PREDICTION OF WATER USE IN PUERTO RICO

Phase I: Mayagüez

by

Nazario D. Ramirez-Beltran
Industrial Engineering Department
University of Puerto Rico
Mayagüez, P.R. 00681

Project No.
Grant Agreement No. 14-08-0001-G-2043

Final Technical Report
Submitted to the Office of Water Resources Research
U.S. Department of the Interior
Washington, D.C. 20240

The activities on which this report is based were financed in part by Department of the Interior, through the Puerto Rico Water Resources Research Institute.

May, 1993.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to the United Estates Department of the Interior. Their financial support is greatfull appreciate. He specially wishes to express his gratitude to the Water Resources Research Institute of the University of Puerto Rico, institution that also financially supported this project. The author wishes to appreciate Nestor M. Torres, associate director of financial and consumer services of the Puerto Rico Aqueduct and Sewer Authority, who provided the water use series for the city of Mayaguez Puerto Rico. The author also appreciates the administrative efforts and support provided by Dr. Rafael Muñoz, Director of the Water Resources Research Institute.

Mike F. Acuña, a graduate student at the Department of Industrial Engineering was dedicated to conduct the computational tasks. His cooperation and professional efficiency are highly appreciated by the author of this report. Gloribel Perez Perez and Ileana Rivera Pagán undergraduate students of Industrial Engineering Department efficiently helped in the data collection process and the computational tasks.

This research was conducted by Dr. Nazario D. Ramirez-Beltran who is working as an associate professor of the Industrial Engineering Department of University of Puerto Rico. Computer facilities of the University Computer Center and the Industrial Engineering Computer Center were intensively used during the realization of this project. The computer packages: SAS/ETS, Autobox, MTS, Quatro Pro, Lotus 123, and Statgraphics were used to derive the present results.

ABSTRACT

A sampling procedure was used to collect socioeconomic data which is correlated with residential water use. One hundred and thirty houses were visited twice to estimate the residential water use. Eight years of monthly water use of three variables were analyzed. The variables studied were: industrial, residential, and commercial water use. Number of workers and climatological series were used as input variables to explain the three underlying water use variables.

Regression and times series analysis are discussed in general form. Three regression identification techniques are described: stepwise regression, directed search on t, and all possible regressions. Multicollinearity problem, detection outliers and residual analysis are also discussed. Five time series identification techniques are detailed presented: dynamic regression, univariate ARIMA, transfer function, vector ARMA, and state space techniques.

Linear and nonlinear econometric models were developed to predict the residential water use. Five different time series models were identified to each time series. Detailed fitting techniques are presented in this final report. Univariate and multivariate time series technique were used to represent the stochastic behavior of the monthly water use for the city of Mayaguez Puerto Rico. The identified stochastic difference equations are: dynamic regression model, autoregressive moving average model, transfer function model, vector autoregressive moving average model, and the state space model. Fitting and prediction capabilities were studied to identify the best model for expressing the industrial, residential, and commercial time series. The best models were: the dynamic regression model for the residential water use, the transfer function model for the commercial water use, and the transfer function model for the industrial

water use. Multicollinearity problem and outliers detection was investigated on the selected models. Residual analysis to check model adequacy was also conducted on the selected models. Prediction for twelve months ahead were computed with the selected models.

TABLE OF CONTENTS

use. The econometric model uses some economical variables which reveals evidences of affecting the water consumption. This model is independent of time and may be used to predict the annual water use for a household.

2. OVERVIEW OF PREVIOUS AND RELATED WORK

The rapid growth of cities cause sever strains on their municipal water supply systems. To develop a rational urban growth plan it is necessary to estimate the water needs for either new urbanizations or new industrial parks. Kuttan (1970) pointed out that the enormous development in the industry and the population of Puerto Rico have demanded careful planning of the island's water requirements and resources. Guilbe (1972) pointed out that Puerto Rico's booming population has caused competition for water to become more intense and thus the corresponding water problems have increased in number of complexity. Lobb (1975) indicated that extra capacity costs are in the range of 30-40% of the total annual cost of a typical water utility. Therefore, timely and accurate prediction of water use is required for deriving an efficient urban plan for the rational growth of the major cities in Puerto Rico.

Some statistical models have already been used to predict the water use in Puerto Rico (Guilbe, 1972; Attanasi, et al, 1975). However, those models cannot be used with the actual data and conduct the required prediction. Some of the models are poorly identified so that prediction may not be adequate. For instance the collected data by Guilbe (1972) were processed using the his model and R^2 was equal to 0.21. Using the same data it is possible to improve the coefficient of determination, R^2=0.27. A linear model that improves the regression fit to the Guilbe data is the following:

$$y = a + bx^{5/2} + cx^3 \tag{2.1}$$

It should be made clear that further improvement can be easily obtained; however, this is not the main concern of this research. It should be noted that the Guilbe's model is independent of time and is designed to conduct long term prediction. If monthly water use schedule is needed, then the Guilbe model is unable to provide forecasts for that time period, i.e., a time dependent function is needed to obtain such forecasts. US Geological Survey (1975) with the Environmental Quality Board of the Commonwealth of Puerto Rico developed a system analysis for water resource planning (Attanasi, et al., 1975). Time series model were used to develop models to study residential, commercial and industrial water use.

Variations of urban water use over time are caused in part by socioeconomic factors such as city population, household income, and water price, and in part by climatic factors such as rainfall, evapotranspiration, and temperature. However, water consumption responds to these factors on different time scale. The population growth in a city, for instance, causes a gradual increase in water use, which may be apparent over a span of several years, while the effects of climatic factors are evident in a short horizon (Maidment, and Parzen, 1984).

Most of the demand modeling techniques are base upon multiple regression analysis. In most of the cases monthly and annual aggregated flow from treatment plants data have been used. Most models are used for policy assessment or long-term prediction of water requirements. Many of the explanatory variables in water demand models do not contribute to short-term predictive ability. To forecast residential water demand Whitford (1972) uses the following explanatory variables: housing type, household income, population density, regulation on water use, and pricing policy. Meyer and Mangan (1969) developed a sophisticated model, which uses a set of multiple regression equations. Thompson (1976) pointed out that Meyer-Mangan's model has shown a large forecast error. Obviously, regression is not the best water demand forecasting technique, specially when the involved variables are auto and cross-correlated. It is well known that least square approach fails to identify the appropriate model, since underestimate the parameter variance. Thus, the t-statistic of the estimated parameter may indicated that the underlying parameter is significantly different from zero, when in fact it is not (Granger and Newbold, 1974; Domokos, Weber, and Duckstein, 1976).

Other authors have used time-series regression approaches with lagged demand in the model to capture the time dependent behavior of the process. Agthe and Billings (1980) employed lagged past demand and climatic variables to develop a dynamic model. Hansen and Narayanan (1981) developed a multivariate regression model to study monthly variations in municipal water demand. Water demand was the dependent variable and the explanatory variables were: lagged monthly water use, price, average temperature, total precipitation, and percentage of daylight hours. He claims that his model is useful to study the impact of management decisions such as price changes, and conservation programs.

Maidment, et al., (1985) developed a time series model to forecast the maximum daily demand rate during summer water conservation period. The water use data of different metropolitan cities seems to exhibits strong influences of temperature and rainfall on the demand for treated water. They pointed out that climatic variables most influence daily water use, especially during summer months. Maidment and Miaou (1986) developed a time series model with intervention to represent the daily water use.

Smith (1988) presents a time series model of daily municipal water use. He claim that his model is a conditional autoregressive process with randomly varying mean. The varying accounts for changes in water use that result from interaction over time of the price of water, plumbing code provisions, and customer income.

Weber (1989) uses a regression model to forecast water demand, the input data are: monthly water consumption, evapotranspiration, temperature, rainfall, water price, household income, and consumption patterns. Excellent literature reviews of water use modeling and forecasting approach are given by Maidment and Parzen (1984), Maidment, et al., (1985), and Sastri (1987).

3. DATA COLLECTION

Data involved in this research include water use, climatic variables, and socioeconomic variables. Data can be classified into econometric data, monthly water use, and climatic factors. These data were obtained from different sources: a direct survey, government offices, and journals.

3.1. Sampling data.

A direct survey was designed to collect the required data for the econometric model. It is expected that urban water use over time is correlated with some socioeconomic factors such as population, household income, automobile washing requirements, garden areas, number of bathrooms, etc.

Population of the city of Mayaguez Puerto Rico was defined as the population for the purposes of study. A stratified random sampling was conducted. A non-overlapping group with similar characteristics was defined as a stratum. A sample size in each stratum was determine so that the following cost coefficient function was minimized.

$$C = c_0 + \sum_{i=1}^{20} c_i n_i \qquad (3.1)$$

where the value of c_0 is the general and fix expense, the coefficient c_i is the sampling cost associated with stratum i, the value n_i is the sample size in the stratum i, and C the total sampling cost. The total sample size that minimizes the sampling cost is:

$$n = \frac{(C-c_0)\sum_{i=1}^{20} \frac{N_i S_i}{\sqrt{c_i}}}{\sum_{i=1}^{20} N_i S_i \sqrt{c_i}}$$

and (3.2)

$$n_i = \frac{n N_i S_i / \sqrt{c_i}}{\sum_{j=1}^{20} N_j S_j / \sqrt{c_j}}, \quad for \quad i = 1, 2, \ldots, 20$$

where N_i is the total units in the ith stratum, and S_i is the standard deviation of water use in the ith stratum. The optimum sample size was 130 and the sample size for each stratum is exhibited in Table 3.1

Table 3.1 Sample size for stratum

Stratum	Sample size	Stratum	Sample size
Algarrobo	8	Naranjales	1
Guanajibo	12	Quebrada	8
Juan Alonso	2	Quemado	4
Leguisamo	2	Rio Cañas	4
Limon	1	Rio Cañas arriba	1
Malezas	1	Rio Hondo	3
Pueblo	46	Rosario	2
Pueblo arriba	1	Sabalos	18
Miradero	11	Sabanetas	3
Montoso	1	Batayes	1

A suitable questionnaire was designed (see appendix A) and two visits were conducted to the same house in order to estimate the daily water consumption. Collected data from questionnaire are summarized and shown in appendix B.

3.2. Water use data.

The Aqueduct and Sewer Authority (ASA) was one of the main sources of information for this research. Eight years of bimonthly water use data were provided by ASA. The variables involved in the data provided by ASA are: residential, commercial, and industrial water use. Appendix C exhibits these data and Appendix D shows the plots of these variables. The accuracy of provided data was questioned however, no additional information was provided to measure the quality of the given data. Although, many times and for different media was applied for the daily and annual water use data, this agency does not provided the asking data. This situation may occur because during the project time the ASA was highly criticized because of the water service dose not correspond to the expected public demands.

3.3. Climatological data.

Monthly temperature and rainfall for Mayaguez was obtained from Climatological data of Puerto Rico and Virgin Islands, which is available from the National Oceanic and Atmospheric Administration. Appendix C presents the collected data and Appendix D shows the plots of the underlying data.

4. MODEL BUILDING TECHNIQUES.

Two major techniques for model building are study in this research: regression, and times series analyses. Regression models are used to compute long term prediction. On the other hand, times series models are useful to predict the stochastic behavior of water use in a short time period. This is because times series model identification is base in the autocorrelation function of stationary times series. Autocorrelation function for a stationary time series approach zero for large values of lagged time. Regression techniques were used to identify an econometric model and time series technique were implemented to describe the monthly water use variables.

4.1 Regression Analysis.

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. Regression analysis is based on the method called least squares which takes the best fitting model to be the one that comes closest to the data in the sense of minimizing the sum of squared discrepancies between the observed value and the values given by the model.

Building a regression model that includes only a subset of the available regressors involves two conflicting objectives: (1) It is desirable to have as many regressor variables as possible so that the information content in these factors can influence the predicted value. (2) It is also desirable to include as few regressors as possible because the variance of the prediction values increases as the number of regressor increases. The process of finding a model that is compromise between these two objectives is called the identification procedure. If the model is linear in parameters, there exists various systematic technique to identify the regression model; however, if the model is not linear on parameters then empirical estimation must be conducted since there is not systematic method. The most popular and useful technique for doing model identification are: directed search on t, stepwise regression methods, and all possible regressions.

(1) Directed search on t. This method was proposed by Daniel and Wood (1980) and is useful when the number of candidate regressors is relatively large, for more than twenty regressor variables. This method consist on fitting a full model and ranking the regressors according to decreasing order of magnitude of the | t-statistic |, and then introducing the regressors into the model one at a time in this order and select the model which exhibits that the number of parameters is equal to the Mallows' C_p statistic (Mallows 1973).

(2) All possible regressions. This procedure requires to fit all possible regression equations involving one candidate regressor, two candidate regressors, and so on. The best regression equation is the one that has the largest value of adjusted coefficient of determination, R_a^2 (Montgomery, 1982). If the intercept term is included in all equations, then if there are K candidates regressors, there are 2^K total equations to estimate and examine. Thus, this method is recommended when few regressors are available, K<10.

(3) Stepwise regression methods. Because evaluating all possible regressions can be burdensome computationally, various methods have been developed for evaluating only small number of regression models by adding or deleting regressors one at a time. These methods are referred as stepwise-type procedure, which are classified into three broad categories: (1) forward selection, (2) backward elimination, and (3) stepwise regression which is a combination of procedures (1) and (2) (Montgomery, 1982).

Once model identification task is accomplished it is necessary to conduct model diagnostics and residual analysis. The most important diagnostics are detection of outliers and multicollinearity problems. Some times the regressors are nearly linearly related, and in such a case the inferences based on the regression model can be misleading or erroneous. When there are near linear dependencies between the regressors, the problem of multicollinearity is present. There are two basic method for conducting multicollinearity diagnostics: the variance inflation factor, and the condition number (Montgomery, 1982). When the variance inflation factor is larger than five or the condition number is larger than one hundred, serious multicollinearity problem has been detected. The problem of multicollinearity can be solved by conducting orthogonal transformation to the regressor variables. The most useful orthogonal transformation are the ridge and the principal component regressions (Draper and Smith, 1981).

Outliers are data points that are not typical of the rest of the data. Depending on their location, outliers can have moderate to sever effects on the regression model. Outliers should be carefully investigated to see if a reason for their unusual behavior can be found. Some times outliers are bad values, occurring as a result of unusual but explainable events. In this case, the outlier should be corrected or deleted from data. If the outlier is an unusual but perfectly plausible observation. Deleting this point to improve the fit of the equation can be dangerous, as it can give the user false sense of precision in estimation and prediction. Various test have been proposed for detecting outliers. The most popular tests are: standardized residuals, and Cook's tests (Montgomery, 1982). Residuals that are considerable larger in absolute value than the others, say three to four standard deviations from the mean, are potential outliers.

Residual analysis is one of the strategies to measure the model adequacy. Residual analysis check whether or not the assumption made on model building are satisfied. That is, residual analysis checks whether or not the following assumptions are satisfied:

(a) Relationship between the dependent variable and regressor variables are linear, or at least is well-approximated by a linear relationship.

(b) The error term has zero mean.

(c) The error term has constant variance.

(d) The errors are independent.

(e) The errors are normally distributed.

Thus, a model which has enough regressors to explain the dependent variable with significant parameters, small MSE, without problems of multicollinearity and outliers, and also with approval residual analysis may be considered as a the best regression model.

4.2 Times series analysis.

A time series is an ordered sequence of observations. Although the ordering is usually through time, particularly in terms of some equally spaced time intervals. The ordering may also be taken through other dimensions, such as space. The intrinsic nature of time series is that its observations are dependent or correlated, and therefore the order of the observation is important. Hence, statistical procedures that rely on independent assumption are no longer applicable, and different methods are needed. The body of statistical methodology available for analyzing time series is referred to as time series analysis.

The statistical dependence in data is expressed by the correlation or autocorrelation between successive observations. Therefore the existing methods of time series analysis are based on empirical or estimated autocorrelation or its Fourier transform spectrum. The empirical autocorrelation is a poor estimator of the theoretical autocorrelation. This fact makes the techniques of time series analysis based on such estimates difficult and cumbersome, requiring heaving reliance on ad hoc trial and error procedure.

This difficulty can be avoided by consistently employing methods of linear system analysis and statistical inference. A time series is treated as a realization of the response of a stochastic system to uncorrelated input. The mathematical model for a dynamic system on discrete time reduces a correlated time series output to the independent or uncorrelated input. Thus, the times series methodology can be summarized as finding a model that accomplishes this reduction to independent data and then using the regular statistical techniques for independent observations for estimation, and prediction.

A stochastic process in general is a sequence of random variables that can be described by its corresponding joint probability distribution, $p(w_1,...,w_t)$. To infer such a general probability structure from a single realization, $(w_1, ..., w_t)$, would be very difficult and some times impossible, unless the process is in statistical equilibrium, which is usually called stationary. A stochastic process is said to be strictly stationary if its joint probability distribution is invariant with respect to time. That is, $p(w_1, ..., w_t)=p(w_{1=k}, ..., w_{t+k})$ for any integer k. A process characterized by having a constant mean and a covariance function which is independent of time is called covariance stationary or stationary (Fuller, 1976).

A general approach for modeling univariate or multivariate time series starts by determining whether or not each time series is stationary. A empirical approach consist of plotting the given time series. If the underlying time series exhibits variance changes over time then the process is not stationary and variance stabilization is required. If the variance of a given time series is unstable then its variance can be stabilized by means of the power transformation. On the other hand, if the time series is not stable in the mean then the process is not stationary and the trend must be subtracted either by differencing (w_t-w_{t-1}) or by fitting a polynomial function and then subtracting the fitted function from the observed values.

A general transformation to stabilize the variance was introduced by Box and Cox (1964). This power transformation can be expressed as follows:

$$T(W_t) = W_t^{(\lambda)} = \frac{W_t^{\lambda} - 1}{\lambda} \qquad \textbf{(4.1)}$$

where λ is transformation parameter. Some commonly used values and their associate transformation are:

Table 4.1 Typical transformations.

Transformation parameter (λ)	Transformation $T(W_t)$
-1	$1/W_t$
-0.5	$1/\sqrt{W_t}$
0	$\ln W_t$
0.5	$\sqrt{W_t}$
1	no transformation

It should be noted that transformation parameter can be incorporated as a model parameter to be estimated from data. The maximum likelihood estimate of λ is the one that minimizes the residual sum of squares. Frequently, the power transformation no only stabilizes the variance, but also improves the approximation to normality.

Once the time series are stationary then those series are ready to start the identification procedure. The identification procedure changes depending whether univariate or mutivariate model will be identified. In this research five identification procedures will be discussed: the dynamic regression model, the univariate ARIMA model, the transfer function model, the vector ARMA model, and the state space model.

4.2.1 Dynamic regression models.

Regression with autocorrelated errors is usually called a dynamic regression model (Haugh and Box, 1977). The identification task is actually achieved in three steps. The first step consist of fitting a multivariate regression model ignoring the autocorrelation problem. Conventional regression model identification techniques are used; for instance, stepwise regression, directed search on t, and all possible regressions. At this step, one should be aware that the variance of residuals underestimate the variance of errors, i.e., the values of t-statistic are exaggerated. The second step is subtract the fitted regression model from the observed values and then fit an ARMA model to obtained residuals. The last step is to fit simultaneously the regression and the ARMA model by means of using lagged variables and a nonlinear estimation routine. A computerized routine that can handle this task is called the SYSNLIN and is available at the computer package called SAS/ETS (1985).

The general expression of dynamic regression models is given by:

$$W_t = f(X_t, Y_t, \ldots) + \epsilon_t$$

$$\text{where} \quad \epsilon_t = \frac{\theta(B)}{\Phi(B)} a_t \quad \text{and} \tag{4.1}$$

$$\theta(B) = 1-\theta_1 B - \ldots - \theta_q B^q$$
$$\Phi(B) = 1-\phi_1 B - \ldots - \phi_p B^p$$

where a_t is a sequence of independent and normally distributed random variables with zero mean and constant variance. The letter B is the lag operator $B^k x_t = x_{t-k}$, the root of the polynomials $\theta(B)$, and $\Phi(B)$ lie outside of the unit circle. The values of p, q, θ's and ϕ's are parameters to be estimated from data.

4.2.2 ARIMA models.

Univariate autoregressive integrated moving average (ARIMA) model is a stochastic linear difference equation that expresses a univariate process as a function of its past and the lagged error component.

A general ARIMA(p,d,q) model can be expressed as follows:

$$(1-\phi_1 B - \ldots - \phi_p B^p)(1-B)^d W_t = \theta_0 + (1-\theta_1 B - \ldots - \theta_q B^q) a_t \tag{4.2}$$

where W_t is the stochastic process, ϕ's and θ's are parameter to be estimated from data, a_t is a sequence of independent random variables with zero mean and constant variance, B is the back shift operator, p is number of autoregressive parameters, q is the total of moving average parameters, and d is the order of differencing.

In time series analysis, the most crucial steps are to identify and built a model based on available data. A method of identification can be described as follows:

(1) Plot the time series and determine whether a mathematical transformation is required. The most common transformations are the power transformation and differencing. Since differencing may create some negative values, power transformation is applying first before taking differences.

(2) Compute and examine the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) of the original series to further confirm the necessary degree of differencing. If the sample ACF decays very slowly and sample PACF cuts off after lag one, it indicated that differencing is needed. To remove non-stationary mean it is necessary to considered $(1\text{-}B)^d W_t$, so that, the parameter d is known at the end of this step.

(3) Compute and examine the sample ACF and PACF of the properly and differenced series to identify the orders of p and q. The values of p and q may be obtained as follows: Fit the ARMA(p,p) where p is the largest significant lag from the ACF or the PACF. Next, drop all the parameters whose estimates are not significant, and test whether the residuals behave as white noise. The described procedure is illustrated in Chapter 5.

4.2.3. Transfer function models.

The concept of a transfer function derives from the idea of variations in the independent or input variables transferring into variations in the dependent or output variable. Transfer function models are logical extension of univariate times series models which utilize only the past history of the series for modeling. Assuming that x_t and w_t are the properly transformed series so that they are both stationary. The output series w_t and the input series x_t are related through a general model

$$w_t = \nu(B)x_t + \eta_t$$

$$\text{where} \qquad \nu(B) = \sum_{-\infty}^{\infty} \nu_j B^j \tag{4.3}$$

where $\nu(B)$ is referred as the transfer function, and η_t is the noise series of the system that is independent of the input series x_t. It should be noted that the transfer function model is also known as the ARMAX model. The coefficient of the transfer function, ν_j, are known as the impulse response weights. The transfer function is said to be stable if the sequence of these impulse response weights are absolutely summable. Thus, in a stable system, a bounded input always produces a bounded output. The transfer function is causal if $\nu_j=0$, for $j<0$. Thus in a causal model, the system does not respond to input series until they have been actually applied to the system. Thus the present output is affected by the system's input only in terms of its present and pass values. In practice a causal and stable model is considered.

The purpose of the transfer function modeling are to identify and to estimate the transfer function, $\nu(B)$, and the noise model for η_t, based on the available information of the input series x_t and the output series w_t. The major difficulty is that x_t and w_t are finite and the transfer function contain an infinite of coefficients. To alleviate this difficulty, $\nu(B)$ can be expressed in a rational form:

$$v(B) = \frac{\omega_s(B)B^b}{\delta_r(B)}$$

where (4.4)

$$\omega_s(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$$

$$\delta_r(B) = 1 - \delta_1 B - \ldots - \delta_r B^r$$

and b is a delay parameter representing the actual time lag that elapses before the impulse of the input variable produces an effect on the output variable. For a stable system the roots of $\delta_r(B)$ are assumed to be outside of the unit circle.

The general procedure to identify the transfer function model can be described as follows:

(1) Prewhiten the input series,

$$\alpha_t = \frac{\Phi_x(B)}{\Theta_x(B)} x_t \tag{4.5}$$

where α_t is a white noise series with mean zero and variance σ^2_α.

(2) Calculate the filtered output series. That is, transform the output series w_t using the above prewhitening model to generate the series β_t, where

$$\beta_t = \frac{\Phi_x(B)}{\Theta_x(B)} x_t \tag{4.6}$$

(3) Calculate the cross-correlation function (CCF) between α_t and β_t, to estimate v_k

$$\hat{v}_k = \frac{\hat{\sigma}_\beta}{\hat{\sigma}_\alpha} \hat{\rho}_{\alpha\beta}(k) \tag{4.7}$$

The significance of the CCF is tested by comparing with its standard deviation $(n-k)^{-1/2}$. The values of b, r and s are identified by observing the form of the CCF.

(4) Preliminar estimation of the transfer function is obtained by nonlinear regression.

$$\hat{\nu}(B) = \frac{\hat{\omega}_s(B)}{\hat{\delta}_r(B)} B^b \tag{4.8}$$

(5) Compute noise series. Once preliminar estimates for the transfer function is obtained, noise series is computed as follows:

$$\hat{\eta}_t = w_t - \frac{\hat{\omega}_s(B)}{\hat{\delta}_r(B)} B^b x_t \tag{4.9}$$

(6) Identify the noise model. Based on the sample ACF and PACF of residuals the following structure is identified:

$$\hat{\eta}_t = \frac{\Theta(B)}{\Phi(B)} a_t \tag{4.10}$$

(7) Full model estimation. Nonlinear regression method is used to conduct estimation for the full model.

$$w_t = \frac{\hat{\omega}(B)}{\hat{\delta}(B)} x_{t-b} + \frac{\hat{\Theta}(B)}{\hat{\Phi}(B)} \hat{a}_t \tag{4.11}$$

(8) compute residuals. If residuals satisfied the assumption of independence, normality, and constant variance, and also if all the parameters are significant, then the obtained model is the appropriate transfer function model. If this is not the case, then the processes starts from step (1) over again, until a satisfactory model is developed.

4.2.4. Vector ARMA models.

Times series data in many empirical studies consist of observations from several variables. In transfer function models, a specific relationship between input and output variables was studied. However, in many fields of application the transfer function may not be appropriate model. Thus, a more general class of vector time series models it may be needed to describe relationship among several times series variables. The vector autoregressive moving average (vector ARMA) model permit the testing for lead, lag, independent, contemporaneous and feedback relationships among the variables.

The vector ARMA models were introduced by Tiao and Box (1981) and the general representation can be expressed as follows:

$$\boldsymbol{\phi}_p(B)\,\boldsymbol{\Phi}_P(B^s)\,Z_t = C+\boldsymbol{\theta}_q(B)\,\boldsymbol{\Theta}_Q(B^s)\,a_t \tag{4.12}$$

where $\mathbf{Z}_t$ is a stationary kx1 vector of k time series each having n observations, $\mathbf{C}$ is a kx1 vector of constants, and $\mathbf{a}_t$ is a sequence of white noise vectors, independently and identically distributed as multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The kxk matrices $\phi_p(B)$ and $\theta_q(B)$ are nonseasonal matrix polynomials in the back shift operator B. Finally, the kxk matrices $\boldsymbol{\Phi}_P(B^s)$ and $\boldsymbol{\Theta}_Q(B^s)$ are seasonal matrix polynomials in the back shift operator B and have the following form:

$$
\begin{aligned}
\phi_p(B) &= I-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p \\
\theta_q(B) &= I-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q \\
\Phi_P(B^s) &= I-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps} \\
&\text{and} \\
\Theta_Q(B^s) &= I-\Theta_1 B^s-\Theta_2 B^{2s}-\ldots-\Theta_Q B^{Qs}
\end{aligned}
\tag{4.13}
$$

where **I** is a kxk identity matrix, the ϕ's, Φ's, θ's, and Θ's are kxk matrices of parameters to be estimated from data, and s is the seasonal period.

To ensure the properties of stationarity and invertibility, all roots of the determinantal polynomials $|\phi_p(B)|$, $|\theta_q(B)|$, $|\Phi_P(B^s)|$, and $|\Theta_Q(B^s)|$ are required to lie outside of the unit circle.

As an example, consider the following vector ARMA(2,1) model, let $\mathbf{Y}_t=\mathbf{Z}_t-\mathbf{C}$ be the vector of deviation from **C**, and k=2.

$$
(I-\phi_1 B-\phi_2 B^2)\,y_t = (I-\theta_1 B)\,a_t \tag{4.14}
$$

which is equivalent to

$$
\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi^1_{11} & \phi^1_{12} \\ \phi^1_{21} & \phi^1_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \phi^2_{11} & \phi^2_{12} \\ \phi^2_{21} & \phi^2_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} - \begin{bmatrix} \theta^1_{11} & \theta^1_{12} \\ \theta^1_{21} & \theta^1_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix} \tag{4.15}
$$

In principle, identification of vector time series models is similar to identification of univariate time series models. Thus, for a given observed vector time series $\mathbf{Z}_1$, $\mathbf{Z}_2$, ..., $\mathbf{Z}_n$, its underlying model can be identified from the pattern of its sample correlation and partial correlation matrices, assuming that each time series becomes stationary after proper transformation is applied.

4.2.5. State space models.

The state space of a system is defined to be the minimum set of information from the present and past such that the future behavior of the system can be completely described by the knowledge of the present state and the future input. Thus, the state space representation is based on the Markovian property, which implies that given the present state, the future of the system is independent of its past. Consequently the state space of the system is called the Markovian representation of the system.

Let $\mathbf{Y}_t$ be a stationary multivariate times series of dimension k. The state space representation for $\mathbf{Y}_{t+1}$ is given by

$$\begin{aligned} Z_{t+1} &= F_t Z_t + G a_{t+1}, \\ Y_t &= H Z_t \end{aligned} \tag{4.16}$$

where $\mathbf{Z}_t$ is px1 state space vector, $\mathbf{F}$ is pxp transition matrix, $\mathbf{G}$ is pxk input matrix, $\mathbf{H}$ is kxp observation matrix. The first equation specifies how the state vector evolves through time while the second equation specifies the relation between the time series and the state vector. $\mathbf{Z}_t$ contains the set of present and past information that is correlationally relevant for forecasting. Its first k components comprise $\mathbf{Y}_t$. $\mathbf{a}_t$ is a sequence of independent, zero mean random vectors with constant covariance matrix. $\mathbf{H}$ takes the form [$\mathbf{I}$ $\mathbf{0}$] where $\mathbf{I}$ is a kxk identity matrix and $\mathbf{0}$ is a kx(p-k) zero matrix. The upper kxk submatrix of $\mathbf{G}$ is an identity matrix.

The above state space representation of $\mathbf{Y}_t$ is not unique. Multiplication of (4.16) by any nonsingular matrix will yield another state space representation that is also true. However, Akaike (1976) showed that a multivariate linear stochastic system has a canonical representation which is unique. Identification of the canonical state space model is accomplished in two steps:

(1) determination of the amount of past information to be used in the canonical correlation analysis. This is achieving by fitting successively higher order autoregressive models to the observed series and compute the Akaike information criterion (AIC) for each fitted model. The optimum lag into the past p is chosen as the order of the autoregressive model for which the AIC is minimum. (2) The second step involves the selection of the state vector via canonical correlation analysis between the set of present and past values $\{y_{1,t}, y_{2,t}, \ldots, y_{k,t}, y_{1,t-1}, \ldots, y_{k,t-1}, \ldots, y_{k,t-p}\}$, and the set of present and future values $\{y_{1,t}, y_{2,t}, \ldots, y_{k,t}, y_{1,t+1|t}, \ldots, y_{k,t+1|t}, \ldots, y_{k,t+p|t}\}$, where $y_{j,t+s|t}$ denotes the conditional expectation of $y_{j,t+s}$ at time t. The two previous set are referred as data space and predictor space, respectively. It should be noted that the predictor space does not include the conditional expectation for s>p, since for any vector ARMA(p,q) process, $y_{j,t+s|t}$ for s>p is completely determined by $y_{j,t}, y_{j,t+1|t}, \ldots, y_{j,t+p|t}$. Thus, once the optimum order p is determined, the predictor space consist of conditional expectation of $y_{j,t}, y_{j,t+1|t}, \ldots, y_{j,t+p|t}$. If the coefficient of the vector ARMA are less than full rank, then the predictor space is linearly dependent. However, the state space must be of linearly independent components of the predictor space. Thus, the canonical correlation analysis is used to identify the linearly independent components. Once the component of the state vector are selected the matrices **F** and **G** are estimated. The optimum autoregressive model provides an estimate of the input matrix **G** and the covariance matrix of the innovation. An estimate of **F** matrix is obtained during the canonical correlation analysis.

5. MAYAGUEZ WATER USE MODELS

Increasing interest has been raised in determining the urban water needs. This interest has resulted from the necessity to avoid future water shortages. An adequate potable water supply is a primary factor in the well-being and economic progress of the city of Mayaguez, Puerto Rico.

Model scope is limited to the city of Mayaguez. An extension to other parts of Puerto Rico may require to implement the sampling procedure and/or update the derived models to the current data. Approximate prediction may be obtained from the Mayaguez models, after using the appropriate input values.

This chapter presents the models fitted to the water use of the city of Mayaguez. The behavior of the three variables were study: residential, commercial, and industrial water use. Univariate and multivariate difference equations are fitted and presented in the following order: linear and nonlinear econometric models for residential water use, and monthly models for residential, industrial, and commercial water demands.

5.1. Econometric models.

It is desirable to have an econometric model to predict the residential water use for long term periods and for a specific urban growth. The econometric models are useful for purposes of urban planning. The available resources should be compared with water requirements to generate a reasonable urban growth. Linear and nonlinear models were developed with the socioeconomic data obtained by means of a sampling procedure.

5.1.1. A linear econometric model.

The identified linear model can be expressed as follows:

$$\sqrt{RW} = \beta_0 + \beta_1 (AT)^2 + \beta_2 (NP)^3 + \beta_3 (NP)^6 + \beta_4 HI + \epsilon \qquad (5.1)$$

where the variables RW, NP, AT, NB, HI represent the residential water use, the number of people, the number of automobiles, the number of bathrooms, and the income in each household, respectively, ε is a random variable and normally distributed with zero mean and constant variance. It should be noted that (5.1) is linear on parameters. Parameter estimation and test statistics are exhibit in Table 5.1. This table shows that all parameters were statistical significant at the 95% confidence level. The best linear model still reveals poor fitting since there exists a weak correlation between the explanatory variables and the dependent variable. Although the identified model shows a poor fitting this is an improvement over the model developed by Guilbe whose R^2 was 0.21.

Table 5.1. Parameter estimation for the linear model.

VARIABLE	Estimate	T-ratio
β_0	0.56093	20.14
β_1	0.01117	2.23
β_2	0.0015	4.14
β_3	-1.98E-06	-2.75
β_4	6.42E-06	5.00
R^2 = 0.437	MSE = 0.0348	

It is well known that if a regression model has the multicollinearity problem, then the variance of the estimates become very large and therefore identification and prediction problems may occur. The variance inflation factor (VIF) is one of the well known statistics to test whether or not a regression model shows the multicollinearity problem. The VIF associated to AT^2, NP^3,

NP^6, and IF are: 1.47, 4.36, 4.15, and 1.40, respectively. Therefore, the identified model has not multicollinearity problem since the VIF values were less than five. The condition number, defined as the ratio between the maximum eigenvalue divided by the smallest eigenvalue, is another statistic to detect the multicollinearity problem. The eigenvalues associated to the above variables were λ_1=3.21, λ_2= 1.12, λ_3=0.33, λ_4= 0.22, λ_5=0.09. Thus, the condition number is 35.66. Since the condition number is less than one hundred this result confirm that the regression model does not have the multicollinearity problem.

Once model identification task is accomplished residual analysis is conducted to determine whether or not the postulated assumptions on errors are satisfied. Figure 5.1 shows that the variance is unstable, and therefore, a variance stabilization transformation was implemented. Figure 5.2 shows residuals after implementing the power transformation. Once the variance stabilization transformation was implemented the Bartlett test shows that there is enough evidences to accept the hypothesis that the variance is constant, the Bartlett statistic is 3.815 and the critical value at the 95% confidence is 15.51. The Durbin-Watson statistic, 1.88, shows at 95% of confidence that autocorrelation of first order is not serious. Kolmogorov-Smirnov test shows that there is not enough evidences to reject the hypothesis that residual follows a normal distribution, since the Kolmogorov-Smirnov statistic is 0.0455 and the critical value at the 95% confidence is 0.1206. Figure 5.3 exhibits the normality probability plot which confirms this result. Finally, a rough check of outliers can be made by examining the standardized residuals, since no residuals fall beyond the three sigma there is enough evidence to infer that no outlier is present.

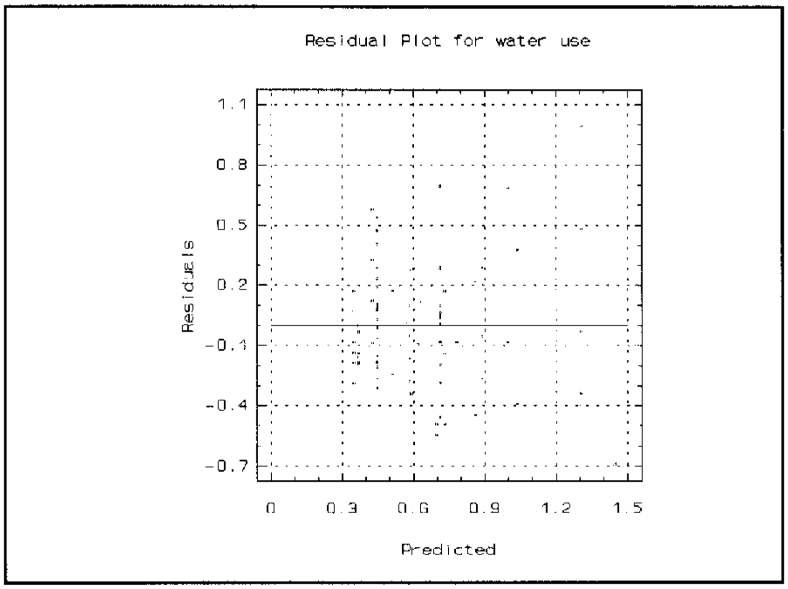

Figure 5.1 Residuals (Original data)

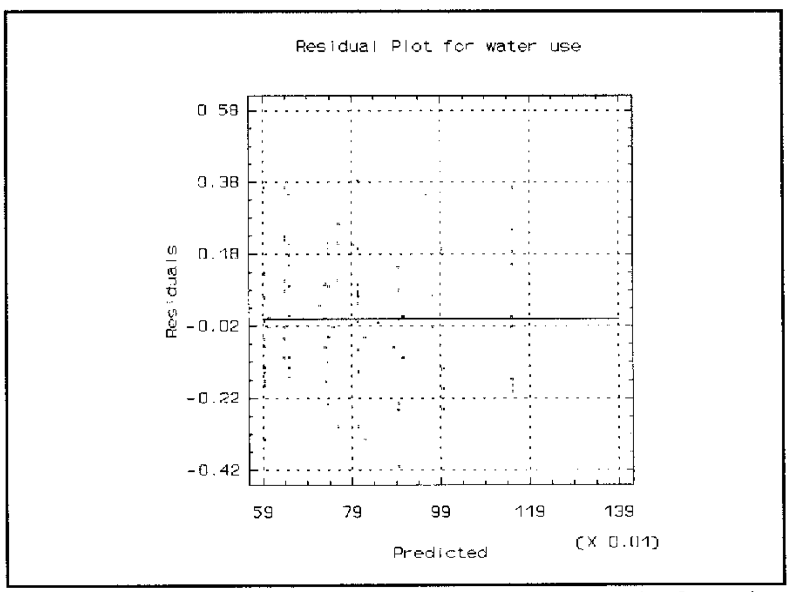

Figure 5.2 Residuals (Transformed data)

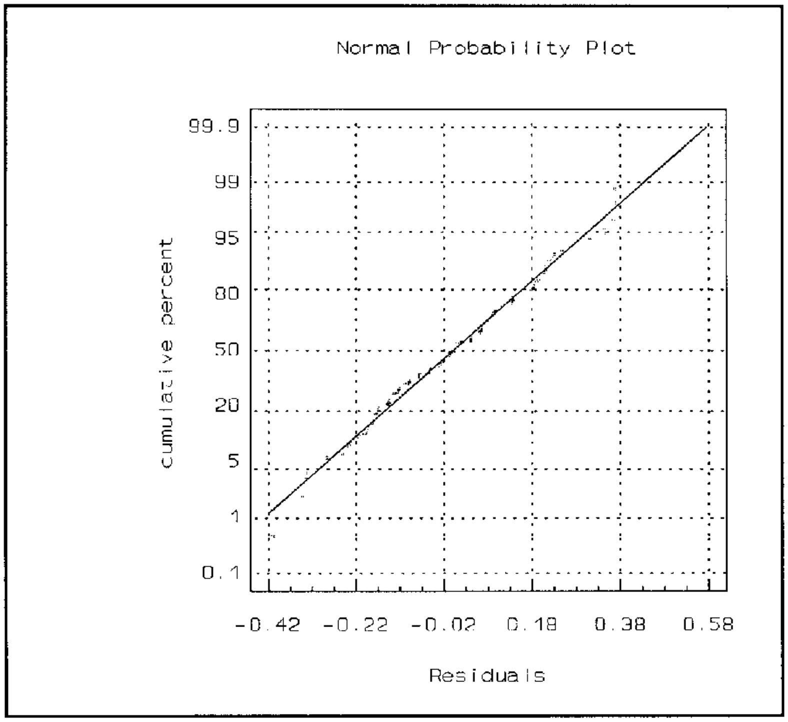

Figure 5.3 Normal probability plot.

5.1.2. A nonlinear econometric model

Since the coefficient of determination for the linear model was small, a nonlinear regression model was studied and the identified model is the following:

$$RW = e^{(a_1 NP)} + \sin\left(\frac{a_2}{NP}\right) + \frac{(e^{b_1} - e^{-b_1})}{2} HI + e^{c_1} e^{B} + e^{(d_1 AT^2)} - \log(e_1) G \qquad (5.2)$$

where the variables RW, NP, HI, B, AT, and G represent the residential water use, the number of people, the annual family income, number of bathrooms, number of automobiles, and garden area in each household, respectively. The values of a's, b_1, c_1, d_1, e_1 are parameters to be estimated from data. A nonlinear optimization algorithm was used to estimate the parameters of the above model and results are presented in Table 5.2.

Table 5.2 Parameter estimation (nonlinear model).

Parameters	**Estimate**	**T**
a_1	-0.0381	-1.93
a_2	-1.4542	-9.23
b_1	9.6E-06	4.04
c_1	-5.5083	-6.21
d_1	-4.0532	-8.43
e_1	0.9350	14.25
R^2 = 0.4774	MSE = 0.0963	

It should be noted that all estimates are significant at the 95% of confidence. The coefficient of determination is better than the linear model. This coefficient shows that 47% of the water-use variability is explained by the model. The Durbin-Watson statistic, 2.014, indicates that data do not reveal evidence of the first order autocorrelation. Kolmogorov-Smirnov statistic, 0.0842, indicates that there is not enough evidence to reject the hypothesis that residuals follow a normal distribution (critical value 0.12 at 95% of confidence). This result is also confirmed with the normal probability plot exhibited on Figure 5.4. The Bartlett statistic, 7.0487, indicates that no power transformation is needed (critical value 15.51, at 95%). Furthermore, standardized residuals do not reveal the presence of outliers.

The eigenvalues associated to the explanatory variables are: λ_1=4.43, λ_2=0.56, λ_3=0.46, λ_4=0.21, λ_5=0.19, and λ_6=0.13. Thus, the condition number is 34.07, which indicates that the regression model does not show the multicollinearity problem.

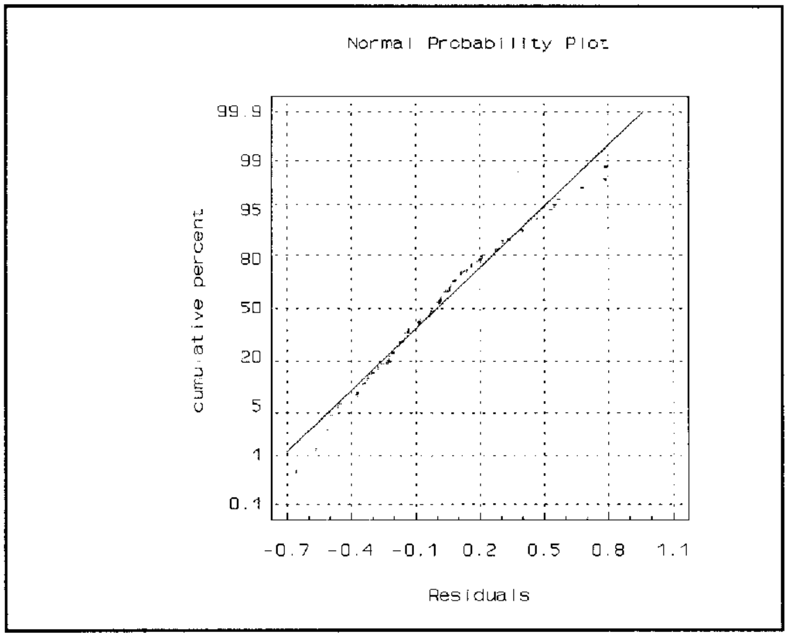

Figure 5.4 Normal probability plot.

5.2. Industrial water use models.

Five different models were fitted to the monthly water use data and the best model in terms of fitting and prediction capabilities is selected. Eight years of monthly water use data were used (January, 1984 to December 1991). The fitted models are: dynamic regression, ARIMA, transfer function, vector ARMA, and the state space models. The variables studied are residential, commercial, and industrial water use. Detailed description for model identification associated with the industrial water use is presented in this chapter.

5.2.1. The Dynamic regression model for industrial water use.

There are many variables which explain the consumption of industrial water use. However, historical records for those variables do not exist. The available monthly records were temperature, rainfall and number of workers. Rainfall variable was not included in the models since this variable has no correlation with the industrial water use. On the other hand, temperature and number of workers were variables used to explain the industrial water use for the city of Mayaguez Puerto Rico. The SYSNLIN routine of SAS/ETS computer package was used. The identified model has the following form:

$$IW_t = \beta_0 IW_{t-1} + \beta_1 IW_{t-2} + \frac{\beta_2}{IW_{t-3}} + \frac{\beta_3}{IW_{t-4}^2} + e^{\beta_4}\,(TE_t^2 + TE_t)$$
$$+ \frac{\beta_5}{NWI_t} + \frac{\beta_6}{NWI_t^2} + \frac{\beta_7}{TE_t^2} + \frac{\beta_8}{NWI_{t-1}^2} + \frac{\beta_9}{NWI_{t-2}^2} + \frac{\beta_{10}}{TE_{t-2}^3} + a_t \qquad \textbf{(5.3)}$$

where the variables IW_t, NWI_t, and TE_t represent the industrial water use, number of worker, and temperature at time t, respectively, a_t is e sequence of independent random variables with zero mean and constant variance. These variables were observed on monthly basis. The ß's are parameters to be estimated from data. Statistics and parameter estimation are presented in Table 5.3.

Table 5.3 Parameter estimation (dynamic model)

Parameter	Estimate	Prob>T
β_0	0.8986	0.0001
β_1	-0.3734	0.0083
β_2	-6.86E13	0.0255
β_3	-3.45E13	0.0001
β_4	4.3152	0.0003
β_5	-2.63E10	0.0001
β_6	2.61E14	0.0001
β_7	1.31E11	0.0001
β_8	-1.71E13	0.0001
β_9	2.23E13	0.0001
β_{10}	2.21E10	0.0001

Durbin-Watson	$\sqrt{MSE}$	R^2	$\sqrt{MS_p}$
2.048	34018.9	0.6761	21186.4

Table 5.3 shows that all the parameters are significant at the 95% of confidence. The coefficient of determination shows that 67.61% of the variance is explained by the model. The Durbin-Watson statistic shows that residuals do not present autocorrelation of the first order. The values of $\surd$MSE and $\surd MS_p$ are statistics that measure the fitting and prediction capabilities of the model.

5.2.2. The ARIMA Model for industrial water use.

Figure 5.5 shows the sample ACF of the industrial water use. This figure exhibits the pattern of a stationary autoregressive process, i.e., differencing is not needed. Figure 5.6 exhibits the sample PACF for the underlying series and suggests that there may exist three significant parameters at lags one, two, and twelve. Thus, after conducting model fitting and hypothesis testing it was found that parameters at lag one and twelve are significant and the one at lag two was not significant. Therefore, the fitting model can be written as follows:

$$IW_t^{(\lambda)} = \mu + \frac{a_t}{(1-\phi_1 B-\phi_{12}B^{12})}$$

$$\text{where} \qquad \textbf{(5.4)}$$

$$E(IW_t^{(\lambda)}) = \mu = \frac{\theta_0}{1-\phi_1-\phi_{12}}$$

where $IW_t^{(\lambda)}$ is the transformed industrial water use, with the parameter of transformation λ=2. The computer program, Autobox, was used to obtain a nonlinear parameter estimation. Table 5.4 presents parameter estimation and some statistics which describe the fitting and prediction capabilities of the model. It should be made clear that all the parameters are significant at 95%

of confidence.

Table 5.4. Parameter estimation (ARIMA model)

Parameter	Estimate	T Ratio
θ_0	45,866.56	14.12
ϕ_1	0.7625	11.92
ϕ_{12}	-0.1557	-2.33

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
2264.07	31500.33	39945.4

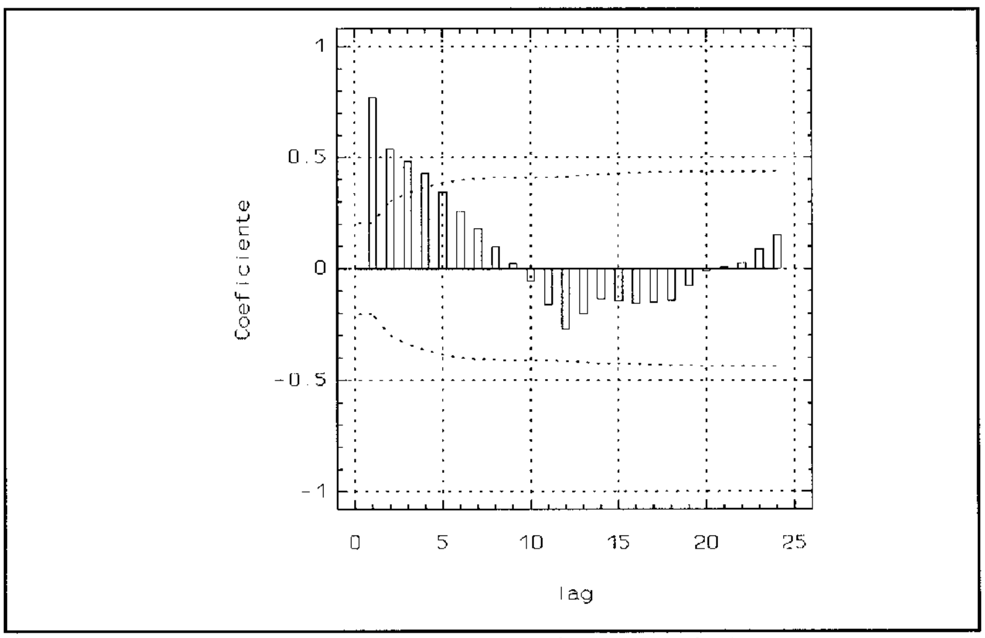

Figure 5.5. Sample ACF for industrial water use.

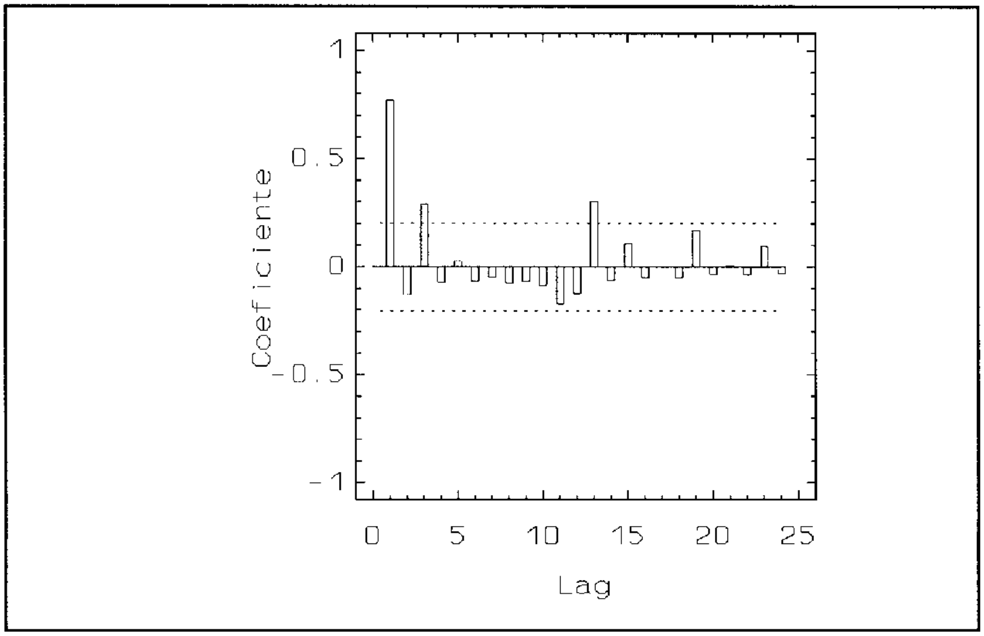

Figure 5.6. Sample PACF for industrial water use.

5.2.3. The transfer function model for industrial water use.

Number of workers and temperature were used as input variables. The structure of the prewhitening models are the following:

$$(1+0.49B)(1-B)NWI_t = \alpha_{1,t}$$

$$and \tag{5.5}$$

$$\frac{(1+1.57B+0.82B^2)(TE_t-7.63)}{1-0.62B} = \alpha_{2,t}$$

where α's are the residuals of the corresponding time series. After implementing the identification procedure described in section 4.2.3. the transfer function is:

where $IW_t^{(\lambda)}$ is the transformed industrial water use, with λ=2, NWI_t is the number of industrial

$$IW_t^{(\lambda)} = -\omega_{0,1}(1-B)NWI_t + (-\omega_{4,2}B^4 - \omega_{6,2}B^6)TE_t + \frac{a_t}{(1-\phi_1 B)(1-\phi_2 B^2)(1-\phi_{12}B^{12})} \quad \textbf{(5.6)}$$

workers at time t, TE_t is the average temperature at time t, and t is the time index given in months. Parameter estimation was obtained by means of using the computer program Autobox. Table 5.5 shows that estimates are significant at 95% of significance. Fitting and prediction statistics are also exhibits in this table.

Table 5.5 Parameter estimation (TF model)

Series	Parameter	Estimate	T Ratio
NWI_t	$\omega_{0,1}$	-1517127	-2.41
TE_t	$\omega_{4,2}$	-0.116E10	-2.15
	$\omega_{6,2}$	-0.155E10	-3.02
$\hat{a}_t$	ϕ_1	0.9634	18.43
	ϕ_2	-0.3752	-3.31
	ϕ_{12}	-0.4679	-4.47

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
1555.32	30653.3	21922.79

5.2.4. The vector ARMA model for industrial water use.

MTS computer program was use to identify and estimate the vector ARMA model. Table 5.6 presents the matrices for the sample ACF and PACF. The positive or negative sign indicates that the associate component is significant at the 95% level. Whereas, the "." indicates that the associate component is not significant. Matrix sample ACF shows an autoregressive pattern and the matrix sample PACF cut off at lag two which indicate that the industrial water use can be expressed as a vector ARMA(0,2).

Table 5.6 Matrix sample ACF and PACF.

Lag	ACF	signs	PACF	signs
0	1.00 -.33 0.17 -.33 1.00 -.18 0.17 -.18 1.00	+ - . - + - . - +	1.00 -.33 0.17 -.33 1.00 -.18 0.17 -.18 1.00	+ - . - + - . - +
1	0.77 -.27 0.16 -.28 0.69 -.24 0.15 -.14 0.76	+ - . - + - . . +	0.77 -.28 0.15 -.27 0.69 -.14 0.16 -.24 0.76	+ - . - + . . - +
2	0.54 -.28 0.08 -.20 0.67 -.29 0.12 -.10 0.43	+ - . - + - . . +	-.13 0.08 -.03 -.14 0.37 0.03 -.16 -.08 -.36	. . . . + . . . -
3	0.48 -.31 -.02 -.20 0.59 -.22 0.13 -.14 0.05	+ - . - + - . . .	0.29 -.10 0.09 -.14 0.07 -.14 -.08 0.17 -.36	+ . . . . . . + -

Thus, the fitted model can be expressed as follows:

$$\begin{bmatrix} IW_t \\ NWI_t \\ TE_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} - \begin{bmatrix} \theta^1_{11} & \theta^1_{12} & \theta^1_{13} \\ \theta^1_{21} & \theta^1_{22} & \theta^1_{23} \\ \theta^1_{31} & \theta^1_{32} & \theta^1_{33} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \\ a_{3,t-1} \end{bmatrix} - \begin{bmatrix} \theta^2_{11} & \theta^2_{12} & \theta^2_{13} \\ \theta^2_{21} & \theta^2_{22} & \theta^2_{23} \\ \theta^2_{31} & \theta^2_{32} & \theta^2_{33} \end{bmatrix} \begin{bmatrix} a_{1,t-2} \\ a_{2,t-2} \\ a_{3,t-2} \end{bmatrix} + \begin{bmatrix} a_{1,t} \\ a_{2,t} \\ a_{3,t} \end{bmatrix} \tag{5.7}$$

Table 5.7 shows estimates for the parameters and the t-statistics which indicated that only the listed parameters are significant at the 95% of confidence.

Table 5.7. Parameter estimation (vector ARMA model)

Parameter	Estimate	t-ratio
c_1	207950	3.72
c_2	19742	12.32
c_3	78.59	3.50
θ_{11}^{1}	0.9496	9.09
θ_{22}^{1}	0.5329	5.16
θ_{33}^{1}	0.7978	7.28
θ_{22}^{2}	0.5337	5.13

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
2040.9	39668.73	31322.86

5.2.5. The state space model for industrial water use.

The identification procedure of state space model was implemented by means of using the routine STATESPACE of the computer program SAS/ETS. The first step is to search for the optimal order in the autoregressive (AR) model fitting of the data. Because AIC is minimum at p=3, the optimal AR order is chosen to be 3. Table 5.8 shows the AIC values for different AR model fits.

Table 5.8. The optimum value of p

p	0	1	2	3	4
AIC	3581.7	3368.3	3348.6	3343.3	3351.9
p	5	6	7	8	9
AIC	3355.1	3368.4	3378.1	3386.5	3392.1

Let $\mathbf{Y}_t=[NWT_t, TE_t, IW_t]'$ for t=1,..., 96 be a stationary multivariate time series. Based on p=3, the canonical correlation analysis between the data space $\{\mathbf{Y}_n, \mathbf{Y}_{n-1}, \dots, \mathbf{Y}_{n-3}\}$ and the predictor space $\{\mathbf{Y}_n, \mathbf{Y}_{n+1|n}, \ldots, \mathbf{Y}_{n+3|n}\}$ is computed. Although $\mathbf{Y}_{n+1|n}, \ldots, \mathbf{Y}_{n+3|n}$ are potential elements of the state vector, this does not mean that all their components $NWT_{n+i|n}$, $TE_{n+i|n}$, $IW_{n+i|n}$, for i=1,2,3 are needed in the final state space vector. Thus, the canonical correlation analysis is performed between

$$\{NWT_{n-i}, TE_{n-i}, IW_{n-i}, i=0,1,2,3\}$$

and

$$\{NWT_{n+i|n}, TE_{n+i|n}, IW_{n+i|n}, i=0,1,2,3\}$$

as is shown in Table 5.9 .

Table 5.9 Canonical correlation analysis.

State vector	Correlation	C	χ^2	df
NWT_n, TE_n, IW_n, $NWT_{n+1|n}$	1.00, 1.00, 1.00, 0.46	5.56	22.45	9
NWT_n, TE_n, IW_n, $NWT_{n+1|n}$, $TE_{n+1|n}$	1.00, 1.00, 1.00, 0.57, 0.45	6.60	21.65	8
NWT_n, TE_n, IW_n, $NWT_{n+1|n}$, $TE_{n+1|n}$, $IW_{n+1|n}$	1.00, 1.00, 1.00, 0.59, 0.46, 0.24	-7.94	5.83	7
NWT_n, TE_n, IW_n, $NWT_{n+1|n}$, $TE_{n+1|n}$, $NWT_{n+2|n}$	1.00, 1.00, 1.00, 0.58, 0.45, 0.20	-9.78	4.06	7
NWT_n, TE_n, IW_n, $NWT_{n+1|n}$, $TE_{n+1|n}$, $TE_{n+2|n}$	1.00, 1.00, 1.00, 0.69, 0.45, 0.28	-5.63	8.03	7

One by one component is added and the information criterion and the chi-square statistic are computed to determine which elements are independent. The component $IW_{n+i|n}$ for $i \geq 1$ must be excluded from the state space vector since C is negative when the component $IW_{n+1|n}$ is considered. Furthermore, when $NWT_{n+i|n}$, $TE_{n+i|n}$, for $i \geq 2$ are considered C becomes negative,

5.3. Residential water use models.

Five different models were fitted to the residential water use for the city of Mayaguez, Puerto Rico. General identification approach was described in Chapter 4, and an example of detailed model identification was presented in the previous section. Comments about model fitting were omitted to avoid repetitions, similar identification and model interpretation were presented in section 5.2.

5.3.1. The dynamic regression model for residential water use.

Model:

$$RW_t = \beta_0 RW_{t-1} + \beta_1 RW_{t-2} + \beta_2 RW_{t-3} + \frac{\beta_3}{RW_{t-4}} + \frac{\beta_4}{(TE_t^3 + TE_t)} + \beta_5 RF^{(1/2)} + \frac{\beta_6}{TE_t^3} + \beta_7 RF_t^3 + \frac{\beta_8}{TE_{t-2}^3} + a_t \quad \textbf{(5.9)}$$

where the variables RW_t, TE_t, and RF_t represent residential water use, average temperature, and rainfall at time t, respectively, t is the time index given in months, and a_t is a white noise process.

Table 5.11 presents the estimates of the ß parameters and some statistics to measure the capabilities of the model.

Table 5.11. Parameter estimation (dynamic model)

Parameter	Estimate	Prob>T
β_0	0.8246	0.0001
β_1	-0.5472	0.0001
β_2	0.4222	0.0009
β_3	4.261E10	0.0001
β_4	6.57E14	0.0001
β_5	-7628.62	0.0299
β_6	-6.58E14	0.0001
β_7	14.266	0.0156
β_8	2.751E10	0.0001

MSE	$\sqrt{MSE}$	R^2	R adj.
827183508	28760.8	0.4965	0.4480

It should be made clear that the estimates are significant at the 95% of confidence.

5.3.2. The ARIMA model for residential water use.

Model:

$$RW_t = \mu + \frac{a_t}{(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)}$$

$$where \qquad (5.10)$$

$$E(RW_t) = \mu = \frac{\theta_0}{(1-\phi_1-\phi_2-\phi_3)}$$

where RW_t represents the residential water use, t is the time index given in months, a_t is a sequence of independent random variables with zero mean and constant variance. ϕ's and θ_0 are parameters to be estimated from data. Estimates and statistics to measure the fitting capabilities are exhibits in Table 5.12.

5.12. Parameter estimates (ARIMA model).

Parameter	Estimate	T Ratio
θ_0	204853.4	3.67
ϕ_1	0.8008	8.37
ϕ_2	-0.46348	-4.17
ϕ_3	0.28144	3.01

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
1919.29	29665.9	40506.63

5.3.3. The transfer function model for residential water use.

Model:

$$RW_t = \omega_{0,1} RF_t + \omega_{4,2} TE_{t-4} + \frac{a_t}{(1 - \phi_1 B - \phi_2 B^2)} \tag{5.11}$$

where RW_t is the residential water use, RF_t rainfall, and TE_t is the average temperature at time t, t is the time index given in months, a_t is a sequence of independent random variables with mean zero and constant variance. Parameter estimation and statistics to test the fitting and prediction capabilities are shown in Table 5.13.

Table 5.13. Parameter estimation (TF model).

Series	Component	Estimate	T Ratio
RF_t	$\omega_{0,1}$	1745.330	2.43
TE_t	$\omega_{4,2}$	6663.76	65.07
a_t	ϕ_1	0.82413	8.2
	ϕ_2	-0.3057	-3.06

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
1862.26	30511.2	31760.06

5.3.4. The vector ARMA model for residential water use.

Model:

$$RW_t = 537000 + \hat{a}_{1,t} + 0.8339\hat{a}_{1,t-1} + 0.1165\hat{a}_{1,t-6}$$

$$RF_t = 6.618 + \hat{a}_{2,t} - 0.1878\hat{a}_{2,t-3} - 2.2858\hat{a}_{1,t-5}$$
$$- 4.2508\hat{a}_{3,t-6} - 10.4529\hat{a}_{3,t-7} + 5.5391\hat{a}_{3,t-11} + 1.774\hat{a}_{1,t-12} \quad \textbf{(5.12)}$$

$$TE_t = 78.59 + \hat{a}_{3,t} + 0.4933\hat{a}_{3,t-1} + 0.2194\hat{a}_{1,t-2} + 0.3255\hat{a}_{3,t-12}$$

where RW_t is the residential water use at time t, RF_t is rainfall at time t, and TE_t is the average temperature at time t, t is the time index given in months, and $[a_{t,1}, a_{2,t}, a_{3,t}]'$ hat is the residual vector at time t. Parameter are significant at 95% confidence level. Fitting and prediction capability of the model are shown in Table 5.14.

Table 5.14 Capabilities of the model (vector ARMA).

AIC	√MSE	$\sqrt{MS_p}$
1862.26	31544.16	29178.31

5.3.5. The state space model for residential water use.

Model:

$$\begin{bmatrix} RF_{(t+1)} \\ TE_{(t+1)} \\ RW_{(t+1)} \\ RF_{(t+2|t+1)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & F_{2,4} \\ F_{3,1} & F_{3,2} & F_{3,3} & F_{3,4} \\ F_{4,1} & F_{4,2} & 0 & F_{4,4} \end{bmatrix} \begin{bmatrix} RF_{(t)} \\ TE_{(t)} \\ RW_{(t)} \\ RF_{(t+1|t)} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ G_{4,1} & G_{4,2} & 0 \end{bmatrix} \begin{bmatrix} a_{1,t+1} \\ a_{2,t+1} \\ a_{3,t+1} \end{bmatrix} \quad \textbf{(5.13)}$$

where RW_t is residential water use at time t, TE_t is the average temperature at time t, RF_t is the rainfall at time t, and t is the time index given in months, $[a_{1,t}, a_{2,t}, a_{3,t}]'$ is a white noise random vector with zero mean and a constant variance covariance matrix. Parameter estimation and statistics to measure fitting and prediction capabilities are given in Table 5.15.

Table 5.15. Parameter estimation (State space model).

Parameter	Estimate	T value
$F_{2,4}$	0.8334	7.23
$F_{3,1}$	-2593.7	-999
$F_{3,2}$	-690.31	-999
$F_{3,3}$	0.5285	6.24
$F_{3,4}$	6040.7	999
$F_{4,1}$	-0.1905	-4.37
$F_{4,2}$	-0.8213	-7.005
$F_{4,4}$	1.6605	27.87
$G_{4,1}$	0.1241	2.86
$G_{4,2}$	0.8927	6.81

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
2062.26	33711.36	33936.77

5.4. Commercial water use models.

Five different models were fitted to the commercial water use for the city of Mayaguez, Puerto Rico. General identification approach was described in Chapter 4, and an example of detailed model identification is presented in section 5.2. Comments about model fitting are omitted to avoid repetitions.

5.4.1. Dynamic model for commercial water use.

Model:

$$CW_t = \beta_0 CW_{t-1} + \frac{\beta_1}{CW^2_{t-2}} + \frac{\beta_2}{CW^2_{t-3}} + \frac{\beta_3}{CW^2_{t-4}} + \beta_4 TE_t + \beta_5 NWC^3_t + \beta_6 NWC_{t-2} + a_t \qquad \textbf{(5.14)}$$

where CW_t is the commercial water use at time t, NWC_t is the number of workers at time t, and TE_t is the average temperature at time t, t is the time index given in months, and a_t is a white noise random process. Parameter estimation and statistics to measure prediction and fitting capabilities are shown in Table 5.16.

Table 5.16 Parameter estimation (dynamic model).

Parameter	Estimate	Prob>T
β_0	0.7785	0.0001
β_1	7.26E13	0.0001
β_2	-1.09E14	0.0001
β_3	7.75E13	0.0001
β_4	622.112	0.0026
β_5	3.69E-08	0.0013
β_6	-6.6290	0.0020

MSE	$\sqrt{MSE}$	R^2	$\sqrt{MS_{p}}$.
44764675	6690.6	0.6351	9706.79

5.4.2. The ARIMA model for commercial water use.

Model:

$$CW_t^{(\lambda)} = \mu + \frac{a_t}{(1-\phi_1 B)}$$

where **(5.15)**

$$E\left(CW_t^{(\lambda)}\right) = \mu = \frac{\theta_0}{(1-\phi_1)}$$

where $CW_t^{(\lambda)}$ is the transformed commercial water use at time t, the transformation parameter λ=-1, and t is the time index given in months, and a_t is a sequence of independent random variables, with zero mean and constant variance. Parameter estimation and statistics to test the fitting and prediction capabilities are shown in Table 5.17.

Table 5.17 Parameter estimation (ARIMA model).

Parameter	Estimate	T Ratio
θ_0	0.247E-05	2.47
ϕ_1	0.768852	11.82

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
1685.11	7034.50	13498.57

5.3.3. The transfer function model for commercial water use.

Model:

$$(1-B)CW_t = \theta_0 + \omega_{0,1}(1-B)NWC_t^{(\lambda)} + \omega_{0,2}TE_t + \frac{(1-\theta_1 B)(1-\theta_{12}B^{12})}{(1-\phi_1 B-\phi_{12}B^{12})}a_t \quad \textbf{(5.16)}$$

where CW_t is the commercial water use at time t, $NWC_t^{(\lambda)}$ is the transformed number of workers at the commercial area at time t, the transformation parameter λ=-1, TE_t is the average temperature at time t, a_t is a white noise process, and t is the time index given in months. Parameter estimation, fitting and prediction capabilities are presented in Table 5.18.

Table 5.18. Parameter estimation (TF model).

Series	Component	Estimate	T Ratio
NWC_t	$\omega_{0,1}$	-0.5947E09	-2.95
TE_t	$\omega_{0,2}$	635.25	2.31
	θ_0	-110559.4	-2.29
a_t	ϕ_1	-0.8653	-7.34
	ϕ_{12}	-0.3301	-3.42
	θ_1	-0.8479	-10.09
	θ_{12}	-0.2658	-2.41

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
1641.47	6561.6	7152.66

5.4.4. The vector ARMA model for commercial water use.

Model:

$$CW_t^{(\lambda)} = 1.0681 + 0.4973\hat{a}_{1,t-1} - 0.0896\hat{a}_{3,t-1} + 0.3205\hat{a}_{1,t-5} + 0.3167\hat{a}_{1,t-6} - 0.3867\hat{a}_{2,t-9} + 0.0721\hat{a}_{3,t-11}$$

$$NWC_t^{\lambda)} = -0.0064NWC_{t-1}^{(\lambda)} - 0.134\hat{a}_{2,t-2} + 0.0834\hat{a}_{1,t-4} - 0.0531\hat{a}_{1,t-6} + 0.0589\hat{a}_{3,t-9} - 0.2272\hat{a}_{2,t-10} + 0.0793\hat{a}_{1,t-11} - 0.0801\hat{a}_{1,t-12} + 0.2443\hat{a}_{2,t-12} \quad \textbf{(5.17)}$$

$$TE_t = 78.63 + 0.5983\hat{a}_{3,t-1} + 0.8108\hat{a}_{2,t-3} - 0.3394\hat{a}_{3,t-6} + 0.5169\hat{a}_{3,t-10} - 1.464\hat{a}_{2,t-12}$$

where $CW_t^{(\lambda)}$ is the transformed commercial water use at time t, $NWC_t^{(\lambda)}$ is the transformed number of workers at the commercial area at time t, the transformation parameter λ=-1, TE_t is the average temperature at time t, and t is the time index given in months, $[a_{1,t}, a_{2,t}, a_{3,t}]'$ hat is the residual vector. Parameters are significant at the 95% of confidence level. Table 5.19 shows the fitting capabilities of the model.

Table 5.19. Fitting capabilities.

AIC	√MSE	$\sqrt{MS_p}$
2041.47	12139.16	13421.91

5.4.5. The state space model for commercial water use.

Model:

$$\begin{bmatrix} NWC_{(t+1)} \\ TE_{(t+1)} \\ CW^{(\lambda)}_{(t+1)} \end{bmatrix} = \begin{bmatrix} F_{1,1} & F_{1,2} & 0 \\ 0 & F_{2,2} & 0 \\ 0 & F_{3,2} & F_{3,3} \end{bmatrix} \begin{bmatrix} NWC_{(t)} \\ TE_{(t)} \\ CW^{(\lambda)}_{(t)} \end{bmatrix} + \begin{bmatrix} a_{1,t+1} \\ a_{2,t+1} \\ a_{3,t+1} \end{bmatrix} \quad \textbf{(5.18)}$$

where $CW_t^{(\lambda)}$ is the transformed commercial water use at time t, the transformation parameter is -1, NWC_t is the number of workers at the commercial area at time t, TE_t is the average temperature at time t, $[a_{1,t}, a_{2,t}, a_{3,t}]'$ is the white noise random vector with zero mean and a constant variance covariance, and t is the time index given in months. Parameter estimation and fitting performances are presented in Table 5.19.

Table 5.19. Parameter estimation (state space model).

Parameter	Estimate	T value
$F_{1,1}$	0.9295	28.05
$F_{1,2}$	34.434	33.82
$F_{2,2}$	0.7553	11.88
$F_{2,3}$	-3E-05	-2.28
$F_{3,2}$	570.98	559.4
$F_{3,3}$	0.7623	11.69

AIC	$\sqrt{MSE}$	$\sqrt{MS_p}$
2018.10	7388.91	8299.19

5.5. The best models to predict Mayaguez water use.

Three variables were studied: industrial, residential and commercial water use and five models were fitted to each monthly water use. Thus, there is a total of fifteen different models. Among those fifteen models three models were selected to represent best the underlaying variables. Fitting and prediction capabilities of each model were measured and the ones that offer the best performances were selected.

Fitting performances is measured by means of the $\sqrt{MSE}$ and prediction capability by means of $\sqrt{MS_p}$. These criterion are defined as follows:

$$MSE = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2 \quad and \quad MS_p = \frac{1}{M}\sum_{i=1}^{M} e_i^2(p) \qquad \textbf{(5.19)}$$

$$where$$

$$e_i = y_i - \hat{y}_i \quad and \quad e_i(p) = y_i - \hat{y}_i(p)$$

where y_i is the observed value, y_i hat is the one-step-ahead forecast computed at origin i (i=1,...95); $y_i(p)$ hat is the forecast with a lead time p (p=1,...,12) made at origin i=84, n=96, M=84, and k is number of parameters involved in the model.

The statistic MSE measures how well the model fits to the given data, and the statistic MS_p measures how well the model predicts the underlying variable. Table 5.20 shows the prediction and fitting capabilities of the fifteen identified models.

Table 5.20. Fitting and prediction capabilities.

Models	Resid. W.		Commer. W.		Indus. W.	
	$\sqrt{MSE}$	$\sqrt{MS_p}$	$\sqrt{MSE}$	$\sqrt{MS_p}$	$\sqrt{MSE}$	$\sqrt{MS_p}$
Dynamic Reg.	28760	23319	6690	9706	34018	21186
ARIMA	29665	40506	7034	13498	31500	39445
Transfer Func.	30511	31760	6561	7152	30653	21922
Vector ARMA	31544	29128	12139	13421	39668	31322
State Space	33711	33936	7388	8299	57800	39063

The model which exhibits the best fitting and prediction capabilities for a particular variable is the one that shows the minimum MSE and the minimum MS_p. Based on this rule the best models are presented in Table 5.21

Table 5.21. The best models.

Times series	The best model	Equation
Residential water use	Dynamic Regression	(5.9)
Commercial water use	Transfer Function	(5.16)
Industrial water use	Transfer Function	(5.6)

5.6. Residual analysis for the best fitted models.

Residuals analysis was conducted on the best models. The purpose of this analysis is to test whether or not the postulated assumptions on the building models are satisfied. Thus, residuals analysis involves to test whether or not residuals follows a normal distribution, with zero mean and constant variance. Residuals also must be independent and should not exhibit outliers.

5.6.1. Residuals from the residential water use model.

The best model for residential water use is given by the dynamic regression model (5.9). Residuals from this model were computed and the following results were found. (1) The Durbin-Watson statistic was 1.927. This statistic indicates that residuals do not present problem of autocorrelation of the first order. (2) Kolmogorov-Smirnov test was used to test normality. The statistic of Kolmogorov-Smirnov was 0.118 and the critical value is 0.12657. Since the statistic is smaller than the critical value, normality cannot be rejected at the 95% confidence level. This result was also confirmed with the normal probability plot, which is exhibited in Figure 5.7. (3) A rough check for outliers was made by examining the standardized residuals, since standardized residuals are smaller than +- three standard deviation it can be conclude that there are not potential outliers.

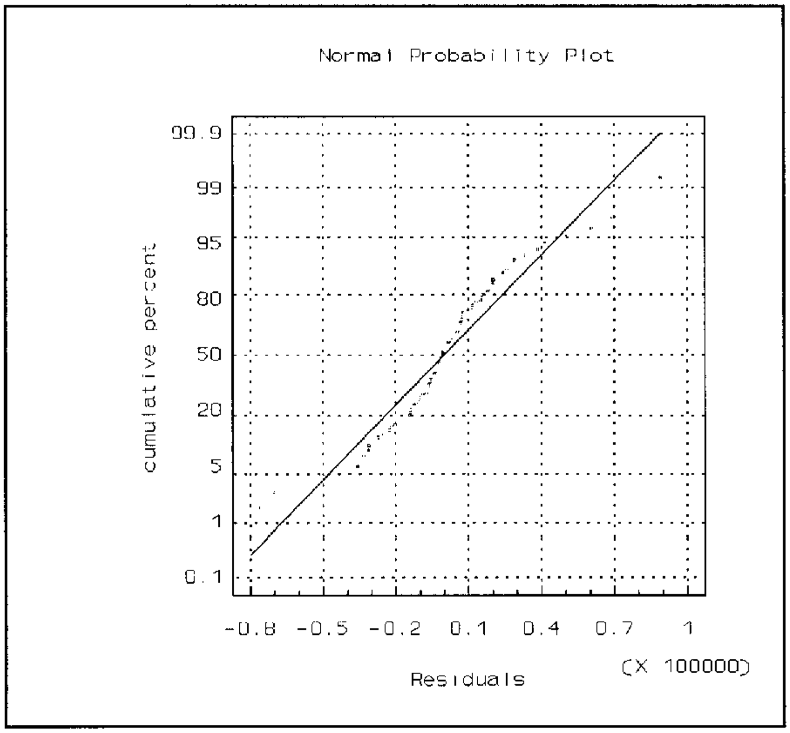

Figure 5.7 Residuals regression model.

5.6.2. Residuals from the commercial water use model.

The assumptions made on the transfer function model building were checked by studying the autocorrelation and cross-correlation functions:

(1) Autocorrelation check. Figures 5.8, and 5.9 show the sample autocorrelation and partial autocorrelation functions, respectively. These figures show that residual behave as a sequence of independent random variables. The Portmanteau test also confirms this result. The Portmanteau statistic is 12.34 and the critical value at the 95% level of confidence is 31.41. Therefore, there is not enough evidence to reject the hypothesis that residuals behave as a white noise series. This fact show evidence that one of the major assumptions made on the error term is satisfied.

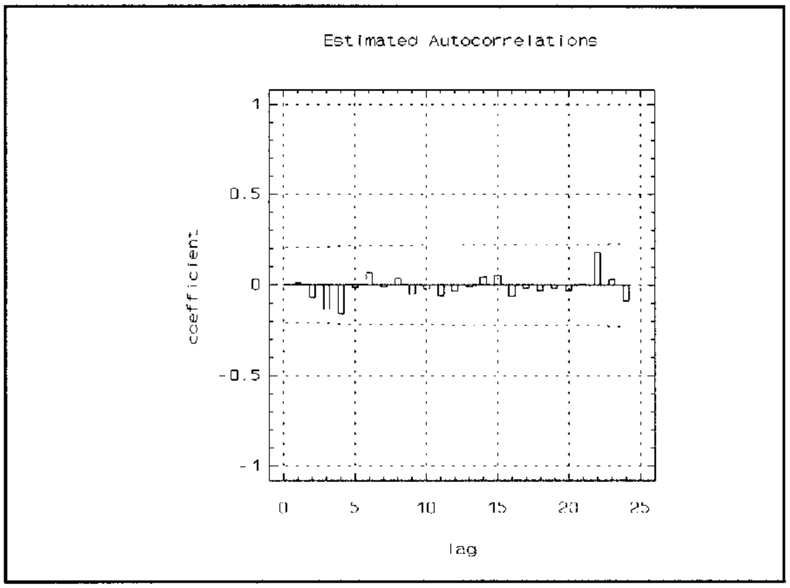

Figure 5.8 Autocorrelation function

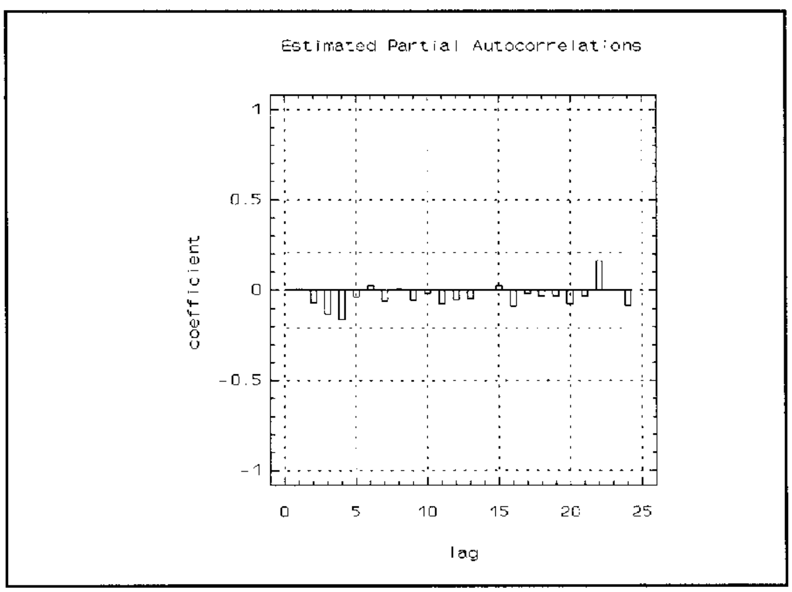

Figure 5.9 Partial autocor. function.

(2) Cross-correlation check.

Another important assumption made on model building was that errors must be independent with respect to the input variables. Tables 5.22 and 5.23 show the sample cross-correlation function between residuals and the input variables. These tables indicate that residuals are independent of the input variables. This result also are confirmed with the Portmanteau test. The Portmanteau statistic is 6.062 and the critical value at the 95% level of significance is 19.67, i.e., it can be conclude that residuals are independent of NWC variable. Furthermore, the Portmanteau test for the temperature reveals the following results. The Portmanteau statistic is 4.015 and the critical value is 19.67 at 95% of level of significance; hence, residuals are also independent of temperature. Therefore, the developed transfer function model for the commercial water use is a satisfactory model.

Table 5.22 Cross-correlation (residuals vs NWC)

Lag	Cross Correl.	Stand. Error.	T Ratio	Lag	Cross Correl.	Stand. Error.	T Ratio
0	-0.027	0.11	-0.25	0	-0.027	0.11	-0.25
1	-0.147	0.11	-1.33	-1	-0.068	0.11	-0.62
2	0.062	0.111	0.56	-2	0.205	0.111	1.84
3	0.05	0.112	0.44	-3	0.064	0.112	0.57
4	0.03	0.113	0.26	-4	-0.055	0.113	-0.49
5	-0.045	0.113	-0.39	-5	-0.14	0.113	-1.24
6	-0.1	0.114	-0.88	-6	0.058	0.114	0.51
7	-0.031	0.115	-0.27	-7	-0.026	0.115	-0.22
8	0.042	0.115	0.36	-8	0.015	0.115	0.13
9	0.125	0.116	1.07	-9	0.028	0.116	0.24
10	0.013	0.117	0.11	-10	0.111	0.117	0.95
11	0.021	0.118	0.17	-11	-0.064	0.118	-0.54

Table 5.23 Cross-correlation (residuals and TE).

Lag	Cross Correl.	Stand. Error.	T Ratio	Lag	Cross Correl.	Stand. Error.	T Ratio
0	-0.001	0.104	-0.01	0	-0.001	0.104	-0.01
1	0.024	0.104	0.23	-1	0.011	0.104	0.1
2	0.008	0.105	0.08	-2	-0.173	0.105	-1.65
3	-0.057	0.105	-0.54	-3	0.024	0.105	0.23
4	-0.068	0.106	-0.65	-4	0.007	0.106	0.07
5	-0.058	0.107	-0.54	-5	-0.044	0.107	-0.41
6	-0.013	0.107	-0.12	-6	-0.094	0.107	-0.88
7	0.075	0.108	0.69	-7	0.023	0.108	0.21
8	0.058	0.108	0.54	-8	-0.007	0.108	-0.06
9	0	0.109	0	-9	0.047	0.109	0.43
10	0.092	0.11	0.84	-10	0.036	0.11	0.33
11	-0.096	0.11	-0.86	-11	-0.142	0.11	-1.28

5.6.3. Residuals from the industrial water use model.

The assumptions made on the transfer function modeling were checked on the autocorrelation and cross-correlation functions:

(1) Autocorrelation check. Figure 5.10 shows the sample partial autocorrelation function. This figure shows that residual behave as a sequence of independent random variables. The Portmanteau test also confirms this result. The Portmanteau statistic is 12.45 and the critical value at the 95% level is 32.67, therefore, there is not enough evidence to reject the hypothesis that residuals behave as a white noise series. This fact show evidence that the one of the major assumption made on the error term was satisfied.

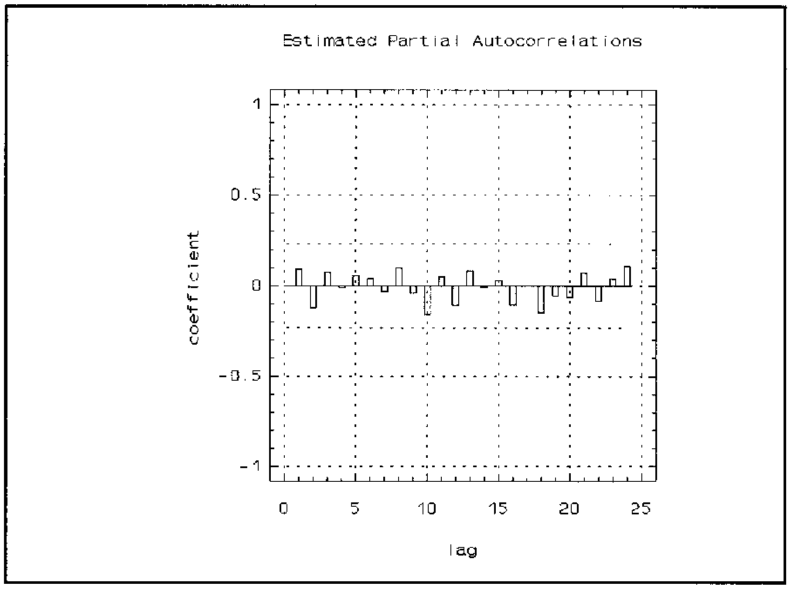

Figure 5.10 Partial autocor. function

(2) Cross-correlation check.

Another important assumption made on model building is that errors must be independent with respect to the input variables. Tables 5.25 and 5.26 show the sample cross-correlation function between residuals and the input variables. These tables indicate that residuals are independent of the input variables. This result also is confirmed with the Portmanteau test. The Portmanteau statistic is 5.909 and the critical value at the 95% level of significance is 19.67, i.e., it can be conclude that residuals are independent of NWI variable. Furthermore, the Portmanteau test for the temperature reveals the following results. The Portmanteau statistic is 15.564 and the critical value is 19.67 at 95% of level of significance; hence, residuals are also independent of temperature. Therefore, the developed transfer function model for the industrial water use is a satisfactory model.

Table 5.25 Cross-correlation (residuals vs NWI)

Lag	Cross. Correl.	Stand. Error	T Ratio	Lag	Cross. Correl.	Stand. Error	T Ratio
0	-0.012	0.115	-0.1	0	-0.012	0.115	-0.1
1	0.024	0.116	0.2	-1	-0.171	0.116	-1.47
2	0.028	0.117	0.24	-2	0.102	0.117	0.87
3	-0.169	0.118	-1.43	-3	-0.001	0.118	0
4	-0.107	0.119	-0.9	-4	-0.071	0.119	-0.6
5	-0.057	0.12	-0.48	-5	0.125	0.12	1.04
6	-0.085	0.12	-0.71	-6	0.193	0.12	1.61
7	-0.108	0.121	-0.89	-7	-0.01	0.121	-0.08
8	0.039	0.122	0.32	-8	-0.109	0.122	-0.89
9	0.073	0.123	0.59	-9	-0.234	0.123	-1.9
10	0.022	0.124	0.18	-10	0.005	0.124	0.04
11	-0.027	0.125	-0.22	-11	0.063	0.125	0.5

Table 5.26 Cross-correlation (residuals vs TE)

Lag	Cross. Correl	Stand. Error.	T Ratio	Lag	Cross. Correl	Stand. Error.	T Ratio
0	0.271	0.115	2.35	0	0.271	0.115	2.35
1	0.193	0.116	1.66	-1	0.03	0.116	0.25
2	-0.014	0.117	-0.12	-2	0.136	0.117	1.16
3	-0.107	0.118	-0.91	-3	0	0.118	0
4	-0.141	0.119	-1.19	-4	0.135	0.119	1.13
5	-0.069	0.12	-0.57	-5	0.05	0.12	0.42
6	0.067	0.12	0.56	-6	0.023	0.12	0.19
7	-0.037	0.121	-0.31	-7	0.023	0.121	0.19
8	-0.162	0.122	-1.33	-8	0.044	0.122	0.36
9	0.106	0.123	0.86	-9	0.068	0.123	0.55
10	0.016	0.124	0.13	-10	-0.226	0.124	-1.82
11	-0.113	0.125	-0.91	-11	0.094	0.125	0.75

5.7. Mayaguez monthly water use forecasts.

(1) Residential water use forecasts.

The forecasts for residential water use can be easily computed by using the following difference equation:

$$\begin{aligned} RW_t = {} & 0.8246*RW_{t-1} - 0.5472*RW_{t-2} + 0.4222*RW_{t-3} + 4.2615E010/RW_{t-4} \\ & + 6.55757E14/(TE_t^3 + TE_t) - 7628.62*RF_t^{(1/2)} \\ & - 6.5744E14/TE_t^3 + 14.26*RF_t^3 + 2.7519E10/TE_{t-2}^3 \end{aligned} \qquad \textbf{(5.20)}$$

using data exhibit in Appendix C forecasts for 1992 were computed and results are shown in Table 5.27.

(2) Commercial water use forecasts.

Forecasts for commercial water use can be easily computed by using the following equation:

$$\begin{aligned} CW_t = & -110559.4+0.1347*CW_{t-1}+0.535*CW_{t-2}+0.3302*CW_{t-3} \\ & +0.59476E09*NWC_t^{(\lambda)}-0.08014E09*NWC_{t-1}^{(\lambda)}-0.318256E09*NWC_{t-2}^{(\lambda)} \\ & -0.196389E09*NWC_{t-3}^{(\lambda)}+635.26*TE+549.71*TE_{t-1}+209.73*TE_{t-2} \\ & +0.8479*\hat{a}_{t-1}+0.2659*\hat{a}_{t-12}+0.2254*\hat{a}_{t-13} \end{aligned} \qquad (5.21)$$

where $NWC_t^{(\lambda)}$ is the transformed number of worker at the commercial area at time t, the transformation parameter equal to -1, CW_t is the commercial water use at time t, TE_t is the average temperature at time t, t is the time index given in months, and a_t hat are the residual at time t. Using data exhibit in Appendix C forecasts for 1992 were computed and results are shown in Table 5.27.

(3) Industrial water use forecasts.

Forecasts for industrial water use can be easily computed by using the following equation:

$$\begin{aligned} IW_t^{(\lambda)} = & \ 0.9635*IW_{(t-1)}^{(\lambda)}-0.3752*IW_{(t-2)}^{(\lambda)}+0.3615*IW_{(t-3)}^{(\lambda)}-0.468*IW_{(t-12)}^{(\lambda)} \\ & +0.4509*IW_{(t-13)}^{(\lambda)}-0.1756*IW_{(t-14)}^{(\lambda)}+0.1692*IW_{(t-15)}^{(\lambda)}-0.1517E07*NWI \\ & +2978751*NWI_t-2025977*NWI_{(t-2)}+1117667*NWI_{(t-3)}-548441*NWI_{(t-4)} \\ & -710015*NWI_{(t-12)}+1394088*NWI_{(t-13)}-950480*NWI_{(t-14)} \\ & +523105*NWI_{(t-15)}-256697*NWI_{(t-16)}-0.11694E11*TE_{(t-4)} \\ & +1126716900*TE_{(t-5)}+1069841120*TE_{(t-6)}-1078009500*TE_{(t-7)} \\ & +584411520*TE_{(t-8)}-563072400*TE_{(t-9)}+527282460*TE_{(t-17)} \\ & +523610160*TE_{(t-18)}-504459360*TE_{(t-19)}+273514560*TE_{(t-20)} \\ & -263545920*TE_{(t-21)} \end{aligned} \qquad (5.22)$$

where $IW_t^{(\lambda)}$ is the transformed industrial water use at time t, the transformation parameter is 2, NWI_t is the number of workers at time t, TE_t is the average temperature at time t, and t is the time index given in months. Using data exhibit in Appendix C forecasts for 1992 were computed and results are shown in Table 5.27.

Table 5.27 Monthly water use forecasts.

Month	Residential Forecasts (m^3)	Commercial Forecasts (m^3)	Industrial Forecasts (m^3)
01/92	551,703	107,600	263,400
02/92	536,909	104,400	265,100
03/92	526,909	97,980	274,100
04/92	530,820	96,980	275,000
05/92	540,546	102,400	266,700
06/92	544,369	103,200	247,800
07/92	546,240	102,900	217,300
08/92	547,633	103,500	208,600
09/92	549,777	104,400	174,800
10/92	548,170	104,600	181,200
11/92	546,126	103,300	194,400
12/92	541,413	103,600	199,400

6. CONCLUSIONS AND RECOMMENDATIONS.

A sampling procedure provides enough information to derive a linear and a nonlinear econometric model to express the residential water use for the city of Mayaguez, Puerto Rico. The nonlinear model has exhibited better prediction capabilities. It was identified that the following variables help in explaining the residential water use: number of automobiles, number of people, annual income, and garden area in each household.

It should be noted that econometric models still reveal a poor coefficient of determination; however, these coefficients are much better than the model found by Guilbe (1969). Thus, if a further study would be conducted, it is recommended that the sample size should be increased and at least four observation should be taken from each house.

Fifteen models were developed to represent the stochastic behavior of residential, commercial, and industrial water use. These models have the capability to obtain water use monthly predictions. Out of these fifteen models three difference equations were selected to express the underlying water use variables. Thus, the dynamic regression model is the model that best represents the stochastic behavior of residential water use. The transfer function model shows better prediction capabilities and fitting conformance to express both the commercial and the industrial water use.

The rapid growth of Mayaguez Puerto Rico cause sever strains on the municipal water supply system. To develop a rational urban growth plan it is necessary to estimate the water need for either new urbanizations or new industrial parks. Increasing interest has been raised in determining the urban water needs. This interest has resulted from the necessity to avoid future water shortages. An adequate potable water supply is a primary factor in the well-being and economic progress of the city of Mayaguez, Puerto Rico. Model scope is limited to the city of Mayaguez. An extension to other parts of Puerto Rico may require to implement the sampling procedure and/or update the derived models to the current data. Approximate prediction may be obtained from the Mayaguez models, after using the appropriate input values.

REFERENCES.

Agthe, D.E., and Billings, R.B. (1980). Dynamic Models of Residential Water Demand." Water Resources Research, 16, 3, 476-480.

Attanasi, E.D., Close, E.R., and Lopez, M.A. (1975). "Techniques for Water Demand Analysis and Forecasting: Puerto Rico, A Case Study." US Department of the Interior Geological Survey, Open-file Report 75-94.

Box, G.E.P., and Cox, D.R. (1964). "An Analysis of Transformations," Journal of the Royal Statistical Society, B-26, 211-253.

Daniel, C., and Wood, F.S. (1980). Fitting Equations to Data, 2nd Ed. John Wiley, New York.

Domokos, M., Weber, J., and Duckstein, L., (1976). "Problems in Forecasting Water Requirements," Water Resources Bulletin, 12, 2, 263-275.

Draper, N.R., and Smith, H. (1981). Applied Regression Analysis, 2nd Ed., John Wiley, New York.

Fuller, W.A. (1976). Introduction to Statistical Time Series, John Wiley, New York.

Granger, C.W.J., and Newbold, P. (1974). "Spurious Regressions in Econometrics," Journal of Econometrics, 2, 111-120.

Guilbe, A., (1969). "Quantitative Analysis of Residential Water Use Patterns, in Mayaguez, Puerto Rico," Water Resources Research Institute, University of Puerto Rico, Project No. A-003-PR.

Guilbe, A., (1972). "Comparative Analysis of Residential Water Use in Puerto Rico," Water Resources Research Institute, University of Puerto Rico, Report A-025-PR.

Hansen, R.D., and Narayanan, R. (1981). "A Monthly Time Series Model of Municipal Water Demand," Water Resources Bulletin, 17, 4, 578-585.

Haugh, L.D., and Box, G.E.P. (1977). "Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series," Journal of the American Statistical Association, Vol. 72, No. 387, 121-130.

Kuttan, A., (1970). "Determination of the Significance of Variables Affecting Residential Water Consumption," Water Resources Research Institute, University of Puerto Rico, Report A-022-PR.

Lobb, H.J., (1975). "Demand-Rate Economics, "Journal of the American Water Works Association, 67, 5, 246-256.

Maidment, D.R., and Parzen, E. (1984). "Cascade Model Monthly Municipal Water Use," Water Resources Research, 20, 1, 15-23.

Maidment, D.R., Miaou, S.P., and Crawford, M.M. (1985). "Transfer Function Models of Daily Urban Water Use," Water Resources Research, 21, 4, 425-432.

Maidment, D.R., Miaou, S.P. (1986). "Daily Water Use in Nine Cities," Water Resources Research, 22, 6, 845-851.

Mallows, C.L. (1973). "Some Comments on C_p," Technometrics, 15, 661-675.

Meyer, J.M. Jr., and Mangan, G.F. (1969). "System Pinpoints Urban Water Needs," Environmental Science and Technology, 3, 10, 904-911.

Montgomery, D.C., and Peck, E.A. (1982). Introduction to Linear Regression Analysis, John Wiley, New York.

Sastri, T. (1987). "Sequential Method of Change Detection and Adaptive Prediction of Municipal Water Demand," International Journal of Systems Science, 18, 6,1029-1049.

Smith, J.A. (1988). "A Model of Daily Municipal Water Use for Short-Term Forecasting," Water Resources Research, 24, 2, 201-206.

Thompson, A.G., Smith, V.E., and Colvin, W.R. (1976). "Development of Commercial Institutional Parameter Units for the MAIN II System of Water Demand Forecasting," Water Resources Research Institute, University of Wyoming.

Tiao, G.C., and Box, G.E.P. (1981). "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, Vol. 76, 802-816.

Weber, A.J. (1989). "Forecasting Demand and Measuring Price Elasticity," Journal of the American Water Works Association, 30, 7, 57-65.

Whitford, P.W. (1972). "Residential Water Demand Forecasting," Water Resources Research, 8, 4, 829-839.

Appendix A: Questionnaire.

Universidad de Puerto Rico
Mayaguez, Puerto Rico
Desarrollo de Técnicas que permiten predecir
el Consumo de Agua para Planificación Urbana

Fecha ___ / ___ / ___

1) CT ___ ___ ___
___ ___ ___

Identificación del Area
a) Sector _______

Nombre ____________
Dirección ____________
____
Teléfono ____________

2) ¿ Cuántas personas viven en la casa ?
- a) uno a tres ( )
- b) cuatro a seis ( )
- c) siete a diez ( )
- d) once o más ( )

3) ¿ Es usted propietario de la casa?
- a) Sí ( )
- b) No ( )

4) ¿ Cuál es su ingreso familiar ?
- a) menos de $1000 ( )
- b) $1001 a $5000 ( )
- c) $5001 a $15000 ( )
- d) $15001 a $30000 ( )
- e) $30000 ó más ( )

5) ¿ De dónde proviene el agua que usted consume ?
- a) Acueducto Público ( )
- b) Pozo privado ( )
- c) Cisterna ( )
- d) Otros ( )

6) ¿ Cuántos baños con ducha hay en la casa ?
- a) cero ( )
- b) uno ( )
- c) dos ( )
- d) tres ( )
- e) cuatro ó más ( )

7) ¿ Cuántos baños sin ducha hay en en la casa ?
- a) cero ( )
- b) uno ( )
- c) dos ( )
- d) tres ( )
- e) cuatro ó más ( )

8) ¿ Usa fregadero en la cocina ?
- a) Sí ( )
- b) No ( )

9) ¿ Cómo lava la ropa ?
- a) a máquina ( )
- b) en pileta ( )

10) ¿ Tiene jardín en su casa ?
- a) Sí ( )
- b) No ( )

11) ¿ Con qué frecuencia riega el jardín ?
- a) Diario ( )
- b) Cada 3 dias ( )
- c) Semanal ( )

12) ¿ Cuál es el área aproximada del jardín ?
- a) 16 m^2 ó menos ( )
- b) 16.1 m^2 a 25 m^2 ( )
- c) 25.1 m^2 a 35 m^2 ( )
- d) 35.1 m^2 a 48 m^2 ( )
- e) 48.1 m^2 ó más ( )

13) ¿ Cuántos automóviles posee ?
- a) cero ( )
- b) uno ( )
- c) dos ó más ( )

14) ¿ Con qué frecuencia lava el carro ?
- a) semanal ( )
- b) quincenal ( )
- c) mensual ( )

15) ¿ Tiene piscina ?
- a) Sí ( )
- b) No ( )

16) ¿ Con qué frecuencia cambia el agua de la piscina ?
- a) cada 3 meses ( )
- b) cada 6 meses ( )
- c) cada 9 meses ( )
- d) anual ( )

17) ¿ Tiene animales en su casa ?
- a) Sí ( )
- b) No ( )

18) ¿ Cuánto paga usted en promedio por el servicio de agua ?
- a) menos de $10 ( )
- b) $10 a $15 ( )
- c) $ 15 a $20 ( )
- d) $ 20 a $30 ( )
- e) $ 30 ó más ( )

19) ¿ Qué tiempo lleva construida su casa ?
- a) 0 a 10 años ( )
- b) 10 a 20 años ( )
- c) 20 a 30 años ( )
- d) 30 años ó más ( )
- e) no sabe ( )

20) ¿ En cuánto estima el valor de su propiedad ?

____

Observaciones : ___________________
____
____
____

Appendix B: Socioeconomic data.

	Water U. (m^3)	N. Peop.	Income ($)	Bathr.	Autom.	Garden
1	0.950	5	10000	1	3	1
2	0.267	2	3000	1	3	0
3	0.516	2	3000	1	0	1
4	0.916	2	3000	1	0	1
5	1.000	5	10000	1	0	1
6	0.433	5	10000	1	1	1
7	0.209	2	3000	1	0	1
9	0.800	5	10000	1	3	1
10	0.868	2	10000	1	3	0
11	0.912	2	22500	1	3	1
12	0.433	2	10000	1	1	1
13	0.367	2	3000	1	1	1
15	0.257	5	10000	2	1	1
16	0.197	2	3000	1	0	1
17	0.288	2	3000	1	0	0
18	0.754	2	10000	1	0	0
19	0.067	2	3000	2	0	0
20	0.426	2	3000	1	0	1
21	1.410	5	10000	2	1	1
22	0.733	2	10000	2	0	1
23	0.377	2	10000	1	0	0
25	0.525	5	3000	1	0	0
26	0.279	2	10000	1	0	0
27	0.534	2	10000	1	1	0
28	0.574	2	22500	1	0	0
29	0.574	5	10000	2	0	1
30	1.006	2	10000	1	0	1
31	0.178	2	3000	2	1	1
32	0.217	2	3000	1	0	0
33	0.176	2	3000	2	0	0
34	0.262	2	10000	1	1	1
35	0.633	2	50000	3	1	1
36	0.525	2	10000	2	1	0
37	0.639	2	10000	3	1	0
38	0.902	2	10000	3	3	1
39	0.426	2	3000	2	0	1
40	0.900	2	22500	2	3	1
42	0.733	2	50000	3	3	1
43	0.316	5	10000	1	0	0
44	0.916	5	22500	1	3	1
45	1.185	2	50000	4	1	1
46	1.050	5	22500	2	3	1
47	0.650	5	22500	1	1	1
48	0.476	2	10000	1	1	1
49	0.985	2	10000	1	1	1
50	1.405	2	50000	4	3	1
51	0.683	2	22500	2	1	1
52	0.835	2	50000	2	1	1
53	0.988	5	22500	1	3	1
54	0.755	8.5	50000	2	3	1
55	1.438	5	22500	2	3	1
56	0.920	2	10000	1	1	1
57	1.000	5	10000	1	1	0
58	1.083	5	22500	1	3	1
59	0.534	2	22500	1	1	0
60	1.076	5	10000	1	3	1
61	0.754	5	10000	1	1	1
62	0.850	2	10000	1	1	1
63	0.762	5	10000	1	1	1
64	2.300	5	50000	3	3	1
65	0.265	2	10000	2	1	1
66	1.700	5	50000	2	3	1

Appendix B: Socioeconomic data.

	Water U. (m^3)	N. Peop.	Income ($)	Bathr.	Autom.	Garden
67	0.852	2	10000	2	3	0
68	0.516	2	22500	2	3	1
69	0.240	2	10000	2	1	0
70	0.720	5	10000	3	1	1
71	0.511	2	10000	2	0	1
72	1.801	5	22500	2	3	1
73	0.902	5	50000	2	3	1
74	0.886	5	22500	3	1	1
75	0.588	2	22500	2	3	1
76	0.252	2	10000	1	3	1
77	0.524	2	3000	2	0	0
78	0.900	5	22500	2	3	1
79	0.237	2	22500	2	3	1
80	1.954	5	50000	2	3	1
81	0.526	2	10000	2	1	1
82	0.688	2	10000	2	3	1
83	0.412	5	10000	1	3	1
84	1.048	5	3000	1	1	1
85	0.518	5	10000	2	1	1
86	0.548	2	10000	1	1	1
87	0.738	2	10000	1	1	1
88	0.308	2	22500	2	1	1
89	1.416	2	50000	1	3	1
90	0.940	5	3000	1	1	1
91	0.189	2	10000	1	1	0
92	0.205	5	10000	1	0	1
93	0.422	2	22500	3	1	1
94	0.132	2	10000	1	1	0
95	0.213	2	3000	1	1	1
96	0.525	2	10000	1	1	1
97	0.235	5	10000	2	3	1
98	0.167	2	3000	1	0	0
99	0.383	2	10000	1	3	1
100	0.267	2	3000	1	0	1
101	0.681	2	10000	1	1	0
102	0.334	2	3000	1	1	0
103	0.147	5	10000	1	0	1
104	0.552	2	10000	1	0	1
105	0.181	2	3000	1	1	1
106	0.213	5	3000	1	1	1
107	1.689	8.5	10000	2	3	1
108	0.230	2	3000	1	1	1
109	0.295	2	10000	1	1	1
110	0.688	2	3000	1	3	1
111	0.165	2	3000	1	0	0
112	0.776	5	10000	1	1	1
113	0.811	5	10000	1	1	1
114	0.345	2	3000	1	0	1
115	0.340	2	10000	1	0	0
116	0.377	2	10000	1	1	0
117	0.703	5	3000	2	3	1
118	0.751	5	3000	1	1	1
119	0.634	5	10000	2	1	0
120	0.476	2	10000	2	3	1
121	1.279	5	50000	2	3	1
122	0.967	5	50000	2	3	1
123	0.436	5	10000	2	3	1
124	0.936	5	50000	2	3	1
125	1.792	5	50000	3	3	1
126	0.905	5	50000	2	3	1
127	0.520	2	22500	4	3	0
128	0.555	2	50000	3	3	1
129	0.644	2	50000	2	3	1
130	1.341	5	50000	3	3	1

Appendix C: Water use and climatological data

RW (m^3/m.)	RF (in.)	TE (°F)	CW (m^3/m.)	NWC (x1000)	IW (m^3/m.)	NWI (x1000)
627873.5	1.95	75.2	114507.5	6300	211487.5	18800
627873.5	3.13	73.4	114507.5	6200	211487.5	19600
490773	0.88	76.2	102634.5	6200	210914.5	20100
490773	1.99	77.5	102634.5	6200	210914.5	20100
552222	11.74	77.4	88113.5	6400	194671.5	19700
552222	10.95	80.2	88113.5	6000	194671.5	18700
562876	10.5	80.1	90403.5	6100	220645.5	19500
562876	6.96	80.8	90403.5	6100	220645.5	19600
590143.5	18.31	79.9	96063	6100	250060.5	19600
590143.5	7.1	80.1	96063	6300	250060.5	19500
573065.5	5.55	77.6	126345	6400	211215.5	19500
573065.5	1.68	77.5	126345	6700	211215.5	19400
551654	2.12	75.3	118970	6500	198813	19600
551654	5.27	75.9	118970	6200	198813	19900
515281.5	8.4	74.5	91127	6100	241540.5	19700
515281.5	6.6	75.4	91127	6100	241540.5	19500
540644	10.69	76.5	94154	6200	272085	19800
540644	3.67	80.2	94154	6200	272085	20100
553992.5	7.23	80.1	99931	6100	298360.5	19800
553992.5	13.49	79.9	99931	6100	298360.5	20200
552504	10.92	78.7	93684.5	6200	282683.5	20300
552504	17.22	78.3	93684.5	6200	282683.5	20500
545881	3.38	78	101325	6300	253987	20600
545881	0.34	78	101325	6500	253987	20900
553553	2.04	75.2	108600.5	6900	211801	19900
553553	2.22	74	108600.5	6700	211801	20400
497784	5.15	76.3	81511	6600	225233	20500
497784	5.64	78	81511	6700	225233	20400
538863.5	15	78.8	89405.5	6900	198013	20300
538863.5	8.03	79.9	89405.5	6900	198013	20400
603203.5	8.45	80.2	99589.5	6800	151873	20200
603203.5	11.97	81	99589.5	6800	151873	20000
518383.5	9.39	80.8	86367.5	6800	196867	20100
518383.5	15.77	79.8	86367.5	6900	196867	20100
507138.5	7.44	79	90580.5	7200	188994.5	20300
507138.5	1.2	77.7	90580.5	7600	188994.5	20400
518452.5	1.49	75.4	87899	7600	187291.5	20000
518452.5	2.39	76.9	87899	7400	187291.5	19800
629098.5	0.92	77.8	87244.5	7500	261621	19900
629098.5	7.58	80.5	87244.5	7600	261621	19100
491652	7.12	81.4	85779	7700	258958	19800
491652	8.3	80.8	85779	7700	258958	17500
518160	14.56	81.7	84904.5	7600	239553	19500
518160	8.07	82.4	84904.5	7500	239553	19000
566678.5	10.06	81.9	88385	7500	279056	20200
566678.5	14.07	79.9	88385	7600	279056	20500
509998.5	7.98	80.5	97337.5	7900	251592.5	20300
509998.5	1.97	79.6	97337.5	8300	251592.5	20600
533200	2.26	75.9	89640.5	7900	230791.5	20600
533200	2.25	75.4	89640.5	7800	230791.5	19700
582653.5	3.65	75.5	82172	7800	80178	21100
582653.5	6.78	78.8	82172	7800	80178	20900
479377	4.79	79.8	87464	7900	237693.5	20800
479377	7.83	80.6	87464	7900	237693.5	18700
569806.5	7.08	80.3	73754	7700	81560.5	20200
569806.5	14.61	78.4	73754	7700	81560.5	20100
507632	10.27	81	77717.5	7800	79655	19100
507632	6.99	78.9	77717.5	7700	79655	20800
445307.5	7.44	77.6	81681	8000	78631.5	20800
445307.5	1.1	75.8	81681	8200	78631.5	21100
461076	2.14	75.3	84018.5	8200	118462.5	20900
461076	2.38	75.3	84018.5	8000	118462.5	21100

Appendix C: Water use and climatological data

RW (m^3/m.)	RF (in.)	TE (°F)	CW (m^3/m.)	NWC (x1000)	IW (m^3/m.)	NWI (x1000)
458512.5	10.29	74.7	90742	8000	200213.5	21600
458512.5	5.3	75.8	90742	8000	200213.5	21600
494888.5	8.26	78.5	82995	8400	122817.5	21500
494888.5	8.77	78.3	82995	8200	122817.5	21300
557493	16.03	81.44	98398	8200	179854	21300
557493	8.83	81.5	98398	8200	179854	21000
537500	9.37	81.7	102320.5	8200	198392	20100
537500	7.95	80.8	102320.5	8300	198392	20300
556367.5	3.99	79.5	97204.5	8400	185568	20600
556367.5	1.27	76.9	97204.5	8900	185568	20600
509574.5	0.54	75.7	92741.5	8400	225292.5	17000
509574.5	2.78	75	92741.5	8300	225292.5	20300
512665	5.01	74.7	91945.5	8300	232814.5	19600
512665	1.83	78.5	91945.5	8300	232814.5	19700
530194	5.11	80.7	93012.5	8300	275508.5	19700
530194	8.97	82.2	93012.5	8300	275508.5	19900
535126	7.59	82	96642	8200	229892	19700
535126	11.1	82.7	96642	8400	229892	19300
534093.5	14.93	81.8	95661	8400	152307	19300
534093.5	14.42	81.7	95661	8500	152307	19100
530833	3.52	79.5	100894	8700	191412	18500
530833	0.91	76.6	100894	9200	191412	18500
529325.5	0.66	75.7	107662.5	9400	205040	18300
529325.5	2.16	76.3	107662.5	9100	205040	18500
543630.5	5.69	75	88835.5	9100	189364.5	17400
543630.5	3.38	78.1	88835.5	9000	189364.5	18000
614746	9.76	81.1	111585	9100	196517	18400
614746	3.04	81.7	111585	9300	196517	18300
527822.5	7.78	82.1	108924	9100	244879	18200
527822.5	4.98	82.1	108924	8900	244879	17500
595330	9.09	82.5	110826.5	9000	280112.5	17400
595330	2.25	80.7	110826.5	9200	280112.5	16800
552272.5	2.81	79.3	107536.5	9600	267117	17800
552272.5	0.51	75.4	107536.5	10200	267117	18000

Appendix D: Plots of data used

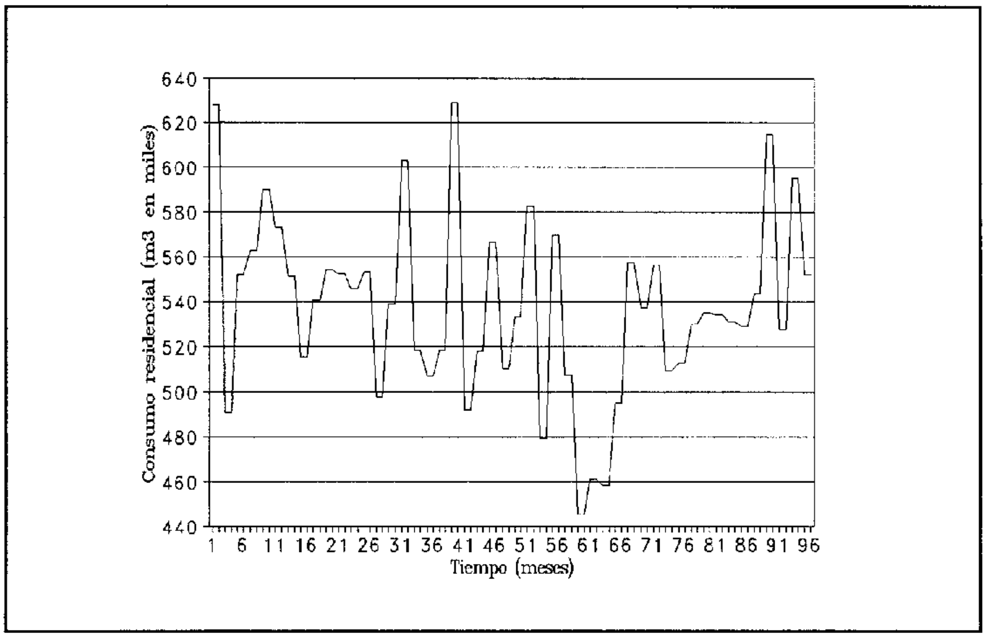

Figure D1. Residential water use (Jan/1984-Dec/1991).

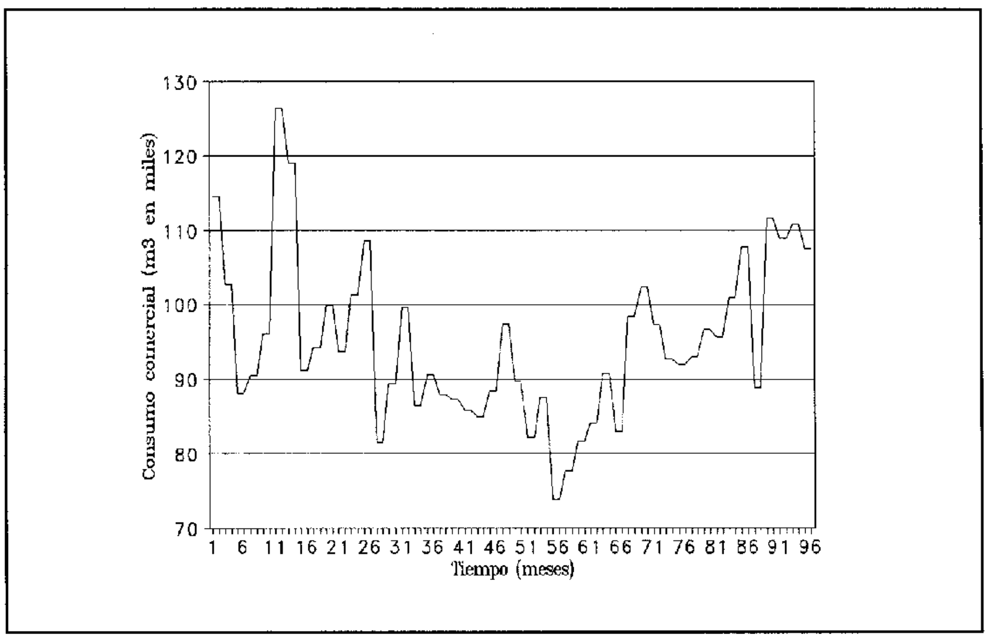

Figure D2. Commercial water use (Jan/1984-Dec/1991).

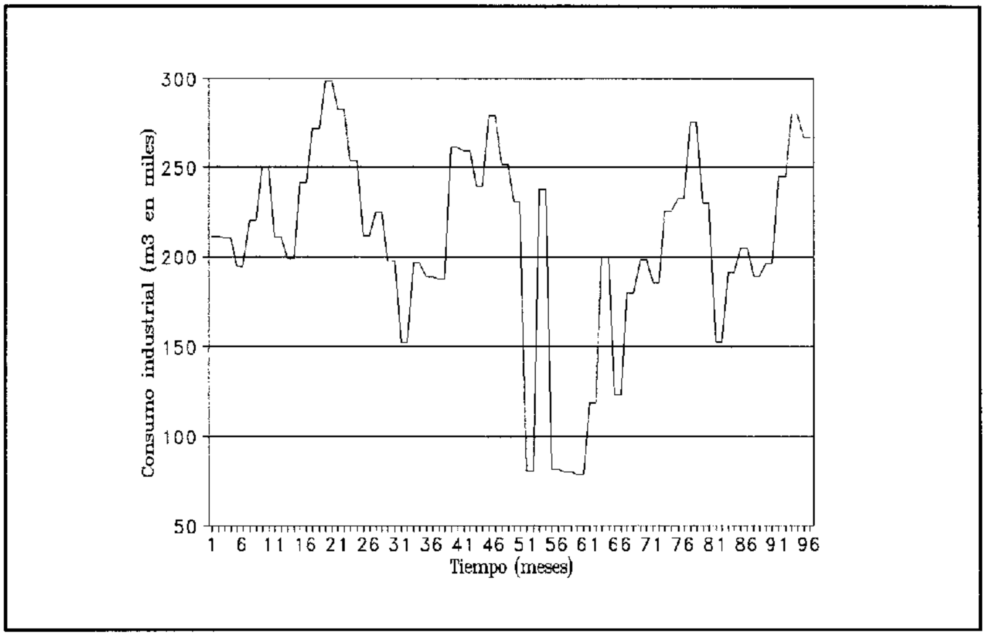

Figure D3. Industrial water use (Jan/1984-Dec/1991).

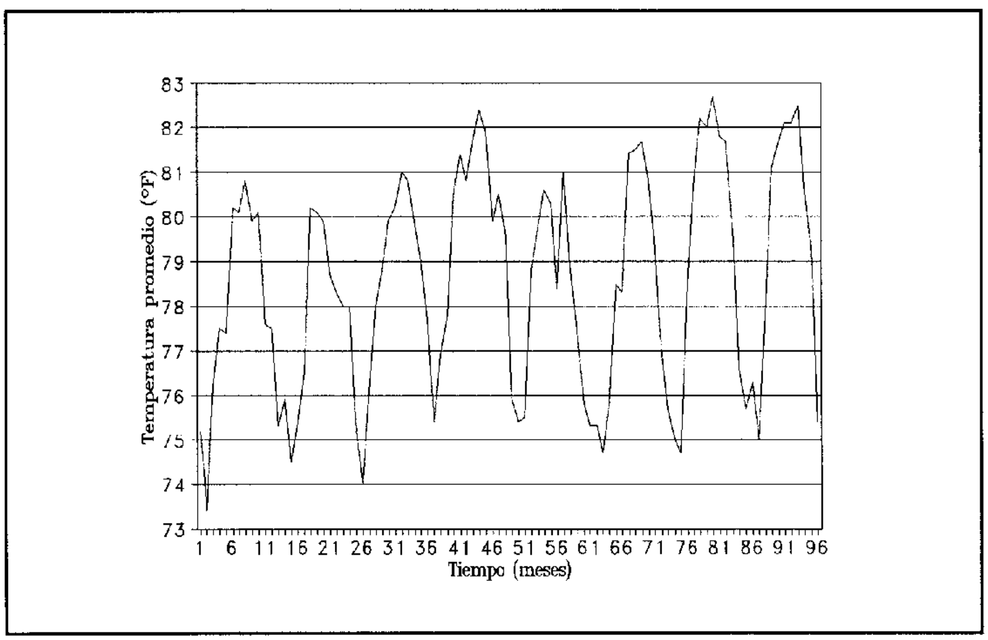

Figure D4. Mayaguez average temperature (Jan/1984-Dec/1991).

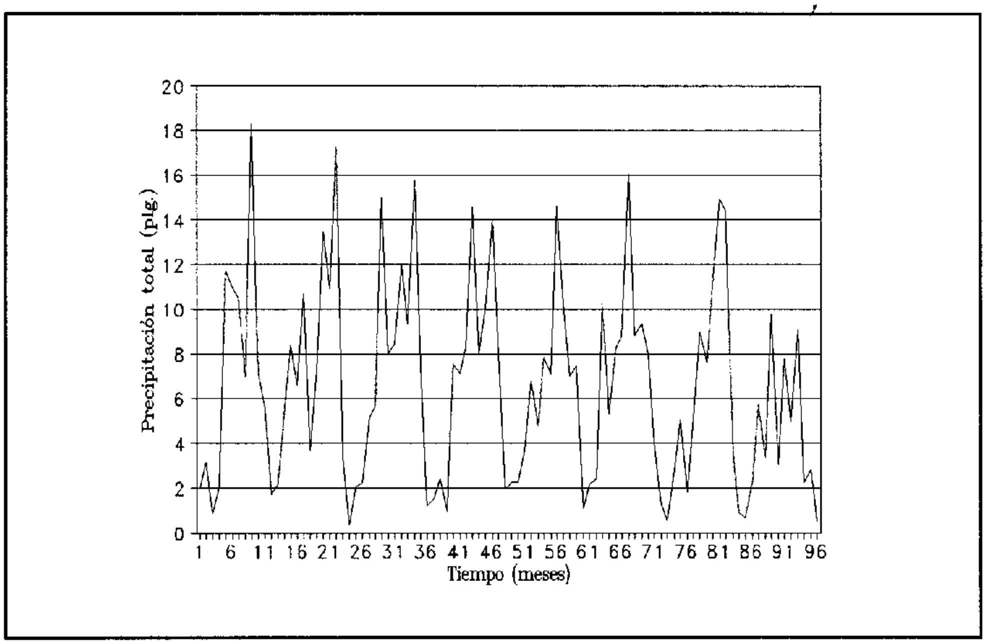

Figure D5. Mayaguez cumulative rainfall (Jan/1984-Dec/1991).

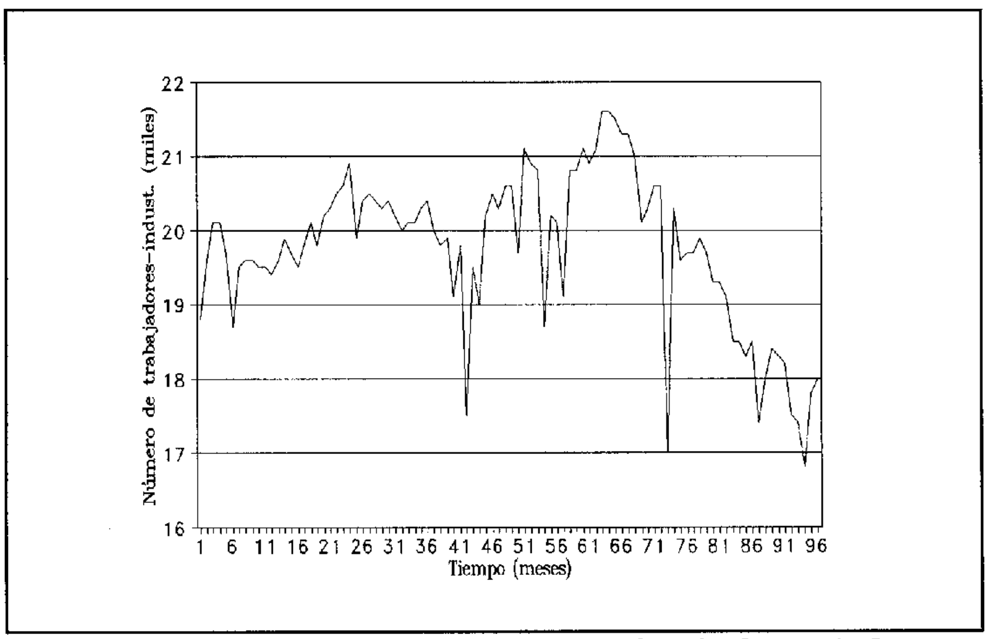

Figure D6. Number of workers at the industrial area (Jan/1984-Dec/1991).

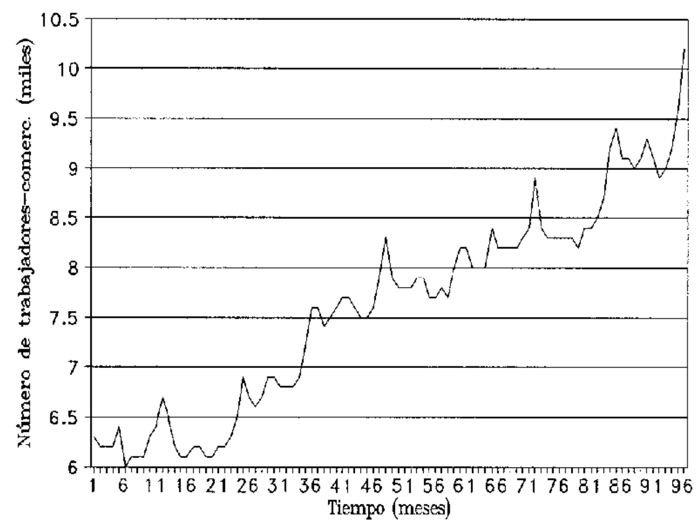

Figure D7. Number of workers at the commercial area (Jan/1984-Dec/1991).