some influence in the data, the steps are referred to as filters. Since the output from one step is used as the input to the next, the entire approach is termed a cascade model. The residual series left after the cascade of transformations is white noise (zero mean, constant variance) and is called the "prewhitened series" (Maidment and Parzen, 1984a, 1984b).

The following procedure outlines the calculation of monthly parameter values. Weekly values are obtained similarly.

Step 1. Regress monthly water use against population.

$$\hat{W}_a(t) = \Phi_0 + \Phi_1 \hat{P}(t) + \Phi_2 \hat{P}^2(t) + \ldots \tag{2}$$

Step 2. Produce detrended monthly water use as the residual of this regression.

$$W_b(t) = W(t) - \hat{W}_a(t) \tag{3}$$

Step 3. Compute average monthly values of $\overline{W}_b(m)$: $\overline{W}_b(1), \overline{W}_b(2), \ldots \overline{W}_b(12)$ (by Fourier series or arithmetic average) and form a cyclic sequence of these values:

$$\hat{W}_b(m) = \hat{\overline{W}}_b(1), \hat{\overline{W}}_b(2), \ldots, \hat{\overline{W}}_b(12), \hat{\overline{W}}_b(1), \ldots \tag{4}$$

Step 4. Produce deseasonalized water use by subtracting the average monthly water use from the observed values.

$$W_c(t) = W_b(t) - \hat{W}_b(m) \tag{5}$$

Step 5. Fit an autoregressive model to $\hat{W}_c(t)$ to allow for "tracking" of unusual deviations from the long-term trend and seasonal variation.

$$\hat{W}_c(t) = \sum_{k=1}^{K} \alpha_k W_c(t-k) \tag{6}$$

Step 6. Produce an independently distributed residual water use series as the residual of this autoregressive model.

$$W_d(t) = W_c(t) - \hat{W}_c(t) \tag{7}$$

Step 7. Take observed monthly rainfall data and repeat steps 3 to 6 (if the rainfall exhibits no long-term trends).

$$\hat{R}_b(m) = \hat{\overline{R}}_b(1), \hat{\overline{R}}_b(2), \ldots, \hat{\overline{R}}_b(12), \hat{\overline{R}}_b(1), \ldots \tag{8}$$

$$R_c(t) = R_b(t) - \hat{R}_b(m) \tag{9}$$

$$\hat{R}_c(t) = \sum_{j=1}^{J} \gamma_j R_c(t-j) \tag{10}$$

$$R_d(t) = R_c(t) - \hat{R}_c(t) \tag{11}$$

Step 8. Determine the relationship between water use and rainfall by regressing the residual series from Equations (7) and (11).

$$\hat{W}_d(t) = \beta_0 + \beta_1 R_d(t) + \ldots \tag{12}$$

Step 9. Produce the final residual series of the rainfall-water use model, $W_e(t)$, and determine its mean, standard deviation and probability distribution in order to check the validity of the analysis.

$$W_e(t) = W_d(t) - \hat{W}_d(t) \tag{13}$$

Discussion of Procedure

Trend refers to any long-term influences which extend over a period of several years. Although these trends are most commonly fitted as low order polynomials, it is possible that long-term cyclic behavior may be present in the data. This behavior can be identified and modeled in a manner analogous to that described above and should not be confused with the annual variation in the data that is referred to as seasonality. This distinction in terminology permits the sequential estimation of the parameters of the trend and seasonal components. It is possible to simultaneously estimate the parameters of both components with linear regression by using dummy variables for the calculation of seasonal parameters, but this method requires the estimation of 12 seasonal parameters. This study utilized sequential estimation in order to apply Fourier analysis to the seasonal data, resulting in the estimation of fewer (four) seasonal parameters.

In relating the two prewhitened residual series, $\hat{W}_d$ and R_d (Equation 12), departures from normal water use are being related to departures from normal rainfall. It is necessary to use the residual series, not the original series, since the technique used to find the relationship (regression) assumes the model errors are independent. If actual water use or rainfall were used, the autocorrelation present in the data may provide a correlated error structure, particularly with weekly data, and thus invalidate inferences on the regression model parameters.

DETERMINATION OF MODEL PARAMETERS FOR DEER FIELD BEACH

The city of Deerfield Beach, Florida, was chosen for this study because of its history of cooperation with outside agencies, its convenient size, its complete and accessible records, and the fact that the city is solely and completely supplied by one water plant, thus ensuring that the community's water demand is accurately reflected in the water plant's production records. There are no households utilizing private wells for drinking water; however, no records are kept on the number of private wells used strictly for lawn irrigation purposes, of which there are a small number. It is known that a few homeowners irrigate lawns from the canals at the rear of their lots.

Both the rainfall and water use series for 1976-80 were passed through the cascade filters to determine their characteristics. These results are presented below. A detailed example of the forecasting procedure follows. The original data for water use and rainfall are reported in Table 1 and displayed in Figures 2 and 3.

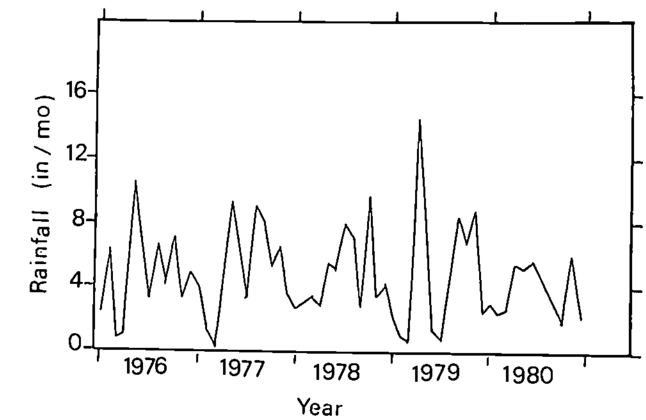

Figure 2. Monthly Rainfall, Deerfield Beach Florida, 1976-1980.

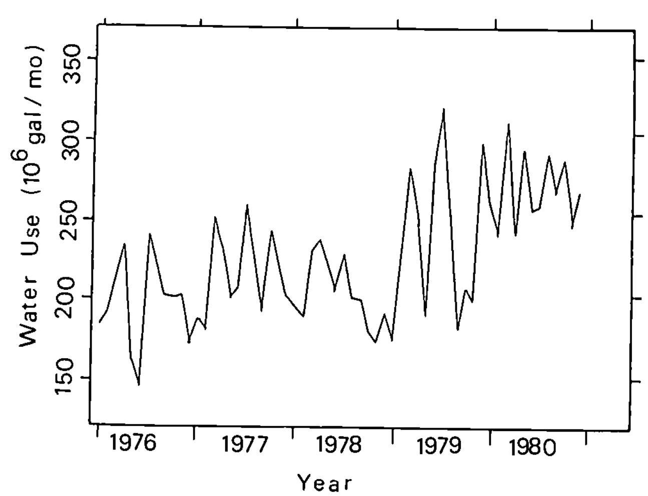

Figure 3. Monthly Water Use, Deerfield Beach, Florida, 1976-1980.

Trend

No long-term trends were apparent in the five years of rainfall data (Figure 2). This does not preclude the possibility of wet and dry cycles with a period of greater than five years, but such cycles are not of concern for the short-term forecasts considered here. A linear long-term trend in water use is evident (Figure 3). Although the trend could have been obtained by regressing water use directly on time, regression on population is chosen to provide information which can be related to commonly reported values of per capita water use and population forecasts.

Population figures for Deerfield Beach are compiled by several sources for differing purposes. The University of Florida compiles estimates which the state of Florida uses for distribution of funds. Broward County compiles estimates for use in the county-wide distribution of funds and other county functions. The city of Deerfield Beach compiles estimates for its own decision-making processes. Table 2 provides this information. Differences in the population estimates are due primarily to differences in estimation of the transient population. This study uses the population figures compiled by the city, primarily because of the frequency of reporting, but also because transients are included in these estimates. The city-estimated population is given in Figure 4. The method used by the city to obtain these estimates is outlined below.

TABLE 2. Population Estimates, Deerfield Beach, Florida.

Date	City Estimate	County Estimate	State Estimate
July 1974	25713	– – –	21858
April 1975	– – –	28100	– – –
July 1975	28864	– – –	25731
July 1976	30802	– – –	28607
April 1977	– – –	32800	– – –
July 1977	34596	– – –	30649
April 1978	35628	36790	– – –
July 1978	36523	– – –	35079
November 1978	38076	– – –	– – –
April 1979	40198	40900	40123
September 1979	41431	– – –	– – –
January 1980	42096	– – –	– – –
April 1980	42969	– – –	– – –
July 1980	43674	– – –	– – –

Source: Unpublished material from the Community Development Department of Deerfield Beach, Florida.

In March 1978 a minicensus was conducted by the Department of Community Development of Deerfield Beach to determine: 1) the number of persons per housing unit; 2) age and sex of occupants; and 3) seasonal occupancy. The census was conducted by first sending the condominiums and large apartment houses questionnaires. Many condominium owners' associations and apartment managers cooperated willingly and returned the questionnaires promptly. The community development staff then divided the city into homogeneously zoned neighborhoods and conducted a field survey. Specific blocks, representative of the neighborhood, were selected for personal interviews. The results of the survey provided an estimate of the March 1978 population. Quarterly updates of the population are based on the Building Department's Certificates of Occupancy (CO's). Although the minicensus cannot compare with the federal census, it does account for the tourists, who are included only at their place of permanent residency in the federal census. Certificates of Occupancy do not guarantee that the dwelling is occupied, but housing speculation was minimal in this time period due to the poor housing market and few builders were carrying large inventories of unsold houses.

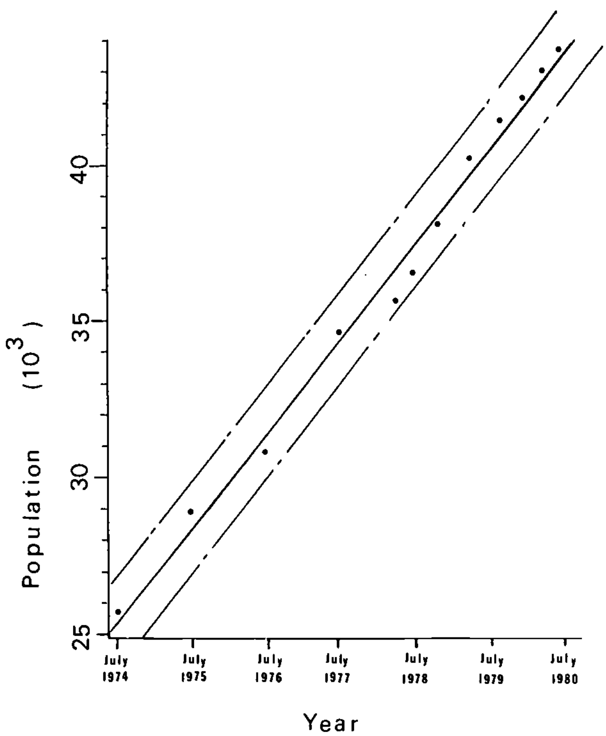

Figure 4. City-Estimated Population and 95 Percent Confidence Limits for Deerfield Beach, Florida, 1974-1980.

A linear population growth rate of 252 persons per month is determined for the data in Figure 4. The regression coefficient, R^2, is 0.99 for the model

$$\hat{P}(t) = \underset{(383)}{25{,}315} + \underset{(7.6)}{252}\,t \qquad (14)$$

where $\hat{P}$ is the population and t is the number of months since July 1974. The standard errors are placed in parenthesis below the estimated parameters. A linear trend in monthly water use is found to be adequate with an R^2 of 0.32 (so $\Phi_2 = 0$ in Equation 2).

$$\hat{W}_a(t) = \underset{(36{,}300)}{34{,}604} + \underset{(0.98)}{5.06}\,\hat{P}(t) \qquad (15)$$

The units of water use are 1000 gallons per month. The slope of the line, 5060 gallons per capita per month, indicates a daily per capita use of 169 gallons. This value is similar to the 167 gallons per capita per day reported as a national average in 1970 by the American Water Works Association (1978). The value of the intercept, 34,604,000 gallons per month, is about 15 percent of total use and may be regarded as a fixed water use related to the irrigation and maintenance of public lands and other uses not dependent on per capita consumption.

Seasonal Variation

Monthly average values of detrended water use, $\hat{W}_b(m)$ (Equation 4), may be found by arithmetic averaging or by fitting Fourier series to the observed data. For a smoothly varying pattern, the Fourier series approach leads to fewer variables to be determined. The Fourier series coefficients are calculated by means of the Discrete Fourier Transform (see Franklin and Maidment, 1983). The mean monthly rainfall could be satisfactorily modeled by two significant cycles of 6 and 12 months:

$$\hat{R}_b(m) = -0.92 \cos \frac{2\pi}{12} m - 1.23 \sin \frac{2\pi}{12} m + 0.22 \cos \frac{2\pi}{6} m - 1.19 \sin \frac{2\pi}{6} m + 4.60 \quad (16)$$

where $m = 1, 2, \ldots, 12$ and $m = 1$ is January.

The annual average rainfall $(\overline{R_b})$ is 4.60 inches per month. Table 3 records the monthly averages obtained by evaluating Equation (16). The annual total, 55.15 inches, agrees closely with the 55 inches per year reported by NOAA (1974) for this area of Florida. Figure 5 reproduces the deseasonalized monthly rainfall, $R_c(t)$, which is the recorded rainfall less its monthly fitted average (Equation 9).

TABLE 3. Fourier Fitted Monthly Average Rainfall, Deerfield Beach, Florida, 1976-1980.

Month	Rainfall (in)
January	2.26
February	1.93
March	3.15
April	4.91
May	5.92
June	5.74
July	5.09
August	4.98
September	5.61
October	6.12
November	5.55
December	3.89
TOTAL	55.15

The detrended water production also has two significant periods, a 12-month and a 4-month cycle, modeled as

$$\hat{W}_b(m) = -14{,}200 \cos \frac{2\pi}{12} m + 5760 \sin \frac{2\pi}{12} m + 8270 \cos \frac{2\pi}{4} m - 16{,}500 \sin \frac{2\pi}{4} m \quad (17)$$

The average value of $\overline{W_b}(m)$ is zero because the expected value of the residuals of the trend regression (Equation 2) is zero. Table 4 records the monthly detrended water use, $W_b(t)$, and Figure 6 displays the deseasonalized water use data, $W_c(t)$.

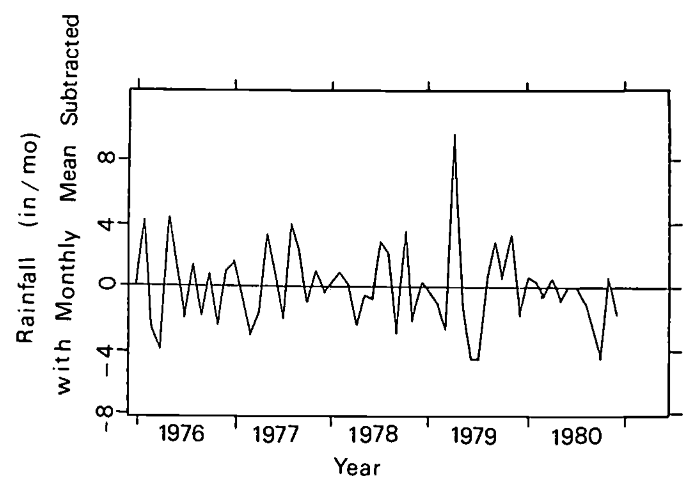

Figure 5. Deseasonalized Monthly Rainfall, Deerfield Beach, Florida, 1976-1980.

TABLE 4. Fourier Fitted Average Detrended Monthly Water Use, Deerfield Beach, Florida, 1976-1980.

Month	Use (1000 gal)
January	−25871
February	−10369
March	22225
April	20351
May	−1289
June	5927
July	25871
August	10369
September	−22225
October	−20351
November	1289
December	−5927
TOTAL	0 (due to detrending)

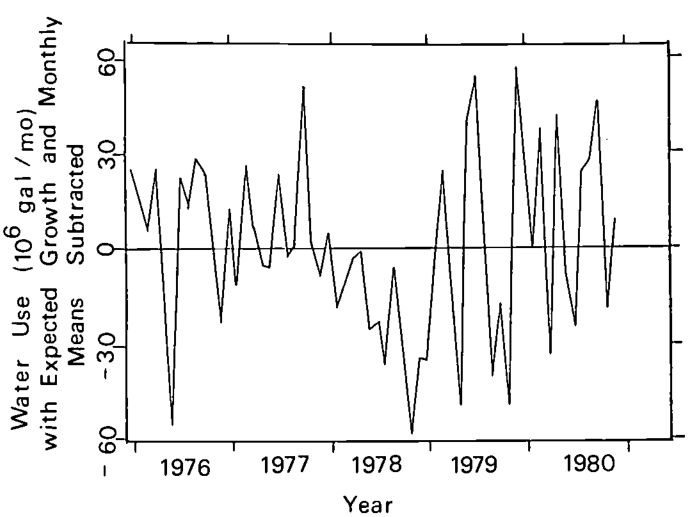

Figure 6. Deseasonalized Monthly Water Use, Deerfield Beach, Florida, 1976-1980.

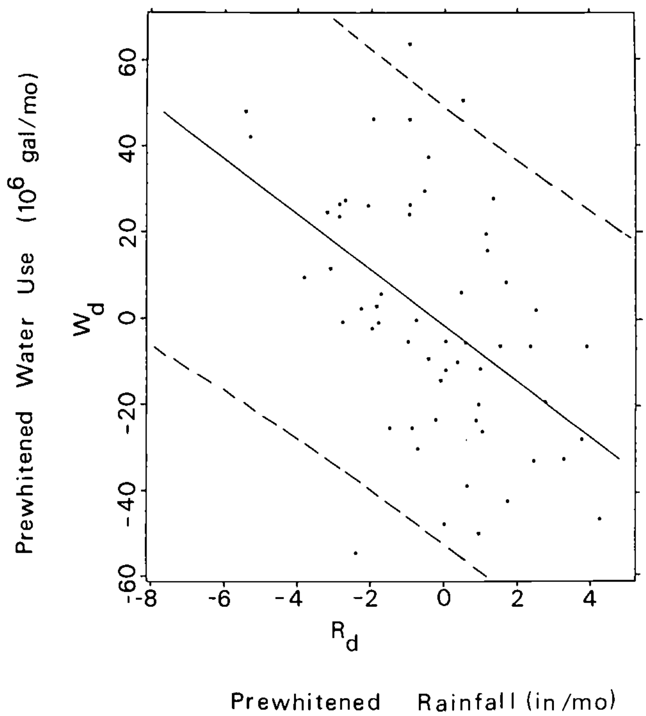

Figure 7. Linear Relationship and 95 Percent Confidence Limits for Prewhitened Water Use and Rainfall, Deerfield Beach, Florida, 1976-1980.

Autoregression

The best fit autoregressive rainfall model as determined by the Criterion Autoregressive Transfer function, CAT (Parzen, 1977), is of order two (J=2 in Equation 10). CAT is a method of determining the optimal number of coefficients to include in an autoregressive model by balancing the decrease in the model error as the model size increases with the increase in the error of estimating the number of coefficients (Parzen, 1977). The model developed is:

$$\hat{R}_c(t) = -0.16\, R_c(t-1) - 0.28\, R_c(t-2) \tag{18}$$

A monthly water use model of order one (K=1 in Equation 6) is suggested by CAT, the model being:

$$\hat{W}_c(t) = 0.14\, W_c(t-1) \tag{19}$$

Rainfall Relationship

The prewhitened variables, ($W_d(t)$ from Equation 7 using $W_c(t)$ from Equation 19 and $R_d(t)$ from Equation 11 using $R_c(t)$ from Equation 18), are compared in Figure 7. It should be noted that the origin in this figure corresponds to the average values of the original variables, adjusted for seasonal effects and autoregression. There is a broad indication of increasing rainfall causing decreasing water use and vice versa. The scatter of the points is too great to easily identify a model, however. Water use was also investigated against lagged rainfall on the presumption that departures from normal water use may be more highly dependent on departures from normal rainfall of a previous month. The lag one and lag two rainfall variables were investigated and were even less regular and displayed even greater scatter about any line.

Although the residuals from the model do not exactly correspond to the levels of the variables themselves, one could still reasonably expect the rainfall-water use relationship to have a limited range of applicability because there may be a minimum water use necessary for indoor residential use and a maximum water use reflecting total fulfillment of all indoor and outdoor requirements irrespective of rainfall. Regardless of the theoretical relationship which may exist between the variables, a linear relationship is considered the most reasonable which could be ascertained from the data available. The model, with an R^2 of 0.19, is

$$\hat{W}_d(t) = -6500\, R_d(t) \tag{20}$$

where $\hat{W}_d$ is the residual series of water use whose units are 1000 gallons per month, while R_d is the residual series of rainfall, in inches. The intercept is zero due to the previous transformations.

The residuals, $W_e(t)$ by Equation (13), from the model were analyzed and met the established criteria for random errors, i.e., zero mean and normal distribution. They have a standard deviation of 24,500 gallons per month, which is 39 percent of the variance of the original data. Thus, the entire cascade model accounts for 61 percent of the variance of the original monthly data. Compared to other cities studied by this or similar approaches (Maidment and Parzen, 1984a; Salas-LaCruz and Yevjevich, 1972), this proportion (61 percent) is low and suggests that Deerfield Beach water use has a significantly larger random component than many other cities. A detailed examination of daily water use in nine cities (Maidment, *et al.,* 1985; Maidment and Miaou, 1986) shows that the proportion of variance in water use that can be modeled is lower in Florida than in Texas because Florida has more regular rainfall and the impacts of individual weather events are less pronounced in Florida than in Texas.

METHODOLOGY FOR FORECASTING WATER USE

Example of 1981 Forecasts for Deerfield Beach Using Actual Rainfall

The parameters of Equations (14) to (20) were determined using monthly data from Deerfield Beach for January 1976 through December 1980. As a check on the validity of the model developed, it is used for providing forecasts for January through December 1981, a period that includes a sharp, severe drought and a 10 percent mandatory reduction in the city water supply. It should be noted that these forecasts are one-step-ahead forecasts and, therefore, utilized actual (realized) data for all but the current time step.

To simplify the example (and to provide some important results discussed in the second section), actual rainfall is used in the forecasts. The rainfall data for 1981 are prewhitened by being passed forward through the cascade of filters to subtract the seasonal and autoregressive components which were evident from the analysis of the 1976-80 data. The rainfall-water use relationship is then utilized to forecast the value of prewhitened water use. This value is passed backwards through the autoregressive, seasonal and growth filters determined from the 1976-80 water use data to produce forecasts which can be directly compared to the actual water use in 1981. The forecast procedure is as follows with the columns referred to being those of Table 5.

Step 1. Find the effect of current rainfall on water use.

(1) Take the actual (or estimated) monthly rainfall in the current month (Column 1 of Table 5) and subtract the Fourier fitted monthly mean rainfall (Column 2) from Equation (16) or Table 3 to produce deseasonalized monthly rainfall, $R_c(t)$ (Column 3) in accordance with Equation (9).

(2) Take the deseasonalized rainfalls for the two previous months, $R_c(t-1)$ and $R_c(t-2)$, and substitute them into Equation (18); determine the residual rainfall in the current month, $R_d(t)$ (Column 4) by Equation 11).

(3) Use the rainfall-water use relationship (Equation 20) to find the weather-related water use in the current month, $\hat{W}_d(t)$ (Column 11).

Step 2. Find the actual deseasonalized water use for the previous month.

(1) Substitute the population for the previous month estimated from Equation (14) (Column 5) into Equation (15) to calculate the corresponding mean monthly water use arising from long-term trend, $\hat{W}_a(t-1)$ (Column 8).

(2) Subtract $\hat{W}_a(t-1)$ from the observed monthly water use, $W(t-1)$ (Column 13) to produce detrended water use $W_b(t-1)$ (Column 6) in accordance with Equation (3).

(3) Subtract the Fourier fitted monthly mean water use (Column 9) (calculated from Equation 17 or Table 4) from $W_b(t-1)$ to produce actual deseasonalized water use $W_c(t-1)$ (Column 7) in accordance with Equation (5).

TABLE 5. 1981 Water Use Predictions for Deerfield Beach, Florida.

	STEP 1					STEP 2						STEP 3			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	$R_a(t)$	$\hat{R}_b(m)$	$R_c(t)$	$R_d(t)$	Pop	$W_b(t)$	$W_c(t)$	$\hat{W}_a(t)$	$\hat{W}_b(m)$	$\hat{W}_c(t)$	$\hat{W}_d(t)$	$\hat{W}(t)$	$W(t)$	Error	Error
Month/Year	(in)	(in)	(in)	(in)		(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(10^6 gal)	(percent)
October 1980	1.52	6.12	−4.60		44215	28.8	49.1	258	−20.3				287		
November	6.19	5.55	0.64		44467	−14.7	−15.9	260	1.3				245		
December	1.96	3.89	−1.93	−3.12	44719	6.2	12.1	261	−5.9				267		
January 1981	0.58	2.26	−1.68	−1.81	44971	22.7	48.6	262	−25.9	1.7	11.8	249	285	−35.4	−12.6
February	2.93	1.93	1.00	0.19	45223	−27.7	−17.4	263	−10.4	6.8	−1.2	258	236	22.2	9.3
March	0.98	3.15	−2.17	−2.48	45475	56.0	33.8	265	22.2	−2.4	16.1	301	321	−20.1	−6.2
April	0.34	4.91	−4.57	−4.64	45727	77.9	57.6	266	20.4	4.7	30.2	321	344	−22.7	−6.6
May	4.97	5.92	−0.95	−2.29	45979	−16.8	−15.5	267	−1.3	8.1	14.9	289	251	37.7	15.1
June	4.42	5.74	−1.32	−2.75	46231	−43.9	−49.9	269	5.9	−2.2	17.9	259	225	34.2	15.1
July	4.67	5.09	−0.42	−0.89	46483	18.4	−7.5	270	25.9	−7.0	5.8	297	288	9.1	3.1
August	11.20	4.98	6.22	5.78	46735	−48.1	−58.5	271	10.4	−1.1	−37.6	243	223	19.8	8.9
September	8.37	5.61	2.76	3.64	46987	−61.4	−39.1	272	−22.2	−8.2	−23.7	217	211	6.0	2.8
October	5.59	6.12	−0.53	1.65	47239	−35.0	−14.7	274	−20.4	−5.5	−10.7	237	239	−2.0	−0.8
November	3.21	5.55	−2.34	−1.65	47491	−21.6	−22.9	275	1.3	−2.1	10.7	285	253	31.9	12.6
December	0.14	3.89	−3.75	−4.27	47743	12.5	18.4	276	−5.9	−3.2	27.7	295	289	5.6	3.1

Step 3. Compute the forecast of water use for the current month.

(1) Substitute $W_c(t-1)$ from Step 2 into Equation (19) to produce the estimated deseasonalized water use in the current month, $\widehat{W}_c(t)$ (Column 10).

(2) Add the long term trend component, $\widehat{W}_a(t)$ (Column 8), the seasonal component, $\widehat{W}_b(m)$ (Column 9), the autoregressive component, $\widehat{W}_c(t)$ (Column 10) and the rainfall-water use component, $\widehat{W}_d(t)$ (Column 11), to produce the one step ahead forecast of water use, $\widehat{W}(t)$ (Column 12) in accordance with Equation (1).

At this point, the forecasts have been generated and for most purposes the task is completed. However, this study was to compare the actual and forecasted values for a one-step-ahead procedure. The remainder of Table 5 completes this task. The error of each forecast is $E(t) = W(t) - \widehat{W}(t)$ (Column 14) and the relative error is $E_r(t) = E(t)/W(t)$ (expressed in Column 15 as a percentage). The measures or the forecast error for the whole of the forecast period (the 12 months of 1981) may be computed: the standard error is

$$S_e = \left[\frac{\sum_{t=1}^{12} E^2(t)}{11} \right]^{1/2} \tag{21}$$

and the average absolute relative error (AARE) is

$$AARE = \frac{1}{12} \sum_{t=1}^{12} \left| E_r(t) \right| \tag{22}$$

where the vertical bars | | indicate the absolute value. The residuals of the comparison of the actual and forecasted values for 1981 are given in Figure 8. Forecasts ranged from overpredicting by 15 percent to underpredicting by 13 percent, the average absolute relative error (AARE) is 8 percent. Approximate 95 percent confidence limits on the forecast are estimated as plus or minus twice the standard deviation of the residual series $W_e(t)$ from Equation (13). This "model error" is 24.5 x 10^6 gal/mo. No observed values fall outside these confidence limits during 1981, although the largest errors are produced following the mandatory reduction in water supply (May and June) and following a period of excessive rainfall (August).

Comments on the Forecasting Procedure

The preceding example used actual rainfall amounts during the forecast period. Although nobody can forecast rainfall perfectly, the exercise is necessary as it provides an indication of the inherent accuracy of the model uncomplicated by inaccuracies in rainfall predictions. This measure is important to water managers as they face the problem of determining the likelihood of demand exceeding supply in the near future in that the manager can simulate the response of water use to several years of known rainfall conditions and obtain a measure of the likely range and variability of water use in forthcoming months or years.

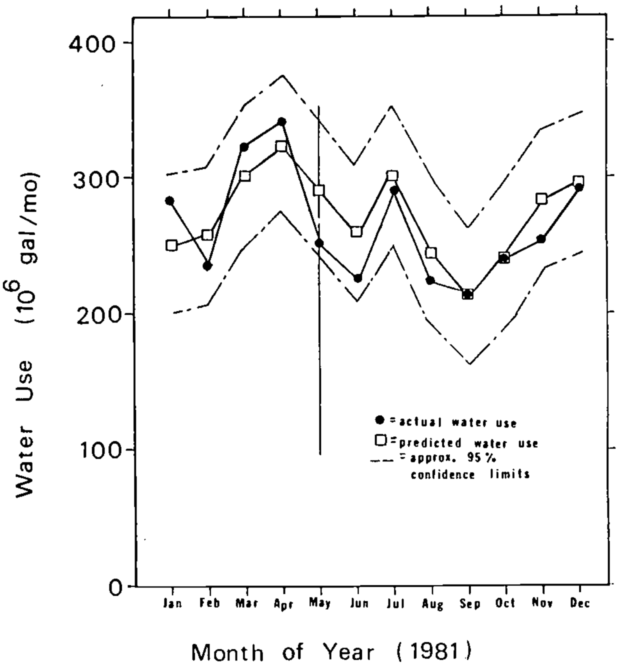

Figure 8. Predicted and Actual Monthly Water Use with Approximate 95 Percent Confidence Limits for 1981 for Deerfield Beach, Florida.

The necessary rainfall predictions may fall into two categories: estimates of the absolute quantity of rain to fall, or, the relative shortage or excess below or above the average rainfall. The first type of rainfall estimate follows the example problem exactly and is most useful when a rainfall forecast, such as from the National Weather Service, is provided or if future water use is to be simulated from past weather conditions. The second type of estimate is most useful in determining responses to deficiencies in average rainfall. For example, the response to a deficiency of two inches below average monthly rainfall may be determined. In this case, forecasts of water use may begin directly with Column (3) of Table 5.

This methodology may also be used for other forecast horizons such as one day ahead or one week ahead, or a period of several days, weeks, or months ahead. Therefore, estimates of utility demand in August may be made in January (with a seven-month-ahead forecast) as long as some estimate of monthly rainfall for that period, for example, from historical data, is provided. In this case, Table 5 would terminate at Column (12) and the autoregressive behavior of $\widehat{W}_c$ would be based on the previous estimate of $\widehat{W}_c$ from Column (10) for all but the first month's estimate (where an actual value of W_c may be used by following the procedure outlined above).

Obviously, the further into the future forecasts are desired, the less accurate they may be.

Comparison of Various Forecasting Scenarios

The example problem indicates that, given the actual monthly rainfall, fairly accurate forecasts of monthly water use can be produced. Since rainfall prediction is an uncertain science, three scenarios regarding future weather conditions are developed to indicate how well the model might perform under conditions when future rainfall is not known in advance. One forecast (the "mean rainfall" forecast) is made assuming a return to normal weather by predicting the coming month's rainfall as the historical average for that month. A second forecast (the "residual rainfall" forecast) is made assuming the past month's relative weather conditions continue into the coming month by predicting equal rainfall excesses or deficiencies over normal (i.e., assuming $R_c(t) = R_c(t-1)$. A third forecast is based on historical mean monthly water use without a relationship to rainfall. The weekly analysis was performed and weekly forecasts also generated in the manner described above. The relative magnitudes of the components of the cascade monthly forecast can readily be seen from Columns (8) through (11) of Table 5. The long-term growth accounts for the majority of the variance in monthly water use, confirming its importance in long-term forecasting.

Table 6 presents a comparison of seven forecasting methods. Method 1 is a simple predictor using the mean value of water use from 1976-1980 in each period as the forecast value for 1981; Methods 2 to 4 are successively more complete time series models of water use without reference to rainfall; Methods 5 to 7 represent various assumptions regarding the knowledge of future rainfall.

For both monthly and weekly data, the forecasts of Method 1, which are based on historical mean water use values, consistently underpredict the observations for 1981 even though rainfall ranged from 4.6 inches less than average 5.8 inches greater than average during 1981. The time series Method 4 ($\hat{W}_a(t) + \hat{W}_b(t) + \hat{W}_c(t)$) gives significantly better forecasts than Method 1 but is not significantly more accurate than Method 3 ($\hat{W}_a(t) + \hat{W}_b(t)$) for the monthly case. For the weekly time step, the inclusion of the autocorrelation term ($\hat{W}_c(t)$) in the model does improve model performance. A further substantial reduction in forecast error results when future rainfall is known (Method 7) but forecasts based on estimated rainfall are little better than ignoring rainfall altogether for both monthly and weekly data.

Comparing the monthly and weekly forecasts it can be seen that the inclusion of the autocorrelation term considerably improves the forecast accuracy of the weekly data but not for the monthly data. The lower limit on forecast accuracy is about the same (8 percent) for both weekly and monthly data.

SUMMARY AND CONCLUSIONS

A Cascade Forecasting model is presented as a tool for producing and evaluating one-step-ahead forecasts of municipal water use based on the past history of water use, rainfall and population. The model divides water use into four components: long term (deterministic) components – long-term growth and seasonal cycle, and short-term (stochastic) components – autocorrelation and correlation with precipitation. The deterministic components are related to slowly changing parameters, such as the change in per capita consumption, and therefore can project the basic water use pattern fairly accurately for a period of a few months ahead. The stochastic components are directly affected by rapidly fluctuating variables, such as daily weather and previous water use, and therefore can adjust the deterministic pattern to account for unusual circumstances.

Analysis of the monthly water use for Deerfield Beach, Florida, from January 1976 through December 1980 indicates a linear long-term growth rate associated with population growth, and an annual seasonal cycle of water use. Also present is a weak influence on current water use by water use in the previous month and a relationship with rainfall. The rainfall-water use relationship is tentatively identified as linear, although there is a great deal of scatter in the points. However, the contribution to forecast accuracy of even this tentative relationship is significant. Analysis of weekly data indicate similar relationships, although the more variable

TABLE 6. Comparison of Accuracy of Various Forecasting Components, Deerfield Beach, Florida, 1981.

Method	Series Used for Forecasting	AARE (percent) Monthly Model	AARE (percent) Weekly Model
1	Historical Mean	14.6	15.8
2	Long-Term Trend	14.2	16.7
3	Trend Plus Seasonality	12.1	15.5
4	Trend Plus Seasonality Plus Autocorrelation	12.0	11.9
5	Total Model "Residual Rainfall" Forecast	13.6	12.8
6	Total Model "Mean Rainfall" Forecast	11.7	11.9
7	Total Model Actual Rainfall	8.0	7.7

nature of the weekly data produced weaker deterministic components and stronger stochastic components.

The cascade model is used to produce one-step-ahead forecasts for 1981, a period of abnormal weather. In spite of the fact that the abnormal weather patterns were not present during the 1976-80 period used to specify the parameters, the forecasts compared favorably to the actual water use during 1981. The increased forecast accuracy due to the addition of each model component is determined. Compared to forecasting the historical mean water use, forecast accuracy is increased by accounting for the deterministic components for monthly water use, but no similar increase is realized for weekly water use. Weekly water use exhibits much stronger autocorrelation than monthly use, so the inclusion of autocorrelation in the model increases forecast accuracy noticeably. The inclusion of estimated average rainfall does not substantially increase the accuracy of either the monthly or weekly model. However, a relationship between water use and rainfall exists, as evidenced by an increase in forecast accuracy when actual rainfall is used in the models, and can be utilized to determine the expected variability of future water use by running multiple sequences of historical rainfall through the modeling process. The robustness of the model is evident in that the forecasts 'track' the water use even during the drought period.

The conclusions of the study are:

(1) In spite of the large random error, forecasts based on the time series model performed better than assuming historical mean water usc, evcn though the forecast period experienced a period of abnormal weather not present in the period used to specify the model parameters. The observed values fall within the 95 pcrccnt confidencc limits of the forecasts. The largest errors occurred during the drought and following a period of excessive rainfall. This may be the result of the lincar climatic model's inability to reproduce the physical reality of the effect of extreme events on water use.

(2) Forccasts of one-month-ahead water use for Deerfield Beach, Florida, based on a full-time series model, without the rainfall component, were only marginally more accurate than forccasts based solely on the long-term components of the model. One-week-ahead forecasts were improved by the addition of the autocorrelation components.

(3) The addition of a rainfall component improves model performance. The addition of this component allows the model to be used for risk analysis by determining a range of futurc possible use based on the results of multiple historical rainfall sequences.

ACKNOWLEDGMENTS

This research was partially supported by the South Florida Water Management District and the Water Resources Center of the University of Florida. The assistance of James Heaney, Dale Holinbeck and Carl Woelche is gratefully acknowledged.

LITERATURE CITED

Franklin, S. L. and D. R. Maidment, 1983. Forecasting Municipal Water Use During a Drought: A Case Study of Deerfield Beach, Florida. Tech. Report CRWR-188, Center for Research in Water Resources, Austin, Texas, 103 pp.

Maidment, D. R. and E. Parzen, 1981. A Cascade Model of Monthly Municipal Water Use. Texas Engineering Experiment Station, College Station, Texas.

Maidment, D. R. and E. Parzen, 1984a. Time Patterns of Monthly Water Use in Six Texas Cities. Journal Water Resources Planning and Management, Am. Soc. Civil Engineering 110(1):90-106.

Maidment, D. R. and E. Parzen, 1984b. A Cascade Model of Monthly Municipal Water. Water Resources Research 20(1):15-23.

Maidment, D. R. and S.-P. Miaou, 1986. Daily Water Use in Nine Cities. Water Resources Research (in press).

Maidment, D. R., *et al.*, 1985. Analysis of Daily Water Use in Nine Cities. Tech. Report CRWR-201, Center for Research in Water Resources, Austin, Texas, 67 pp.

National Oceanographic and Atmospheric Administration (NOAA), 1974. Climates of the States Volume 1. Watcr Ccnter Information, Inc., Port Washington, New York.

Oh, H. S. and H. Yamauchi, 1974. An Economic Analysis of the Patterns and Trends of Water Consumption Within the Service Area of the Honolulu Board of Water Supply. Report 84, Hawaii Water Resources Research Center, University of Honolulu, Honolulu, 97 pp.

Parzen, E., 1977. Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes. *In:* Multivariate Analysis, IV, P. Krishnaiah (Editor). Symposium on Applications of Statistics, North-Holland, Amsterdam, pp. 283-295.

Salas-LaCruz, J. D. and V. Yevjevich, 1972. Stochastic Structure of Water Use Time Series. Hydrology Paper 52, Colorado State University, Fort Collins, Colorado, 71 pp.

Seidel, H. F., 1978. A Statistical Anahysis of Water Utility Operating Data for 1965 and 1970. Journal American Water Works Association 70(6):315-323.