JKAU: Eng. Sci., vol. 8, pp. 61-71 (1416 A.H./1996 A.D.)

Modeling Municipal Water Demand Using Box-Jenkins Technique

KHALED H. AL-DHOWALIA
Civil Engineering Department, College of Engineering, King Saud University, Riyadh, Saudi Arabia

ABSTRACT. The Box-Jenkins time series technique has been used to establish a monthly average water demand model for Riyadh city, the capital of Saudi Arabia. The model is based on the actual monthly average water demand observed from January 1978 to December 1987. In an ex-post forecast the model accurately predicts monthly variations in Municipal Water demand. For medium term forecasts the developed model is expected to give more accurate future water demand forecasts than the trend extrapolation method adopted by the Riyadh Water Department.

1. Introduction

Careful planning for future municipal water demand is very important for cities with limited and expensive water sources. The traditional demand forecasting has been trend extrapolation but lately there has been increasing concern about the risks of over-provision associated with this approach and other alternative methods have been investigated.

Many attempts have been made to develop models that explain the large variability in water use by expressing demand as a function of various parameters affecting use. Major parameters include the price of water, average temperature, number of connections, income and water use. Multiple regression and time series models were use by some researchers for municipal water forecasts as shown in studies by Wong[1], Agthe[2], Hughes[3], Hansen and Narayanan[4], Power *et al.*[5], Hanke *et al.*[6], Archibald[7], Maidment *et al.*[8], Cochran *et al.*[9], Kher[10], Metzner[11] and Weber[12]. Al-Mohawas and El-Razaz[13], El-Razaz, *et al.*[14], and El-Razaz and Mazi[15] used Box-Jenkins time series technique successfully to forecast power demand for fast developing power systems.

The daily water consumption for Riyadh, the capital of Saudi Arabia, exceeded one million meter cube in the summer of 1989. The average water demand rate grew by more than 15 percent per year in the last few years and this growth is expected to persist for sometime due to high population growth, fast industrial and commercial developments, and increase in per capita water consumption. Thus the water demand forecasting for Riyadh is of critical importance due to the nature of limited and costly water sources supplying the city.

This paper discusses the method of establishing a model based on Box-Jenkins time series technique suitable for the monthly average water demand for Riyadh city. Medium term forecasts for the monthly average are obtained and compared with the actual average demand experienced in 1988 and 1989.

2. Method and Results

2.1 Time Series Data

Riyadh city is located in the center of Saudi Arabia. The climate of Riyadh is arid and of a typical desert nature with extreme temperature differences, both between summer and winter and from day to night. January is the coldest month with an average temperature of about + 14°C and a minimum of – 5°C. July and August are the hottest months with an average temperature of + 35°C and a maximum of + 50°C. Rainfall is highly variable from year to year with an average annual rainfall of less than 100 mm. Rainy season extends from November to May. The relative humidity is low in summer with an average of about 15% and moderate in winter with a range from 35% to 45%. The Riyadh Region Water Department (RRWD) serves about 1.5 million city residents. The city is supplied by two major sources from desalination plants which are located at the eastern province 500 kilometers from Riyadh and this represents about 70% of the total water supplied. The balance is supplied from five water treatment plants, treating ground water, located within the city limits. Although RRWD uses block structure for costing, the customers pay only a marginal cost which represents only about 10 percent of the actual cost and the rest is subsidized by the government. Because of this, the effect of some important parameters such as water price and income on water demand will be minimal. So in the absence of some important data and parameters a different approach for modeling Riyadh water demand must be investigated.

The time series models were fitted to data for the years 1978 to 1987 and data from 1988 to 1989 are used in an ex-post forecast. Average monthly water demand for Riyadh (m^3 day) for the period of January 1978 to December 1987 is shown in Fig. 1.

2.2 Model Formulation

Delong *et al.*[16] produced a rational structured approach to modeling and forecasting stationary time series. For the short and medium terms, Box-Jenkins technique leads to better forecasts than other statistical forecasting methods[16]. The general Box-Jenkins model is an autoregressive-integrated moving average process abbreviated as ARIMA $(p, d, q)(P, D, Q)_s$, where p, d, q are nonseasonal au-

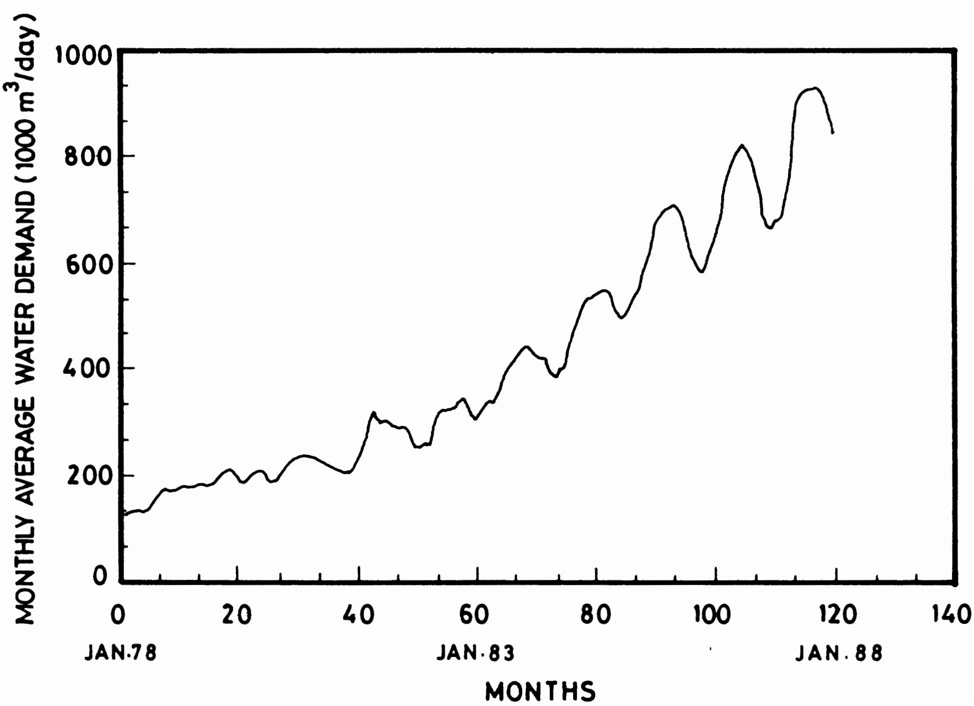


FIG. 1. Monthly average municipal water demand for Riyadh.

toregressive order, number of regular differences, non-seasonal moving-average order respectively and P, D, Q, s are seasonal autoregressive order, number of seasonal differences, seasonal moving-average order and order of seasonal differences respectively. This model is described by the following general equation

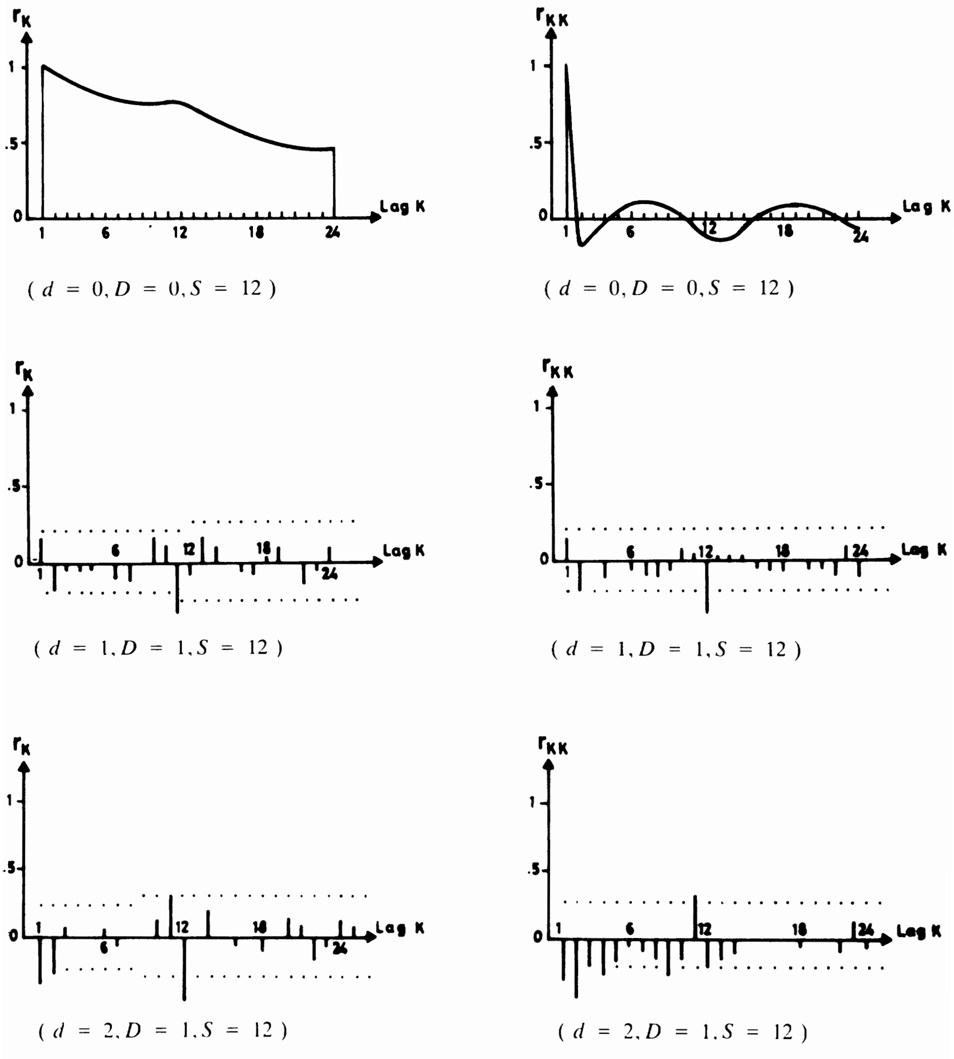
$$(1 - \phi_1 B - \cdots - \phi_p B^P) (1 - \Gamma_1 B^s - \cdots - \Gamma_p B^{sp}) (1 - B^s)^D (1 - B)^{dY_t}$$
$$= \delta_0 + (1 - \theta_1 B - \cdots - \theta_q Bq) (1 - \Delta_1 B^s - \cdots - \Delta_Q B^{SQ}) \varepsilon_t$$

where ϕ_i, Γ_i, θ_i and Δ_i are regular autoregressive (AR), seasonal (AR), regular moving average (MA) and seasonal (MA) parameters respectively. Also B, Y_t, δ_0 and ε_t are back shift operator, time series variable, deterministic trend component and random variable respectively.

The procedure consits of three main stages, namely identification, estimation and diagnostic checking, and forecasting.

2.3 Identification

The first step in developing a model is to examine the sample autocorrelation (AC) and sample partial autocorrelation (PAC) functions. The autocorrelation and partial autocorrelation functions were computed and plotted using SAS/ETS[17]. Sample autocorrelation for the data sequence ($S = 12$) with no regular or seasonal differences up to 24 lag have been obtained as shown in Fig. 2. This sample autocorrelation is oscillatory and a nondecaying function indicating a non-stationary process. In order to obtain a stationary process a number of differencing schemes were tested. Stationarity appears to have been achived for the cases ($d = 1, D = 1, S = 12$) and ($d = 2, D = 1, S = 12$). The AC and PAC functions of these two cases are also shown in Fig. 2. After examining the AC and PAC functions six ARIMA models have been identified and are given in Table 1. The formulations of the identified models are shown in Table 2.



Autocorrelation functions

Partial autocorrelation functions.

FIG. 2. Autocorrelation and partial autocorrelation functions for several differences ($S = 12$).

TABLE 1. Identified water demand models.

Model no.	p	d	q	P	D	Q
1	1	1	1	1	1	1
2	0	1	1	1	1	1
3	0	1	1	0	1	1
4	1	2	1	1	1	1
5	0	2	2	0	1	2
6	0	2	1	0	1	1

TABLE 2. Formulations of water demand models.

Model no.	Model
1	$(1-B^{12})(1-B)(1-\phi_1 B)(1-\Gamma_1 B^{12})\,Y_t = \delta_0 + (1-\theta_1 B)(1-\Delta_1 B^{12})\,\varepsilon_t$
2	$(1-B^{12})(1-B)(1-\Gamma_1 B^{12})\,Y_t = \delta_0 + (1-\theta_1 B)(1-\Delta_1 B^{12})\,\varepsilon_t$
3	$(1-B^{12})(1-B)\,Y_t = \delta_0 + (1-\theta_1 B)(1-\Delta_1 B^{12})\,\varepsilon_t$
4	$(1-B^{12})(1-B)^2(1-\phi_1 B)(1-\Gamma_1 B^{12})\,Y_t = \delta_0 + (1-\theta_1 B)(1-\Delta_1 B^{12})\,\varepsilon_t$
5	$(1-B^{12})(1-B)^2\,Y_t = \delta_0 + (1-\theta_1 B-\theta_2 B^2)(1-\Delta_1 B^{12}-\Delta_2 B^{24})\,\varepsilon_t$
6	$(1-B^{12})(1-B)^2\,Y_t = \delta_0 + (1-\theta_1 B)(1-\Delta_1 B^{12})\,\varepsilon_t$

2.4 Estimating and Diagnostic Checking

In order to choose among the several candidate models, it was necessary to estimate the parameters of each of them and examine their properties. The estimates of the unknown parameters in the models 1 throughout 6 were calculated using the least squares method[17] and are shown in Table 3 together with results of the diagnostic checks.

Three diagnostic criteria were used to test the adequacy of the tentative models namely :

1) The t-test to measure the stability and importance of the parameters.

2) The chi-square statistics CQ, to measure the autocorrelation of the residuals. For good fit, residuals have to be randomly distributed and therefore, not correlated. CQ value is calculated using the formula[16,18].

$$CQ = (n-d)\sum_{\ell=1}^{K} r_\ell^2(\varepsilon)$$

where n is the number of observations and $r_\ell(\varepsilon)$ is the sample autocorrelation of the residuals at lag ℓ.

TABLE 3. Parameters estimates and their corresponding test statistics.

Model	Parameters estimates	t-test statistics	Degree of freedom	CQ statistics	SE
1	δ = 1570.92		19	22.63	14716
	θ = 0.5042	1.71		(30.1)*	
	Δ = 0.1627	0.58			
	ϕ = − 0.2457	0.74			
	Γ = − 0.3009	1.11			
2	δ = 1265		20	22.63	14681
	θ = − 0.2801	2.89		(31.4)*	
	Δ = 0.1836	0.68			
	ϕ = − 0.2987	- 1.14			
3	δ = 991.3		21	24.59	14699
	θ = − 0.2995	3.11		(32.7)*	
	Δ = 0.4359	4.19			
4	δ = 92.03		19	20.54	15312
	θ = 0.9846	34.87		(30.1)*	
	Δ = 0.0719	0.25			
	ϕ = 0.2605	2.48			
	Γ = − 0.3719	1.37			
5	δ = 87.96		20	21.03	15259
	θ_1 = 0.6646	6.79		(31.4)*	
	θ_2 = 0.308	3.16			
	Δ_1 = 0.3928	3.6			
	Δ_2 = 0.0856				
6	δ = 103.56		21	26.71	18617
	θ = 0.4796	2.54		(32.7)*	
	Δ = − 0.1486	0.72			

*Critical values of chi-square at 5%.

3) The standard error SE defined as[18].

$$SE = \sqrt{\frac{\sum_{t=1}^{n} (Y_t - \hat{Y}_t)^2}{n - m}}$$

where: Y_t is the actual observation at time t.
$\hat{Y}_t$ is the estimated value from the model.
m number of parameters in the model.

Comparing the chi-square statistics applied to the first 24 autocorrelations with the critical values of chi-square at 5 percent shows that all mix models satisfy the chi-square test which indicates the adequacy of the models. Although model 4 has the lowest chi-square value, the standard error value (SE) is higher than the first three models. Based on standard error values models 1, 2 and 3 are superior. Since all the candidate models are acceptable it is better to compare the forecasts of these models in order to choose the appropriate model.

2.5 Forecasting

Models 1 through 6 were used to generate monthly average water demand for Riyadh city for two years starting January 1988[17]. Table 4 gives forecasts error of the six Box-Jenkins seasonal models for the year 1988. Models 1, 2 and 3 gave similar and more accurate forecasts than models 4, 5 and 6.

TABLE 4. Forecasts percentage error for 1988.

Month	$\% \text{ Error} = \frac{\lvert \text{Forecast} - \text{Actual} \rvert}{\text{Actual}} \times 100$					
	Model no.					
	1	2	3	4	5	6
Jan.	1	1.1	1.5	1.5	1.9	2
Feb.	2.8	3	3.5	4	4.5	4.7
Mar.	2.84	3	3	4.7	4.6	5.5
Apr.	2.34	2.5	2.8	4.7	4.9	5.5
May	5.16	5.2	5.5	8.1	8.4	7.5
June	5	5	5.1	8.4	8.4	7.6
July	7.4	7.5	7.2	11.6	11.2	10.5
Aug.	9	9.1	8.7	13.6	13.2	13
Sept.	9.3	9.4	9.3	14.5	14.4	14.4
Oct.	9.3	9.4	9.5	15.1	15.1	15.5
Nov.	12.9	13.1	13.6	19.9	20.4	21
Dec.	14.5	14.6	15.5	22.5	23.4	24.2
Average	6.8	6.9	7.1	10.7	10.9	11

Since the differences between the monthly forecasts of the three models 1, 2 and 3 and their averages are relatively small, the choice between these three models is not critical. Model 3 was chosen because it is less complicated. Therefore, one may conclude that the monthly average water demand for Riyadh can be modeled as

$$(1 - B^{12}) (1 - B) Y_t = 991.3 + (1 + 0.2995 B) (1 - 0.4359 B^{12}) \varepsilon_t$$

The forecasts of the chosen model (3), and their lower and upper 95% confidence limits are shown in Table 5 and plotted in Fig. 3. Also Table 5 compares the actual monthly average water demand in 1988 with the forecasts and the lower confidence limits. The model resulted in overestimated forecasts since the Box-Jenkins technique assumed that pattern of the time series will persist. The water demand for Riyadh is expected to continue to grow with a decreasing rate. However, the lower 95% confidence limits is less than the actual demand. Therefore one may conclude that the forecast of the next two years for Riyadh city will be the average between the forecasts and their lower 95% confidence limits.

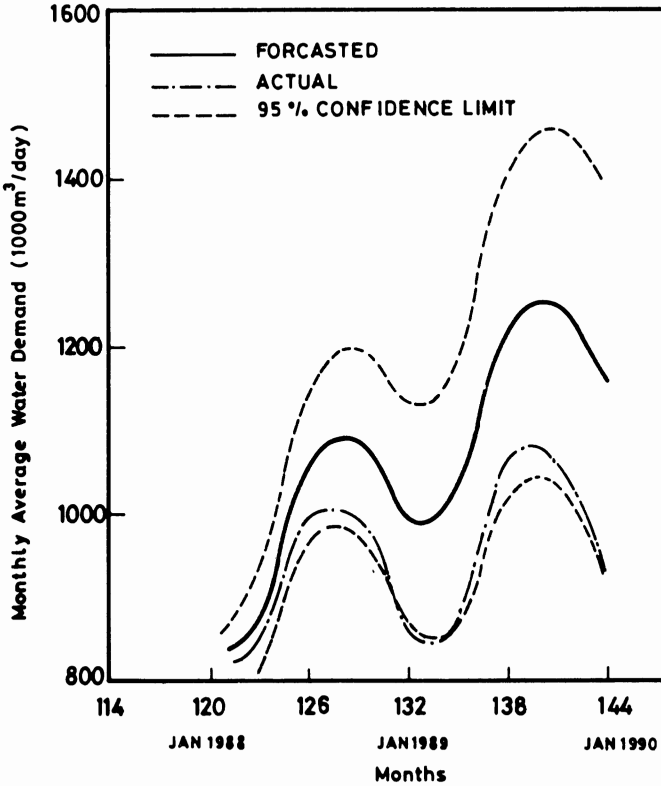


FIG. 3. Actual and forecasted water demand.

TABLE 5. Comparison of the actual and forecasts of 1988.

Month	Actual (A)	L 95*	Forecasts (F)	U 95**	$\frac{L\ 95 - A}{A} \times 100$	$\frac{F - A}{A} \times 100$
Jan.	820211	803515	832320	861126	- 2	1.5
Feb.	820537	801791	849024	896257	- 2.3	3.5
Mar.	841623	806897	867164	927432	- 4.1	3.0
Apr.	890658	844661	915607	985553	- 5.2	2.8
May	952308	924435	1004551	1084867	- 2.9	5.5
Jun.	999803	961855	1050385	1138905	- 3.8	5.1
July	1004947	981693	1077782	1173892	- 2.3	9.0
Aug.	1000832	984979	1088120	1191262	- 1.6	8.7
Sep.	994702	978105	1087829	1197553	- 1.7	9.4
Oct.	977898	954627	1070561	1186494	- 2.4	9.5
Nov.	908175	909791	1031518	1153445	+ 0.2	13.6
Dec.	863113	869683	997131	1124579	+ 0.8	15.5

* Lower 95% confidence limit.
**Upper 95% confidence limit.

3. Conclusion

The forecasting methodology presented in this paper should be useful for predicting medium term municipal water demand. The univariate time series models have been applied successfully to forecast the average monthly water demand for fast developing water system of Riyadh city. However one should be careful in applying these models for long term water demand forecasts because it tends to overestimate the demand by large margin. The forecasts obtained were compared with those experienced in 1988 and 1989. Although the model resulted in over-estimated forecasts the lower 95% confidence limit is less than the actual demand. Decision makers may select the appropriate forecasts among those presented in Table 5. However, one may conclude that the average between the forecasts and the lower 95% confidence limit is an appropriate one.

The ARIMA procedure is very useful in developing water forecasting models because it is simple and requires only water demand historical data. However it would not be possible to evaluate the effect of management options on these models.

Due to the nature of the critical water situation in Riyadh City accurate forecasting is a must. The developed model will give more accurate future water demand forecasts than the trend extrapolation method adopted by RRWD.

Acknowledgement

Special thanks are given to the Riyadh Water Department for providing me with all the water demand data necessary for this work. Thanks are also extended to Engineer Mohammad Mansoure for the assistance in the computer work.

References

[1] **Wong, S.T.**, A model on municipal water demand: A case study of Northeastern Illinois, *Land Economics*, **48**(1): 34-44 (1972).

[2] **Agthe, D.E.** and **Bilings, R.B.**, Dynamic models of residential water demand, *Water Resources Research*, **16**(3) June: 476-480 (1980).

[3] **Hughes, T.C.**, Peak period design for small western U.S. water supply systems, *Water Resources Bulletin*, **16**(4) August: 661-667 (1980).

[4] **Hansen, R.D.** and **Narayanan, R.**, A monthly time series model of municipal water demand, *Water Resource Bulletin*, **17**(4) August: 578-585 (1981).

[5] **Power, A.N.** *et al.*, Deterministic models for predicting residential water consumption, *Water Resource Bulletin*, **17**(6) December: 1042-1044 (1981).

[6] **Hanke, S.H.** and **de Mare, L.**, Residential water demand: A pooled, time series, cross section study of MALMO, Sweden, *Water Resources Bulletin*, **16**(4) August: 621-625 (1982).

[7] **Archibald, G.G.**, Forecasting water demand: A disaggregated approach, *Journal of Forecasting*, **2:** 181-192 (1983).

[8] **Mailment, D.R.** and **Parzen, E.**, Cascade model of monthly municipal water use, *Water Resources Research*, **20**(1) January: 15-23 (1984).

[9] **Cochran, R.** and **Cotton, A.W.**, Municipal water demand study Oklahoma city and Tulsa, Oklahoma, *Water Resources Research*, **21**(7) July: 941-943 (1985).

[10] **Kher, L.K.**, Identification of water demand models from noisy data, *Water Resource Research*, **22**(3) March: 322-330 (1986).

[11] **Metzner, R.C.**, Demand forecasting: A model for San Francisco, *J. AWWA*, **15**(February): 56-59 (1989).

[12] **Weber, J.A.,** Forecasting demand and measuring price elasticity, *J. AWWA,* **18**(May): 57-58 (1989).
[13] **Al-Mohawes, N.** and **El-Razaz, Z.,** Modeling monthly peak load in a fast growing city using Box-Jenkins technique, *Journal of Engineering Science, King Saud University,* **13**(2): 245-254 (1987).
[14] **El-Razaz, Z.** *et al.*, Load duration curve forecasting using time series approach, *Electric Machines and Power Systems,* **15**: 397-410 (1988).
[15] **El-Razaz, Z.S.** and **Mazi, A.A.,** Unified weekly peak load forecasting for fast growing power system. *IEEE Proceedings,* **136,** pt. c, (1) January: 397-410 (1989).
[16] **Delony, P.M., Box, G.** and **Jenkins, G.,** *Time series analysis: Forecasting and control,* Holden Day Inc., London, 189 p. (1976).
[17] **SAS Institute Inc.,** *SAS/ETS User's Guide Econometric and Time-Series Library,* Cary (1980).
[18] **Bowerman, B.L.** and **O'Connell, R.T.,** *Forecasting and Time Series,* Duxbury Press, North Massachusetts, 481 p. (1979).

Notation

B = back shift operator.
D = number of seasonal differences.
d = number of regular differences.
K = lag time unit.
m = number of parameters in the model.
n = number of observations.
P = seasonal autogressive order.
p = nonseasonal autogressive order.
Q = seasonal moving-average order.
CQ = chi-square statistic.
q = nonseasonal moving-average order.
$r_\ell(\varepsilon)$ = the sample autocorrelation of the residuals at lag ℓ.
SE = standard error.
s = order of seasonal differences.
Y_t = the average monthly water demand (m^3 / day).
Δ_i = seasonal moving-average parameter.
δ_o = deterministic trend component.
θ_i = regular moving average parameter.
ϕ_ι = regular autogressive parameter.
Γ_t = seasonal autogressive parameter.
ε_t = random variable.

نمذجـــة طلب الميـــاه بالبلـــديات باستخـــدام طريقـــة بوكس - جنكـنز

خالد حمد الضويلع
قسم الهندسة المدنية ، كلية الهندسة ، جامعة الملك سعود
الريـــــاض - المملكة العربية السعودية

المستخلص . استخـدمت طريقـة السـلاسل الزمنية لبوكس - جنكنز لتطوير نموذج رياضي لتمثيل الـطلب الشهـري للمياه لمدينـة الـرياض ، عاصمـة المملكـة العـربية السعودية . النموذج مبني على أساس معدل الطلب الشهري الحقيقي المسجل خلال الفترة من يناير **١٩٧٨** إلى ديسمبر **١٩٨٧**م . باستخدام هذا النموذج أمكن التنبؤ بدقة بالتغير الشهري لطلب المياه واستنباط الطلب المستقبلي للمياه للمدى المتوسط بدقة أكبر من طريقة التنبؤ النمطي المتعارف عليها ، والمستخدمة من قبل مصلحة المياه بمدينة الرياض .