Forecasting Hourly Water Demands by Pattern Recognition Approach

By Leonid Shvartser,[1] Uri Shamir,[2] Member ASCE, and Mordechai Feldman[3]

Abstract: Hourly water-demand data is forecasted with a model based on a combination of pattern recognition and time-series analysis. Three repeating segments are observed in the daily demand pattern: "rising," "oscillating," "falling," then "rising" again the following day. These are called "states" of the demand curve, and are defined as successive states of a Markov process. The transition probabilities between states are "learned," and low-order auto-regressive integrated moving average (ARIMA) models fitted to each segment, using a modest amount of historical data. The model is then used to forecast hourly demands for a period of one to several days ahead. The forecast can be performed in real time, on a personal computer, with low computational requirements, at any time the system state deviates from the planned, or when new data become available. The process of model development, application, and evaluation is demonstrated on a water system in Israel.

Introduction

The task for on-line control of a water distribution system is to prepare and execute a plan for operating the system. The objective of the plan is to meet the demands at the lowest cost, and it is therefore referred to as an optimal operation plan.

Typically, the operating plan is prepared for a period of 24 h ahead. There are several reasons for this: (1) The demands display a pronounced daily cycle; (2) energy tariffs are based on time of day; and (3) there is frequently a time of the day at which a boundary condition on the state of the system (reservoir levels) can be set in advance with reasonable certainty. While the normal planning period is therefore 24 h, it is sometimes necessary to look further ahead, for example to the end of the current week when larger storage reservoirs have a weekly cycle. There may be cases when the forecast is needed only to the end of the current day.

The necessary condition for preparation of an operating plan is a forecast of water demands for the entire planning period. The temporal definition of the forecast is determined by the needs of the operational planning process. Here, we deal with a continuous trace over time, from which values for an hour or even parts thereof may be taken. We shall call this temporal definition "hourly demands," for convenience and also because an hour is a common time step for planning the operation. The spatial definition of the forecast, i.e., for which points in the network we need it, depends on the mathematical model of the network used. In this paper we do not deal explicitly with this aspect. Suffice it to say that the demand we are forecasting is for a population that is large enough to smooth out the randomness of

---

[1]Command and Control Unit, Mekorot Water Co., 9 Lincoln St., Tel Aviv, Israel.
[2]Prof., Fac. Civ. of Engrg., Technion, Haifa 32000, Israel; also Consultant to Mekorot Water Co., Israel.
[3]Command and Control Unit, Mekorot Water Co., 9 Lincoln St., Tel Aviv, Israel.
Note. Discussion open until May 1, 1994. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on September 28, 1992. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 119, No. 6, November/December, 1993. ©ASCE, ISSN 0733-9496/93/0006-0611/ $1.00 + $.15 per page. Paper No. 4869.

one individual's behavior, yet small enough to have a common pattern not smeared out by the combination of very different demand patterns.

In most water-distribution systems there is storage, which is operated on a daily cycle to reduce energy charges, by shifting pumping away from times of high energy tariffs. This operation of the storage smooths the instantaneous peak demands; and, from the point of view of optimal pumping schedules, reduces the importance of short-duration peaks. Therefore, we adopt as criteria for evaluation of the demand forecasting method, first the total daily volume and then its temporal distribution.

We have found few reports of forecasting water demands on an hourly basis and some on daily forecasts. Moss (1978) and Gray (1978) used time series analysis methods to develop a forecast of demands over the day. Maidment and Panzer (1984), and Maidment et al. (1985) developed transfer functions that describe the influence of temperature and rainfall on daily water consumption. This method seems to require large amounts of data; 20 years were used by Maidment et al. Sterling and Bagrelia (1985) used a two-stage auto-regressive integrated moving average (ARIMA) procedure: first an ARIMA with a seven-day integration period to determine the daily trend, then another ARIMA model with a 24-h integration period on the data with the daily trend removed, to determine the deviation from the daily mean. Hartley and Powell (1991) continue this work in an interesting way. They combine this method with a heuristic approach: The mathematical algorithm provides the base forecast that is augmented by a knowledge base containing information pertaining to any abnormal events likely to affect demand over the prediction period. Jowitt and Xu (1992) analyzed the influence of meteorological data on water demand, and concluded that they cannot find a specific correlation. They therefore used the time-series approach suggested by Sterling and Bagrelia (1985).

At Mekorot, Israel's national water-supply company, work has been conducted in recent years on methods for on-line operation of water supply systems. The analysis is conducted by simulation and optimization, which requires demand forecasting. Initially, we used a fixed daily demand pattern, and forecasted only its "scale," i.e., the average daily value. The work reported in this paper is the second step, which uses methods of pattern recognition and time-series models. We are presently adding consideration of explanatory variables—such as temperature, day of week, days since last rainfall—using cluster analysis methods, and embedding the whole package in an expert system. This work will be reported in a future paper.

BASIC APPROACH

The variation of water demands over the day depends on many parameters—temperature, humidity, time since last rainfall (for irrigation and lawn watering), day of the week, etc. Yet, we detect a markedly similar pattern over the day, one which shows a typical "signature" of each consumer type. Fig. 1 is from a study of Boston water system (*Boston Water* 1966), with the data given as dimensionless demand curves (DDCs) normalized by the mean daily value. Fig. 2 shows data from the Sorek system in Israel, which supplies a combination of agricultural and domestic demands.

In all cases we can detect a similar pattern: a low demand at night, rising during the morning, remaining high but with some variability for part of the day, then decreasing. The transitions between these parts, as well as the shape of each part, may vary among consumers, yet we can generalize

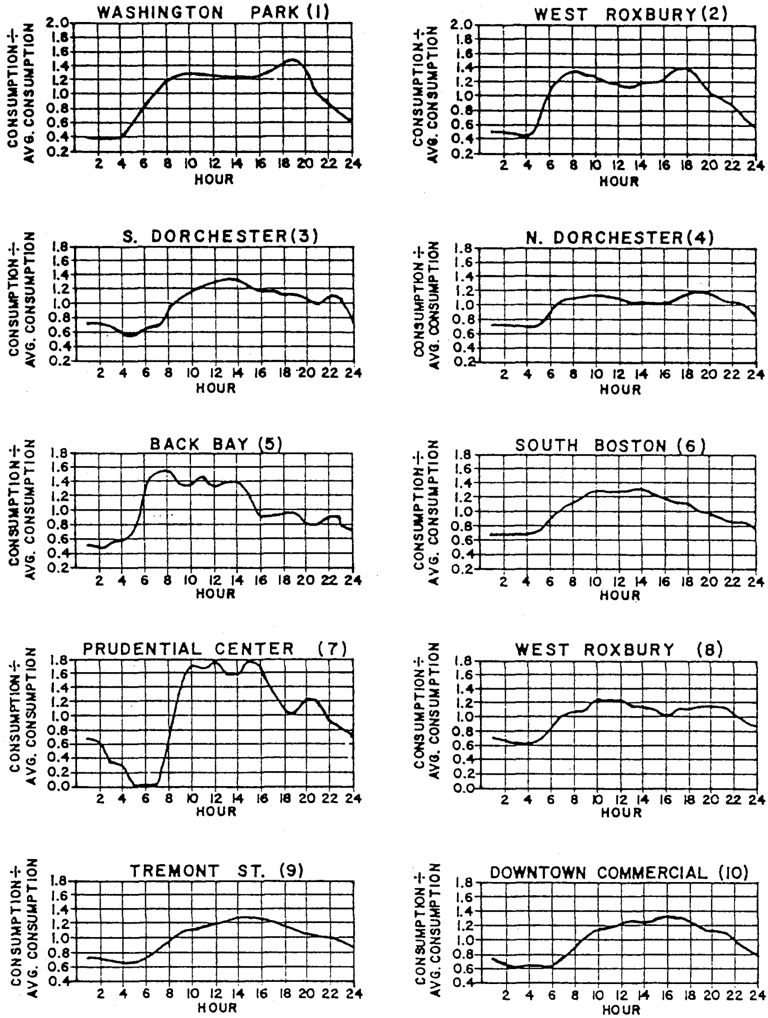

FIG. 1. Dimensionless Demand Curves (*Boston Water Study* 1966)

that the daily pattern is made of a "rising" segment, an "oscillating" segment, and a "falling" segment, as depicted schematically in Fig. 3. Our method is based on assuming this general pattern, then identifying the points of transition between segments and constructing a time-series model for each segment.

This paper deals with a statistical model, which is based on the assumption of stable meteorological and other environmental parameters. In the next phase of our work we are going to take into consideration the effects of changing conditions.

We follow the approach proposed by Mottl' et al. (1983), which is based on principles of pattern recognition. The daily demand pattern is assumed to be a stochastic process with segments, where the transition between them

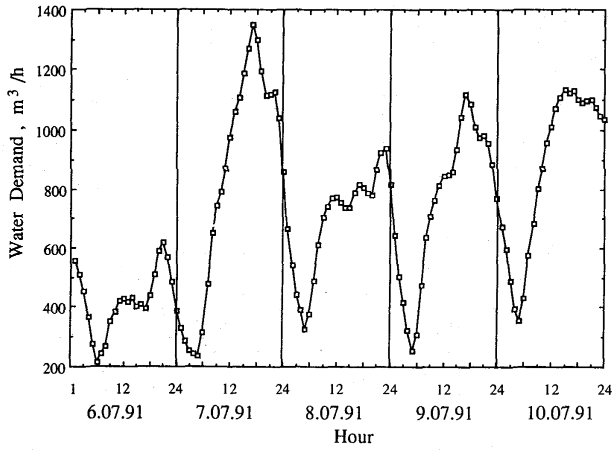

FIG. 2. Water Demand: Sorek Water Supply District, Israel

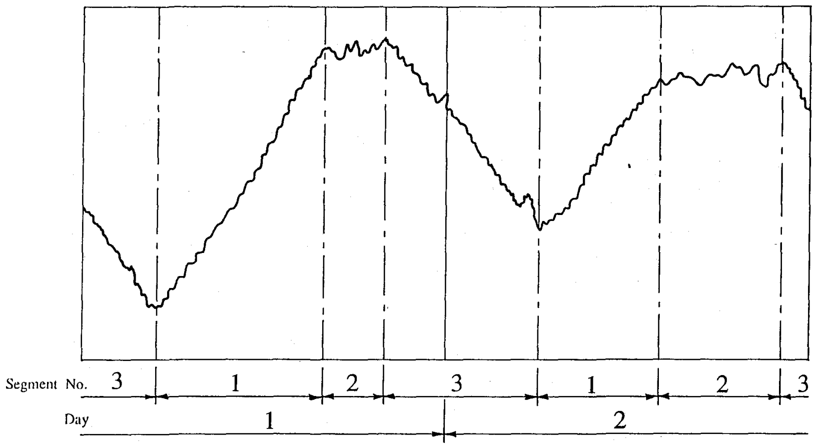

FIG. 3. Schematic Description of Three-Segment Process: "Rising" (1), "Oscillation" (2), "Falling" (3)

is a Markov chain. Within each segment the demand is described by an autoregressive process.

The analysis and forecasting contain the following steps:

1. Assuming: Assume the number of segments of the daily demand pattern. We have used three: "rising," "oscillating," and "falling," and have also tried four segment types, as will be discussed herein. The segment type at each time is termed the system "state." The demand in each segment (state) is assumed to be an autoregressive model of low order.

2. Learning: A portion of the demand data is used, and divided by observation into the segments selected in the "assuming" process just stated. The initial values of the transition probabilities between states, and the

initial values of the parameters in the autoregressive model for each segment are computed.

3. Recognition: The entire demand data set is used, to determine the optimal set of parameters—posterior probabilities for being in each state, transition probabilities between states, and parameters of the autoregressive models.

4. Forecasting: The posterior probabilities of being in each state, and the autoregressive models for the segments are used to forecast the demands for the next day.

The data requirements of the model are modest, as will be shown. It is unfortunate, but in very few systems are demands monitored, and recorded continuously (on time scales of less than a day) and especially, kept in some accessible form. The only data we could get is from Omaha, Nebraska (Christensen and Macdissi 1989). For the systems in Israel that we analyzed, demands were computed from a water balance, using continuously monitored pump operations and reservoir level changes.

DEMAND BEHAVIORAL MODEL

Stochastic Process with Segments

The demand is considered a stochastic process with three repeating segments. Transition from one segment to the next is a Markov chain, and therefore the length of the segments is random. Within each segment the demand is described by an autoregressive (AR) model of low order.

Let ν_t, $t = \ldots, -1, 0, 1, 2, \ldots$, be the state of the demand curve at time t, which may take on m different values, i.e., $\nu_t = 1, \ldots, m$. The transition probabilities from one state to the next are

$$\mathbf{P} = [p^{ij} | p^{ij} = \mathbf{Q}(\nu_t = j | \nu_{t-1} = i), i, j = 1, \ldots, m] \quad (1)$$

The demand at time t, x_t, is a normal autoregressive process of order n:

$$x_t = c_{0t} + \sum_{i=1}^{n} c_{it} x_{t-i} + b_t \zeta_t \quad (2)$$

where $c_t = (c_{0t}, c_{1t}, \cdots, c_{nt})$ = a vector of time-dependent coefficients; ζ_t = a normal variable with zero expectation and unit variance; and b_t = the standard deviation of the disturbing white noise, which is also time dependent.

The vector of parameters in (2), $\boldsymbol{\theta}_t = (c_t, b_t)$, at any time t can assume one of the m values, such that when $\nu_t = k$, $k = 1, \ldots, m$ then $\boldsymbol{\theta}_t = \boldsymbol{\theta}^k = (c_0^k, c_1^k, \ldots, c_n^k, b^k)$. The order of the autoregressive model, m, can be allowed to vary between segment types, to provide more flexibility for model fit. We have data, which is assumed to be a realization of this process, $\mathbf{X}_1^N = (x_1, \ldots, x_N)$. With this data we have to solve two interrelated problems:

1. To estimate the parameters $\mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m$, and
2. To identify the state of the process, ν_t, at any point in time $t[\mathbf{V}_1^N = (\nu_1, \ldots, \nu_N)]$.

Following Yacovlev and Vorob'yov (1986) we use the maximum likelihood method. The parameters are determined from

$$P^*, \boldsymbol{\theta}^{1*}, \ldots, \boldsymbol{\theta}^{m*} = \underset{(\mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)}{\arg\max} \log f(\mathbf{X}_1^N | \mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m) \quad \ldots\ldots\ldots\ldots (3)$$

where $f(x_1^N|\mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)$ is the conditional probability density of the vector $\mathbf{X}_1^N$.

The computation is carried out be an iterative process, to be detailed herein, which generates at iteration s the values $(\mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)$. The sequence of iterates has the following properties (Yacovlev and Yorob'yov 1986):

1. The sequence of numbers $\log f(\mathbf{X}_1^N|\mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)$ is monotone non-decreasing,
2. The estimates $\mathbf{P}_s^*, \boldsymbol{\theta}_s^{1*}, \ldots, \boldsymbol{\theta}_s^{m*}$ that satisfy

$$\log f(\mathbf{X}_1^N|\mathbf{P}_{s+1}, \boldsymbol{\theta}_{s+1}^1, \ldots, \boldsymbol{\theta}_{s+1}^m) - \log f(\mathbf{X}_1^N|\mathbf{P}_s^*, \boldsymbol{\theta}_s^{1*}, \ldots, \boldsymbol{\theta}_s^{m*}) = 0 \quad \ldots (4)$$

are arguments of a local maximum of the likelihood function $f(\mathbf{X}_1^N|P, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)$.

The iterative equations are:

$$\boldsymbol{\theta}_{s+1}^k = \underset{\boldsymbol{\theta}^k}{\arg\max} \sum_{t=1}^{N} p(\nu_t = k|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)\cdot \log f(x_t|\mathbf{X}_1^{t-1}, \boldsymbol{\theta}^k) \quad \ldots (5)$$

$$p_{s+1}^{i,j} = \frac{\sum_{t=1}^{N} p(\nu_{t-1} = i, \nu_t = j|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)}{\sum_{u=1}^{m}\sum_{t=1}^{N} p(\nu_{t-1} = i, \nu_t = u|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)} \quad \ldots\ldots\ldots\ldots (6)$$

where

$$f(x_t|x_1^{t-1}, \boldsymbol{\theta}^k) = \frac{1}{b^k\sqrt{2\pi}} \exp\left\langle \frac{-1}{2(b^k)^2}\left(x_t - c_0 - \sum_{i=1}^{n} c_i^k x_{t-i}\right)^2\right\rangle \quad \ldots\ldots (7)$$

where $p(\nu_t = k|x_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)$ = the posterior probabilities of process x_t being in state k at time t. The algorithm for estimating these is based on properties of a Markov chain and on (7). For details of the algorithm, see Mottl' et al. (1983). At every iteration, (4) is checked, for convergence to the desired solution. The joint distribution is

$$p(\nu_{t-1} = i, \nu_t = j|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m) = p(\nu_{t-1} = i|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)$$
$$\cdot p(\nu_t = j|\nu_{t-1} = i, \mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)$$
$$= p(\nu_{t-1} = i|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m)\cdot p_s^{i,j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

According to Yacovlev and Vorob'yov (1986)

$$\log f(\mathbf{X}_1^N|\mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m) = \sum_{t=1}^{N} \log f(x_t|x_1^{t-1}, \boldsymbol{\theta}^{\nu_t})$$
$$+ \sum_{t=1}^{N} \log p\nu_{t-1}\nu_t - \sum_{t=1}^{N} \log p(\nu_t|x_1^N, \mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m) \quad \ldots\ldots\ldots\ldots\ldots (9)$$

for each state distribution $\mathbf{V}_1^N = (\nu_1, \ldots, \nu_N)$, which makes (4) computable.

[It is a consequence of the presentation of the joint distribution of $\mathbf{X}_1^N$ and $\mathbf{V}_1^N$ in two equivalent ways

$$f(\mathbf{X}_1^N|\mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)\cdot p(\mathbf{V}_1^N|\mathbf{P}, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)$$
$$= p(\mathbf{V}_1^N|\mathbf{P})\cdot f(\mathbf{X}_1^N|\mathbf{V}_1^N, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m)$$

and properties of model (1) and (2) (by taking logarithms of this equation)].

We thus have the following algorithm for estimating the model parameters $P, \boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^m$.

Initial Step: Learning

1. Examine the given data, $\mathbf{X}_1^N$ of length $t = 1, \ldots, N$, and decide how many segment types, m, to use. m is a small number, e.g., $m = 3$. It is desirable to have $N > 200$–300 points.

2. Select an initial part of the data, $\mathbf{X}_1^{N_1}$, for $t = 1, \ldots, N_1$, with $N_1 \ll N$.

3. By observation, divide the selected set into segments, thus setting for each time t its state ν_t. Calculate the initial matrix of transition probabilities as the sample frequencies. Set

$$p(\nu_t = k|\mathbf{X}_1^N, \mathbf{P}_0, \boldsymbol{\theta}_0^1, \ldots, \boldsymbol{\theta}_0^m) = \begin{cases} 1, & \text{if } \nu_t = k \\ 0, & \text{otherwise} \end{cases} \quad \ldots\ldots\ldots\ldots\ldots (10)$$

for $t = 1, \ldots, N_1$. Also set $p(\nu_0 = k|\mathbf{X}_1^N, \mathbf{P}_0, \boldsymbol{\theta}_0^1, \ldots, \boldsymbol{\theta}_0^m)$, arbitrarily. For example

$$p(\nu_0 = k|\mathbf{X}_1^N, \mathbf{P}_0, \boldsymbol{\theta}_0^1, \ldots, \boldsymbol{\theta}_0^m) = p(\nu_1 = k|\mathbf{X}_1^N, \mathbf{P}_0, \boldsymbol{\theta}_0^1, \ldots, \boldsymbol{\theta}_0^m) \quad \ldots (11)$$

4. Calculate $\boldsymbol{\theta}_1^k$, $k = 1, \ldots, m$ according to (5), using (7) for $f(x_t|\mathbf{X}_1^{t-1}, \boldsymbol{\theta}^k)$ and posterior probabilities from (8). All these operations use the data for $t = 1, \ldots, N_1$ only. Calculate the probabilities $p_1^{i,j}$, $i, j = 1, \ldots, m$ according to (6).

Iterative Step: Recognition

For this step the entire length of the data set, $\mathbf{X}_1^N$, for $t = 1, \ldots, N$.

1. Calculate the posterior probabilities

$$p(\nu_t = k|\mathbf{X}_1^N, \mathbf{P}_s, \boldsymbol{\theta}_s^1, \ldots, \boldsymbol{\theta}_s^m) \qquad t = 1, \ldots, N,\ k = 1, \ldots, m \quad .. (12)$$

2. Calculate $\boldsymbol{\theta}_{s+1}^k$, $p_{s+1}^{i,j}$, using (5) and (6).

This is repeated until conditions for a (local) optimum of the likelihood function are satisfied. According to the terms commonly using in pattern recognition, the initial step is called "learning," while the iterative step is called "recognition."

A "loss function" is defined for the classification actually obtained, $\hat{\mathbf{V}}_1^N = (\hat{\nu}_1, \ldots, \hat{\nu}_N)$ relative to the (unknown) true classification $\tilde{\mathbf{V}}_1^N = (\tilde{\nu}_1, \ldots, \tilde{\nu}_N)$. The minimum of this function is sought. The simplest rule to achieve this is

$$\nu_t(\mathbf{X}_1^N) = k^* \qquad \text{if } k^* = \arg\max_k p(\nu_t = k|\mathbf{X}_1^N, \mathbf{P}^*, \boldsymbol{\theta}^{1^*}, \ldots, \boldsymbol{\theta}^{m^*}) \quad .. (13)$$

Mottl' et al. (1983) have shown that this yields

$$\hat{\mathbf{V}}_1^N(\mathbf{X}_1^N) = \arg\min_{\mathbf{V}_1^N} \sum_{t=1}^{N} \bar{\lambda}(\nu_t, \hat{\nu}_t) \quad \text{(14)}$$

where

$$\bar{\lambda}(k, k) = 0, \qquad \bar{\lambda}(k, l) = 1, k, l, = 1, \ldots, m, k \neq 1 \quad \text{(15)}$$

This simple decision rule may cause some "lack of definition" at boundaries between segments, i.e., the change from one type to the next is sometimes missed. This can be remedied by more complex minimization criteria. Mottl' and Muchnik (1984) have suggested such a criterion, which was then minimized by dynamic programming. For our application, it seems that the use of the simple criterion given in (13) produces adequate results.

APPLICATION

The data are two months of hourly water demands in the Sorek water system, in central Israel, during the summer of 1991. The system serves the town of Yavne, of about 20,000 inhabitants, which use 20% of the water, with the remaining 80% for agricultural irrigation. Total daily demand averages 22,000 m^3.

The demand data were actually computed, rather than measured directly, from continuous records of pump station outputs and reservoir level changes, by a simple water balance. Because there were some unreasonable variations in this data, it was decided to smooth it, using a three-point moving average. The resulting series is thus considered to be the demand data. The period of record is from 1.06.91 to 31.07.91, a total of 61 days, or 1,464 h. A five-day segment is shown in Fig. 2.

The number of states selected is $m = 3$: 1 = "rising," 2 = "oscillations," and 3 = "falling." The first N_1 = eight days of the data used in the "learning" step. The order of the autoregressive model was allowed to vary between the segment types, and the optimal orders (according to standard errors and subjective observations) were $n_1 = 3$, $n_2 = 3$, $n_3 = 1$; i.e., the "rising" state will be an autoregressive model of order 3, and so on. The standard errors of these three models were: $S_1 = 37.7$, $S_2 = 33.0$, $S_3 = 37.4$ as compared with a standard deviation of 265 and a mean of 697 in the whole data series. The coefficients of the three autoregression models appear in Table 1.

The "learning" step is seen in Fig. 4. At the bottom is the value of $\nu_t = 1, 2, 3$, which delineates the corresponding states "rising," "oscillations," and "falling," respectively. It demonstrates that the pattern is indeed detected, and the transitions are always "smooth," i.e., (··· 111222 ··· 222333

TABLE 1. Coefficients for Autoregressive Models with Three States

Segment (state) (1)	C_0 (2)	C_1 (3)	C_2 (4)	C_3 (5)
1	33.85	1.54	−0.92	0.34
2	45.7	1.79	−1.23	0.43
3	−8.996	0.88	—	—

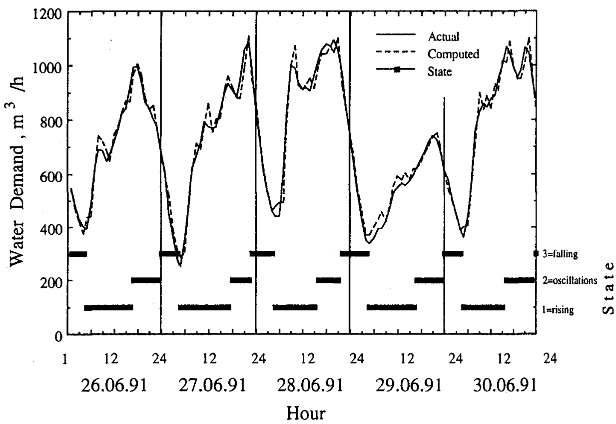

FIG. 4. Approximation and Segmentation Results of Three-State Model

TABLE 2. Autoregressive Model Coefficients and Standard Error for Four-State Model

State (1)	Model Coefficients				Standard error, S (6)
	C_0 (2)	C_1 (3)	C_2 (4)	C_3 (5)	
Rising slowly	21.38	1.9	−1.41	0.53	25.8
Rising rapidly	86.8	1.6	−0.99	0.32	36.9
Oscillations	33.85	1.54	−0.92	0.34	33.3
Falling	−8.996	0.88	—	—	37.4

··· 33111 ···). Thus no difficulties are encountered at the boundaries between states. The modeled demand shows good fit to the actual demand.

June 29, 1991 is a Saturday, and it has a significantly different demand pattern than the weekdays. Note that July 6, 1991, which appears in Fig. 2, is also a Saturday, and shows a similar pattern. On Saturday the rate at which the "rising" ($\nu = 1$) segment rises is lower than on the other days. We therefore tried a model with four states: "rising slowly," "rising rapidly," "oscillation," and "falling." The coefficients for the four autoregressive models, and the standard errors of the four types appear in Table 2.

The "falling" model remained unchanged; the "oscillations" model changed only slightly. The improved fit of the four-state model over the three-state model can be seen in Fig. 5, for Saturday, July 6, 1991.

FORECASTING

State Prediction

The estimation algorithm produces posterior probabilities of each state at all times. The sequence of values of these estimations can be viewed as m time series, as shown for the three-state model in Fig. 6. (We can see values slightly greater than 1 on the figures of posterior probabilities. It seems that for the purpose of the actual application, such a precision of the estimation is sufficient. But in the real-time system, it is necessary to take care of the data precision for prevention of cumulative errors.) These time

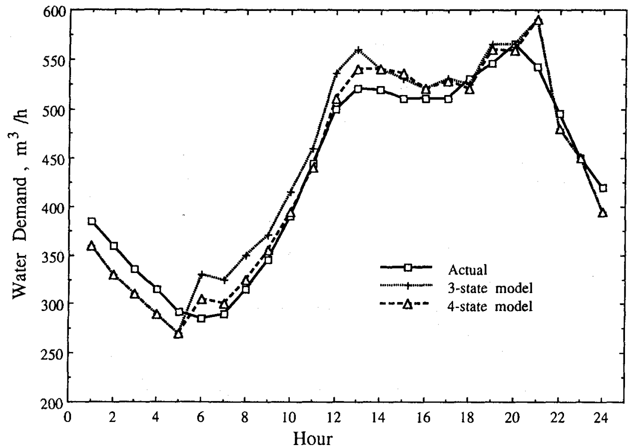

FIG. 5. Results of Three-State and Four-State Models for Saturday, July 6, 1991

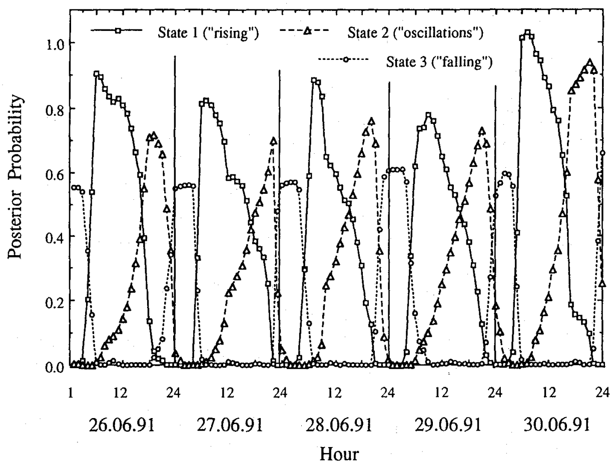

FIG. 6. Estimated Posterior Probabilities of Affiliations of Point x_t to One of States (Three-State Model)

series are seen to have a fixed period of 24 h, and can be modeled by an ARIMA model (Box and Jenkins 1970), which is stationary and periodic, with period T (= 24 h), of the form.

$$(1 - B^T)[1 - \phi_1(B)]p_t = [1 - \Gamma_1(B^T)][1 - \Gamma_2(B)]\zeta_t \quad (16)$$

where p_t = posterior probability, viewed as a stochastic process; ζ_t = normal variable (0, 1); B^T = backshift operator ($B_{x_t} = x_t - x_{t-1}$); and ϕ_1, Γ_1, Γ_2 = polynomials.

The necessary and sufficient condition that this process be stationary is that all the zeros of its polynomials lie outside the unit circle, i.e., the absolute values of all its roots are greater than 1. Using the demand data for the period of June 1–20, 1991, we obtained the following ARIMA models for the three time series of posterior probabilities:

State ν = 1: "Rising"

$$(1 - B^{24})(1 - 1.6B^1 + 0.74B^2)p_t^{(1)} = (1 - 0.61B^{24})\zeta_t \qquad (17a)$$

Standard error = 0.038; and R^2 = 0.98.

State ν = 2: "Oscillations"

$$(1 - B^{24})(1 - 1.6B^1 + 0.69B^2)p_t^{(2)} = (1 - 0.62B^{24})\zeta_t \qquad (17b)$$

Standard error = 0.034; and R^2 = 0.98.

State ν = 3: "Falling"

$$(1 - B^{24})(1 - 1.3B^1 + 0.57B^2)p_t^{(3)} = (1 - 0.69B^{24})\zeta_t \qquad (17c)$$

Standard error = 0.029; and R^2 = 0.97, where R is the correlation between data (posterior probabilities from the fitting model) and the forecast [values according to (17)]. It is easy to check that stationarity is indeed detected, i.e., that the absolute values of all roots are greater than 1.

The results of these models are demonstrated for state ν = 1 ("rising") in Fig. 7. The forecast is produced each day at 0:00 for 24 h ahead. With these forecasted probabilities, the state for each time interval is determined through application of the decision rule (13), i.e., at each time the state with highest probability is selected.

To assess the performance of this method for forecasting the states, we proceeded as follows, using the time period of July 1–31, 1991. First we determined the state of the process, ν_t, at every hour t, by the process which ends with (13). Then we forecasted for the same time period the state ν_t, as described in this section. The measure of fit is

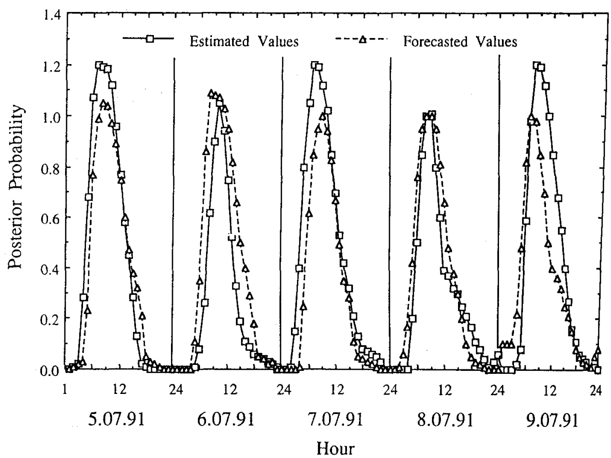

FIG. 7. Forecasting of Posterior Probabilities by ARIMA Model; State 1 ("Rising"); Forecasting: Each Day at 0:00 O'clock to 24 h Ahead

$$\rho = \frac{1}{N}\sum_{t=1}^{N} \lambda(\nu_t, \hat{\nu}_t) \qquad (18)$$

$$\lambda(\nu_t, \hat{\nu}_t) = 1, \quad \text{if } \hat{\nu}_t = \nu_t \qquad (19a)$$

$$\lambda(\nu_t, \hat{\nu}_t) = 0, \quad \text{otherwise} \qquad (19b)$$

The result for the 31 × 24 = 744 h was ρ = 0.82, i.e., 82% of the forecasted values match. Note that the ARIMA models (17*a*)–(17*c*) were based on data for the dates of June 1–20, 1991, whereas the performance of these ARIMA models was examined for a later period.

Futhermore, the sequence of states moved "smoothly" from 1 to 2, from 2 to 3, then back to 1. Only six times in the 93 "boundaries" between states did we get "nonsmooth" sequences like (··· 1112122 ···), and the effect of this was minimal.

Demand Forecasting

Next we forecast the demands for the period of July 1–31, 1991, using the forecasted states. Two forecasts were produced:

1. By the three-state model, where all days have the same "rising" model, and
2. By the four-state model: first, the states were produced by the three-state model, then all weekdays were forecasted by the "rising-rapidly" model, whereas Saturdays were forecasted by the "rising-slowly" model.

For estimation (recognition) we used a window of 15 previous days. It means that for the recognition step, we used the interval $\mathbf{X}_{N-360}^{N}$.

Results for July 5–9, 1991 can be seen in Fig. 8. The second method is seen to give somewhat better results, although sometimes the first is better, as is the case on Friday, July 5, 1991.

Table 3 contains results of several measures of the forecast produced for

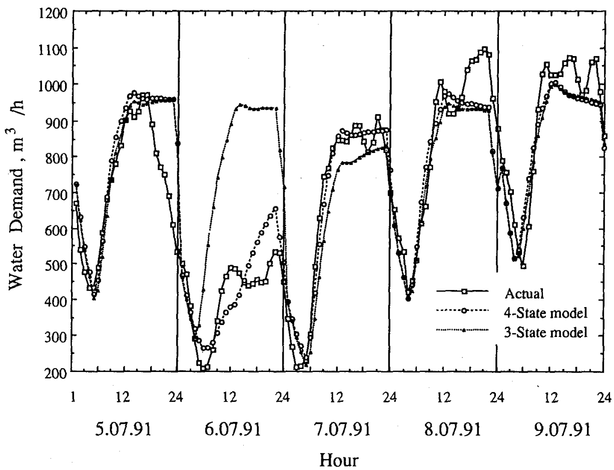

FIG. 8. Three-State and Four-State Models Demand Forecasting: Each Day at 0:00 O'clock for 24 h Ahead

TABLE 3. Evaluation of Three-State and Four State Models' Forecast for Day of July 1991

Date (July) (1)	Day (2)	Actual daily volume (m^3) (3)	Four-State Model				Three-State Model			
			Daily volume (m^3) (4)	Deviation (%) (5)	Standard error (6)	Maximum error (7)	Daily volume (m^3) (8)	Deviation (%) (9)	Standard error (10)	Maximum error (11)
1	Monday	15,618.8	17,997.8	+15.2	133.3	257.0	17,669.57	+13.13	113.4	226.13
2	Tuesday	18,450.3	18,143.8	−1.6	52.2	99.2	17,767.12	−3.702	50.722	113.54
3	Wednesday	19,889.2	16,884.3	−15.1	152.8	260.2	16,508.77	−17.03	169.07	300.48
4	Thursday	19,882.7	17,867.9	−10.1	152.8	340.0	17,541.13	−11.77	159.74	348.11
5	Friday	17,302.5	18,306.2	+5.8	63.1	236.3	17,709.98	+2.354	50.170	98.782
6	Saturday	10,117.5	11,586.9	+14.5	97.3	136.8	16,586.71	+63.93	312.95	465.89
7	Sunday	19,908.0	14,788.9	−25.7	278.9	246.3	13,301.63	−33.18	333.36	624.53
8	Monday	16,576.1	20,095.9	+21.2	170.3	559.0	19,711.73	+18.91	156.16	270.18
9	Tuesday	18,157.1	18,799.9	+3.5	99.6	293.8	18,197.11	+0.219	94.573	209.18
10	Wednesday	20,943.3	19,137.0	−8.6	105.6	185.8	18,712.23	−10.65	118.92	287.68
11	Thursday	21,644.6	24,761.5	+14.3	186.5	464.7	25,189.73	+16.37	196.13	472.46
12	Friday	17,314.8	19,407.3	+12.0	134.7	210.6	18,946.49	+9.423	131.97	349.27
13	Saturday	9,929.1	10,440.3	+5.1	83.0	348.2	17,636.13	+77.61	370.76	494.58
14	Sunday	16,004.5	16,222.2	+1.3	47.6	160.1	14,815.3	−7.430	86.804	180.02
15	Monday	20,136.3	19,102.1	−5.1	90.7	92.4	18,632.42	−7.468	93.448	166.35
16	Tuesday	21,469.5	20,525.3	−4.3	79.0	167.7	20,361.59	−5.160	75.239	116.83
17	Wednesday	20,886.9	19,435.6	−6.9	137.7	136.9	18,901.4	−9.506	149.00	276.32
18	Thursday	19,655.0	22,719.1	+15.5	227.6	259.3	22,971.75	+16.87	226.30	505.95
19	Friday	20,116.7	19,120.6	−4.9	79.6	502.4	18,493.5	−8.069	108.84	231.48
20	Saturday	11,854.9	11,480.8	−3.1	89.2	251.1	18,429.21	+55.45	311.12	433.58
21	Sunday	20,052.8	16,861.7	−15.9	163.6	190.8	15,487.92	−22.76	220.69	359.73
22	Monday	19,201.1	19,088.0	−0.5	71.6	162.8	18,650.63	−2.867	74.059	169.75
23	Tuesday	20,533.4	22,254	+8.3	164.8	292.9	22,500.4	+9.579	171.33	311.08
24	Wednesday	17,324.7	18,846.8	+8.7	106.2	309.4	18,043.19	+4.146	90.038	164.47
25	Thursday	17,575.2	19,080.1	+8.5	109.3	281.9	18,482.18	+5.160	104.05	195.88
26	Friday	16,867.7	18,575.3	+10.1	142.4	170.8	17,853.66	+5.844	129.72	326.83
27	Saturday	10,653.7	10,569.5	−0.7	83.0	225.2	16,133.6	+51.43	297.33	451.47
28	Sunday	24,620.9	18,270.5	−25.7	339.1	338.5	17,784.09	−27.76	360.04	543.19
29	Monday	22,227.0	21,929.4	−1.3	80.5	182.4	22,155.21	−0.323	80.914	162.37
30	Tuesday	21,419.5	19,397.7	−9.4	135.8	531.5	19,055.66	−11.03	155.59	296.84
31	Wednesday	16,876.5	19,910.4	+17.9	151.8	152.7	19,565.3	+15.93	144.36	325.63

July 1991. Column 3 gives the actual total daily demand. Columns 4–7 are for the four-state model, and columns 8–11 for the three-state model. In each case the four columns are: 24-h totals relative difference between the two (%), standard error of the hourly values, and maximum error of the hourly values. The results show that the four-state model performs better, but its improvement over the three-state model varies over the days of the week.

Table 4 summarizes the results of Table 3, by days of the week. The results, by column order, are: averages for each day of the week of the corresponding columns in Table 3 (columns 5, 6, 7, 9, 10 and 11; column 2 of Table 3—averages of absolute values of column 5, Table 4.). The results indicate an advantage of four-state model, obviously on Saturday.

EVALUATION OF FORECASTING METHOD

Recall that we are forecasting hourly water demands for the purpose of determining the optimal operating schedule of a supply system. The forecast is produced in real time, at any desired time, usually for a period of one day ahead, sometimes up to a week. The forecast is used to plan the operation, and this plan is then put into action. The actual performance of the system is monitored, for example water levels in the reservoirs. If the actual state of the system deviates from the calculated state, we conclude that actual demands differ from the forecast, and we must update them.

To assess the performance of the proposed method in real time, we conducted the following experiment:

- Every forecast was based on hourly demand data over the last 15 days only.
- For every hour from July 1, 1991 to July 30, 1991, we produced the forcast to the end of the current day and for 24 h ahead.

Computation time for each forecast of 24 h ahead, which includes model "learning" and "recognition," fitting the ARIMA models, then predicting the states and forecasting, on a PS/2 model-70 personal computer, is approximately 60 s.

As a measure of performance we used the deviation of the total volume for the forecasting period from the actual volume. This selection is based

TABLE 4. Evaluation of Three-State and Four-State Models' Forecasts: Averages by Day of Week

	Four-State Model			Three-State Model		
Day (1)	Deviation (%) (2)	Standard error (3)	Maximum error (4)	Deviation (%) (5)	Standard error (6)	Maximum error (7)
Sunday	17.1	207.3	233.9	22.7	250.2	426.8
Monday	8.7	109.3	250.7	8.5	103.5	198.9
Tuesday	5.4	106.3	277.0	5.9	109.4	209.4
Wednesday	11.4	130.8	209.0	11.4	134.2	270.9
Thursday	12.1	169.0	336.5	12.5	171.5	380.6
Friday	8.2	105.0	280.0	6.4	105.1	251.5
Saturday	5.9	88.1	240.3	62.1	323.0	461.3

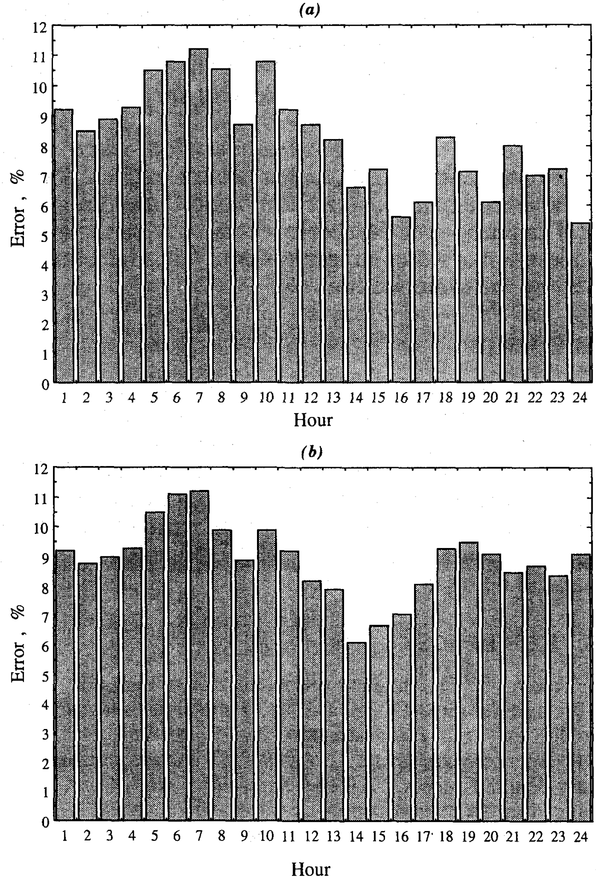

FIG. 9. Relative Deviation of Total Volume: (*a*) Forecasting to End of Day; and (*b*) Forecasting for 24 h Ahead

on the observation from Tables 3 and 4 that this measure is in good correlation with all other measures. We computed the average of this relative difference in volumes, for each hour of the day, over all 30 days of July. The results are seen in Figs. 9(*a*) and (*b*). The first is for the period to the end of the current day, the second for a 24-h period ahead.

The difference ranges from about 6% to 11%, in both cases. The decrease in error towards the end of Fig. 9(*a*) is explained by the fact that the remaining period becomes smaller.

From our experience with optimal operation of the water supply system we have concluded that this accuracy is quite adequate.

CONCLUSIONS

The method presented in this paper is based on concepts taken from pattern recognition and time series analysis. The daily water-demand pattern is observed to be made up of three segments, called "states" since they are modeled as a Markov process, and in each the demand is described by a low-order ARIMA model. The posterior probabilities of the Markov chain and the parameters of the the ARIMA models are determined from a relatively small sample of data, and a forecast for the coming 24 h can be produced with this model.

The model is well suited for real-time operation, where the state of the system is monitored continuously, and as deviations from the planned state are detected, a new demand forecast can be produced easily, with updated information. The purpose of the forecast is to provide the input to methods for optimal operation of water-supply systems, although other purposes may also be served by the same method.

ACKNOWLEDGMENTS

The work presented herein is part of an extensive project at Mekorot Water Co. Ltd., in the Command and Control Unit headed by Y. Orenstein. The objective is to develop the "plant software" for real-time control and operation of the water-supply system. The software includes: network simulation for single and multiquality systems, optimal operation by various optimization methods, and expert systems. In parallel, an advanced Supervisory Control and Data Acquisition (SCADA) software system has been developed, and is being implemented.

APPENDIX I. REFERENCES

Boston Water Study. (1966). Maguire Associates, Boston, Mass.

Box, G. E. P., and Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. Holden Day, San Francisco, CA.

Christensen, P. E., and Macdassi, T. J. (1989). "Computerized analysis and prediction of hourly demand curves." *AWWA Speciality Comp. Conf.*, American Water Works Association, Denver, Colo.

Gray, D. F. (1978). "Use of consumption predictors." *Tech. Rep. 4/78*. Engrg. Dept., Cambridge Univ., Cambridge, England.

Hartley, J. A., and Powell, R. S. (1991). "The development of a combined demand prediction system." *Civil Engrg. Systems*, 8(4), 231–236.

Jowitt, P. W., and Xu, C. (1992). "Demand forecasting for water distribution systems." *Civ. Engrg. Systems*, 9, 105–121.

Maidment, D. R., Miaou, S. P., and Crawford, M. M. (1985). "Transfer function models for daily urban water use." *Water Resour. Res.*, 21(4), 425–432.

Maidment, D. R., and Parzen, E. (1984). "Time patterns of water use in six Texas cities." *J. Water Resour. Plan. Mgmt.*, ASCE, 110(1), 90–106.

Moss, S. M. (1978). "On line optimal control of a water supply system." PhD thesis, Cambridge, Univ., Cambridge, England.

Mottl', V. V., and Muchnik, I. B. (1984). "Segmentation of structural curves using dynamic programming methods." *Automation and Remote Control*, 45(1/Part 2), 88–94 (translated from Russian).

Mottl', V. V., Muchnik, I. B., and Yacovlev, V. G. (1983). "Optimal segmentation of experimental curves." *Automation and Remote Control*, 44(8), 1035–1044 (translated from Russian).

Sterling, M. J. H., and Bagrelia, A. B. (1985). "Adaptive forecasting of daily water demands." *Comparative models for electrical load forecasting*," D. W. Bunn and E. D. Farmers, eds., John Wiley and Sons, New York, N.Y., 213–215.

Yacovlev, V. G., and Vorob'yov, S. A. (1986). "Estimation of model parameters of random processes with instantly changing properties." *2nd IFAC Symp. Stochastic Control*, International Federation of Automatic Control, 244–248.

APPENDIX II. NOTATION

The following symbols are used in this paper:

B = backshift operator;
b_t = standard deviation of disturbing white noise;
$f(\mathbf{X}_i^j|\cdot)$ = conditional probability density of demand values vector;
$f(x_t|\cdot)$ = conditional probability density of demand value at time t;
n_i = order of the autoregressive model in state i;
$\mathbf{P}$ = the set of transition probabilities;
$\mathbf{P}_t^{(\cdot)}$ = posterior probability of process x_t being in state $(\cdot)$ at time t;
p^{ij} = $\mathbf{Q}(\nu_t = j|\nu_{t-1} = i)$—transition probability from state i at time $t - 1$ to state j at time t (independent of time);
$p(\mathbf{V}_i^d|\cdot)$ = conditional probability of state vector;
$p(\nu_t|\cdot)$ = conditional probability of state at time t;
$p(\nu_{t-1} = i, \nu_t = k|\cdot)$ = conditional probability of the process x_t being in state i at time $t - 1$ and in state k at time t;
q_t = $(c_{0t}, c_{1t}, \ldots, c_{nt}, b_t)$—time-dependent autoregressive parameters;
R = correlation coefficient;
S_i = standard error of the model in state i;
s = iterative step index;
t = $\ldots, -1, 0, 1, 2, \ldots$—time;
$\mathbf{V}_1^j$ = $(\nu_i, \nu_{i+1}, \ldots, \nu_j)$—time series of states from $t = i$ to $t = j$;
$\hat{\mathbf{V}}_1^N$ = computed state vector;
$\tilde{\mathbf{V}}_1^N$ = unknown true state vector;
$\mathbf{X}_i^j$ = $(x_i, x_{i+1}, \ldots, x_j)$—time series of demands from $t = i$, to $t = j$;
x_t = the demand process variable;
Γ_1, Γ_2 = polynomials;
ζ = normal variable with zero expectation and unit variance;
$\boldsymbol{\theta}^k$ = autoregressive parameters vector at state k;
$\boldsymbol{\lambda}$ = diagonal unit matrix;
$\bar{\boldsymbol{\lambda}}$ = antidiagonal unit matrix;
ν_t = state of demand curve at time t;
ρ = the measure of fit of forecasted state vector to estimated state vector; and
ϕ_1 = polynomial.