INTERVENTION ANALYSIS TO ESTIMATE PHOSPHORUS LOADING SHIFTS

K. A. Booman*, L. Pallesen** and
P. M. Berthouex***

*The Soap and Detergent Assoc., 475 Park Avenue South, New York, NY 10016, U.S.A.
**Technical University of Denmark, IMSOR, Lyngby, Denmark
***Dept. of Civil and Environmental Engineering, University of Wisconsin, Madison, WI 53706, U.S.A.

ABSTRACT

Phosphate laundry detergents have been banned in the Great Lakes drainage basin and a few other regions of the U.S.A. New bans are considered from time to time. Any consideration of new bans should include an evaluation of changes that have occurred in past bans. This paper presents a method for estimating such changes. The method is illustrated with treatment plant influent phosphorus mass loading data. All data series analyzed to date were complicated by drift, large variation, and sometimes seasonal patterns. With such data an appropriate time series model is used to estimate the shift. An ARIMA (0,1,1) model has been used successfully with many data sets. This model seems to be robust and has considerable intuitive appeal. A seasonal component was needed in the model for a few data sets. A Bayesian approach was developed to estimate the shift. The effect of a detergent phosphate ban on influent treatment plant phosphorus load appears to be about 0.3 kg/cap.-yr.

KEYWORDS

Intervention analysis, time series models, phosphorus loads, detergent bans, Bayesian analysis.

INTRODUCTION

Intervention in environmental systems is often promulgated without complete knowledge of the system. The promulgators typically have confidence that the intervention will cause a real change, and in the direction of better quality, but without precise knowledge of how large the change will be. The first bans on phosphate laundry detergents in the Great Lakes drainage basin were interventions of this kind. Any consideration of new bans should include an assessment of the existing data which show the effect of the bans. The largest effect of a ban will be on raw sewage phosphorus concentration and treatment plant mass loadings. The decrease in influent phosphorus loads of several treatment plants is estimated and the method used to make the estimates is explained. The general problem is to estimate "the effect of an intervention that has been made with the intent of causing a system to change where the behavior of the system is indicated by a set of data that are a time series, and so the order of the data as well as their magnitude is important (Box and Tiao, 1975)."

Estimating the change in phosphorus load when a detergent ban goes into effect may seem straightforward. Abundant data exist for the period before the ban; some data are available after the ban. It is tempting to assume that there are fixed levels before and after the ban and that the shift due to the ban is the difference between these averages. Having assumed this much, one is likely to also assume that all variation about the levels is random and that a t-test could be used to assess the statistical significance of the shift. This approach, taken by Hartig and Horvath (1982) and Jones and Hubbard (1986), has been criticized by Berthouex et al. (1983), Pallesen et al. (1985), and Booman and Sedlak (1986).

Unfortunately, this method will often give misleading results, particularly for the confidence limits of the estimated shift. Of more than 30 sets of data analyzed to date, none could be modeled as stationary levels with random variation. Most displayed drift or trend, some had seasonal patterns, and all had significant serial correlation. These properties of the data invalidate the approach of subtracting averages and using a t-test. If there is a drift, simple averages are not a fair representation of the data. Observations distant from the intervention would have little relevance to estimating the effect of the ban; data near the ban should carry high weight. If there are seasonal effects, simple averages may be biased by the number of observations in the before and after averages. This bias will be exaggerated if the record length is not a multiple of the length of the season, i.e. 12 months, 24 months, etc. Ignoring serial correlation will make the averages seem more precise than they actually are and bias the significance tests toward indicating a statistically significant shift when none has occurred.

Time series analysis provides the means for properly taking the properties of the data into account. A one parameter model fits most data sets. Where there is a seasonal pattern, a slightly more complicated model is used.

A NONSEASONAL MODEL

The basic model describes the data, which are observed with error, as being generated by a process that drifts as a random walk. This model, called a random walk-white noise model in Pallesen et al. (1985) is

$$Y_t = y_t + e_t \tag{1}$$

$$y_t = y_{t-1} + E_t \tag{2}$$

where y_t is the underlying true but unmeasurable value of the variable, Y_t is the observed value of the variable, e_t is the instantaneous error, and E_t is the shock causing the random walk (drift).

Equations 1 and 2 can be combined to express this as an ARIMA (0,1,1) model of the Box-Jenkins (1976) family of time series models:

$$Y_t = Y_{t-1} + a_t + \theta\, a_{t-1} \tag{3}$$

This representation of the data has considerable intuitive appeal. It is easier to accept than completely random variation about a fixed level. Levels at a particular time are estimated by an exponentially weighted moving average, thus giving recent observations more weight than distant observations. The parameter θ is the weighting factor for the moving averages.

A ban does not become effective immediately. There is a transition period of about four months. The transition period is treated as a gap, i.e. observations within the transition period are omitted. The forecast of the moving average model is a horizontal projection. The shift is estimated by forecasting across the ban, forecasting forward in time from before the ban, and backward in time from the post-ban level. The variance of the shift is the sum of the variances of the estimated pre- and post-ban levels plus the variance of the forecasts across the transition period. Details are given in Pallesen et al. (1985).

THE SEASONAL MODEL

The data sets that are seasonal have an annual cycle. While the seasonality is complex within years (not sinusoidal), it has been consistent year to year and the variance associated with it is very small. It has been successfully represented parsimoniously as a seasonal random walk. A seasonal model that has been useful for several sets of data is

$$Y_t = y_t + s_t + e_t \qquad e_t \in N(0,\sigma_1^2) \tag{4}$$

$$y_t = y_{t-1} + f_t \qquad f_t \in N(1,\sigma_2^2) \tag{5}$$

$$s_t = s_{t-12} + S_t \qquad g_t \in N(0,\sigma_3^2) \tag{6}$$

Instantaneous error, including observation error, is accounted for by Equation 4. Equation 5 represents a random walk component. Equation 6 incorporates the seasonal pattern, also as a random walk. s_t is the seasonal effect component and f_t is the random shock associated with the seasonal effect. This model yields level forecasts.

The parameters in this model (Equations 4, 5, and 6) are the sigmas $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ plus an initial state specification x(0). The Kalman filter-state space approach can be used to calculate the likelihood for any given set of values for σ. Moreover, for a series which includes an intervention, which in turn gives rise to a gap of length G during which the series has been shifted by the amount δ, this permits the likelihood to be calculated given σ and δ. This is an important prerequisite for the Bayesian analysis presented below for the seasonal case.

BAYESIAN ANALYSIS

In the Bayesian framework, the estimate of the shift δ, based on the data Y, can be made unconditionally on the other parameters σ. In the Bayesian framework, the estimate of the shift delta, δ, based on the data Y will be made unconditionally on the other parameters sigma. As will be explained, this is ultimately accomplished by estimating delta as the weighted average of a number of individual estimates of delta in the region of maximum likelihood in log sigma space. This has considerable appeal when the curvature of the likelihood surface is quite low and the point of maximum likelihood is difficult to locate with precision, as it is in the present instance. As well as providing a basis for estimating delta in such situations, the approach provides a realistic estimate of standard error of delta as well.

The marginal posterior distribution $p(\delta|Y)$ based on the noninformative prior distribution for the parameters is

$$p(\delta, \sigma_1, \sigma_2, \sigma_3) \propto p(\delta)p(\sigma_1)p(\sigma_2)p(\sigma_3) \propto c/\sigma_1\sigma_2\sigma_3 \quad (7)$$

$$\propto p(\delta, \log\sigma_1, \log\sigma_2, \log\sigma_3) \propto c \quad (8)$$

In general

$$p(\delta,\sigma|Y) \propto p(Y|\delta,\sigma)\, p(\delta,\sigma) \quad (9)$$

where the first factor on the right-hand side of 9 is the likelihood; the second factor is the prior specified in 7 and 8. It is helpful to factor the joint density as

$$p(\delta,\sigma|Y) = p(\delta|\sigma,Y)\, p(\sigma|Y) \quad (10)$$

where

$$p(\sigma|Y) = \int\delta\, p(\delta,\sigma|Y)\, d\delta \quad (11)$$

Since the parameters σ are really nuisance parameters to be integrated out, $p(\sigma|Y)$ shall not be derived analytically. It is sufficient to approximate it by its density values over a grid of combinations of σ_1, σ_2, σ_3.

It holds in general that the conditional distribution of δ is

$$p(\delta|\sigma,Y) \propto p(\delta,\sigma|Y) \quad (12)$$

Since the noninformative prior for (δ, σ) was equivalent to a locally uniform prior for $(\delta, \log\sigma_1, \log\sigma_2, \log\sigma_3)$, it is clear that the posterior distribution for these parameters is proportional to the likelihood in the parameter space of $(\delta, \log\sigma_1, \log\sigma_2, \log\sigma_3)$.

It is a property of likelihood functions that they tend towards normal (Gaussian) as the number of observations gets large. This means, in particular, that $p(\delta|\sigma,Y)$ can be considered approximately normal. Hence, for given values of $(\log\sigma_1, \log\sigma_2, \log\sigma_3)$, i.e., for known σ, the conditional distribution $p(\delta|\sigma,Y)$ can be constructed from just three values of the likelihood calculated from three δ values in the region around the conditional maximum.

A consequence of the likelihood being approximately normal is that the log likelihood is approximately quadratic. Fitting a second order polynomial through these three points produces a fitted function

$$l(\delta) = A\delta^2 + B\delta + C \tag{13}$$

Comparing this expression to the logarithm of a normal density function,

$$N(\mu,\sigma^2): \ \ln f(\delta) = -(1/2)\ln(2\pi) - \ln(\sigma) - (1/2)((\delta-\mu)/\sigma)^2 \tag{14}$$

yields the equivalence

$$A = -(1/2)\sigma^2 \quad \text{<-->} \quad \sigma^2 = -1/2A$$

$$B = \mu/\sigma^2 \quad \text{<-->} \quad \mu = -B/2A \tag{15}$$

so that, given σ, the posterior distribution of δ is

$$p(\delta|\sigma,Y) \in N(\mu = -B/2A,\ \sigma^2 = -1/2A) \tag{16}$$

In other words, this determines $p(\delta|\sigma,Y)$, including the standardizing factor.

To get the unconditional distribution of δ, note that

$$p(\delta|Y) = \int p(\delta,\sigma|Y)\, d\sigma = \int p(\delta|\sigma,Y)\, p(\sigma|Y)\, d\sigma \tag{17}$$

This shows that $p(\delta|Y)$ is really a weighted sum of conditional distributions $p(\delta|\sigma,Y)$, the weights being proportional to $p(\sigma|Y)$. Since the integration (Equation 17) is to be carried out numerically, we are really making the approximation that

$$\int p(\delta|\sigma,Y)\, p(\sigma|Y)\, d\sigma \approx \sum p\,(\delta|\sigma,Y)\, p(\sigma|Y) \tag{18}$$

where the summation is over a grid of points in the parameter space of σ. This grid should have points laid out uniformly in the space ($\log \sigma_1$, $\log \sigma_2$, $\log \sigma_3$). This will make the weights on the right-hand side of Equation 18 proportional to the likelihood function integrated over the δ-dimension (since the likelihood and the posterior are proportional in that space).

$p(\sigma|Y)$ is calculated in two steps. First, the related magnitudes of $p(\sigma|Y)$ at the grid points are found, the values of which are really posterior probabilities for the regions around each grid point. Then, the relative probabilities are standardized so that they sum to one before the summation (Equation 18) is carried out, i.e., $p(\sigma|Y) = 1$. The relative magnitudes of $p(\sigma|Y)$ at the grid points are found by noting that the maximum density for a $N(\mu = -B/2A,\ \sigma^2 = -1/2A)$ distribution is

$$f(\delta = \mu) = 1/\sigma \sqrt{-A/\pi} \tag{19}$$

For a given value of σ (a given grid point), the nonstardardized posterior had as its maximum

$$\text{max likelihood} = \max\{\exp(A\delta^2 + B\delta + C)\} \tag{20}$$

$$= \exp\{A(-B/2A)^2 + B(-B/2A) + C\} = \exp(C - B^2/4A)$$

For $p(\delta|\sigma,Y)$ to integrate to 1.000, standardize by the factor

$$k = (\sqrt{-A/\pi})/\exp(C - B^2/4A) \tag{21}$$

Consequently, the marginal posterior density of σ at this grid point is inversely proportional to k,

$$p(\sigma|Y) \propto (-A)^{-1/2} \exp(C - B^2/4A) \tag{22}$$

where the common factor π has been dropped for convenience. Combining Equation 22 and Equation 19 yields the weight function $p(\sigma|Y)$ to be used in Equation 17 to find $p(\delta|Y)$.

In summary, $p(\delta|Y)$ is a weighted sum of normal distributions. Both the conditional distributions and weights are found from second-order polynomial fits to log likelihood values at three δ points, this being done over a grid of ($\log \sigma_1$, $\log \sigma_2$, $\log \sigma_3$). The Bayesian approach, applied here to the seasonal case, can be easily adapted to other cases.

RESULTS

The basic white noise-random walk model (ARIMA (0,1,1)) was applied to 31 sets of data. It adequately fit 12 of 21 data sets for Wisconsin treatment plants and 6 of 10 data sets for plants in Michigan.

Figure 1 shows data from Milwaukee, WI. The data are monthly average influent phosphorus loadings, expressed as kg per day, for the Jones Island and South Shore plants combined. There have been three interventions: the July 1979 ban, the July 1982 lift of the ban, and the January 1984 reinstatement of the ban. The law requiring the 1979 ban contained a "sunset" provision for a three-year trial period, during which the effects of the ban were to be evaluated. The ban was abandoned, in accordance with the sunset provision of the statute, because no definite benefits could be shown due to having the ban in force. The ban was reinstated in 1984. This did not result because scientific evidence showed that a ban was necessary. It resulted from political activity by groups which believed that any method of reducing phosphorus would ultimately be rewarded.

There was a steady downward trend from 1975 (month 1 = January 1975) until about January 1982. Obviously, taking simple arithmetic averages of intervals before and after the interventions would be incorrect because such an analysis would not account for this trend. The simple time series model accounts for this trend when the shift is estimated, but this is accomplished indirectly by giving more weight to data closest to the transition period.

The estimation was done using the Bayesian estimation method under the condition that the effect of the three interventions (1979 ban, 1982 ban-lift, and 1984 ban) were equal. For the two bans, the transition periods are considered to include the month the ban went into effect and four preceding months. For the ban lift, the transition period is the month of the ban lift and the three following months.

The estimated shift was 1061 ± 302 kg/day. Assuming 1,100,000 population served, the estimated shift, on an annual per capita basis, was 0.352 kg/cap.-yr. This value differs from Pallesen (1985). In that paper, the data record was shorter, only the ban and ban-lift were considered; and these were estimated separately. It was estimated that the ban decreased loading by 0.522 kg/cap-yr and the ban-lift increased loading by 0.193 kg/cap -yr. The average of these two values is 0.358 kg/cap-yr, which agrees reasonably with the updated estimate presented above.

Figure 2 shows the data and the estimated load for Kalamazoo, Michigan. This example, with an upward trend, is an interesting counterpoint to the downward trend of the Milwaukee data. The model was adequate for both data sets. The estimated shift was 0.833 kg/cap-yr.

Saginaw, Michigan, shown in Figure 3, illustrates how missing data were handled by treating them as though they were an intervention with no associated shift. The model forecasts, which are level for a moving average model (e.g. ARIMA (0,1,1)), are used to fill in the missing values. The estimated shift was 0.508 kg/cap-yr.

Figure 4 illustrates the seasonal model with data from Ann Arbor, Michigan. The estimated shift was 0.581 kg/cap-yr.

In each of the Michigan examples, the transition period was from one month before the ban was fully in effect to three months after any phosphate detergents could be sold. All data during the transition period are ignored by the estimation procedure. The Ann Arbor data happen to show a high influent P load for one month during this transition period (actually the first month that the ban was to have been fully in effect). This is part of the seasonal effect that appears to coincide with the annual cycle of the student population in this university city.

CONCLUSIONS

The magnitude of the decrease in influent sewage phosphorus level due to enforcement of bans on phosphate detergents has been estimated for a number of treatment plants. The estimated shift is 0.3 to 0.4 kg/cap-yr.

Estimating the effect of environmental protection laws requires analyzing data that are in the form of time series. Simply taking arithmetic averages for some interval before and after the

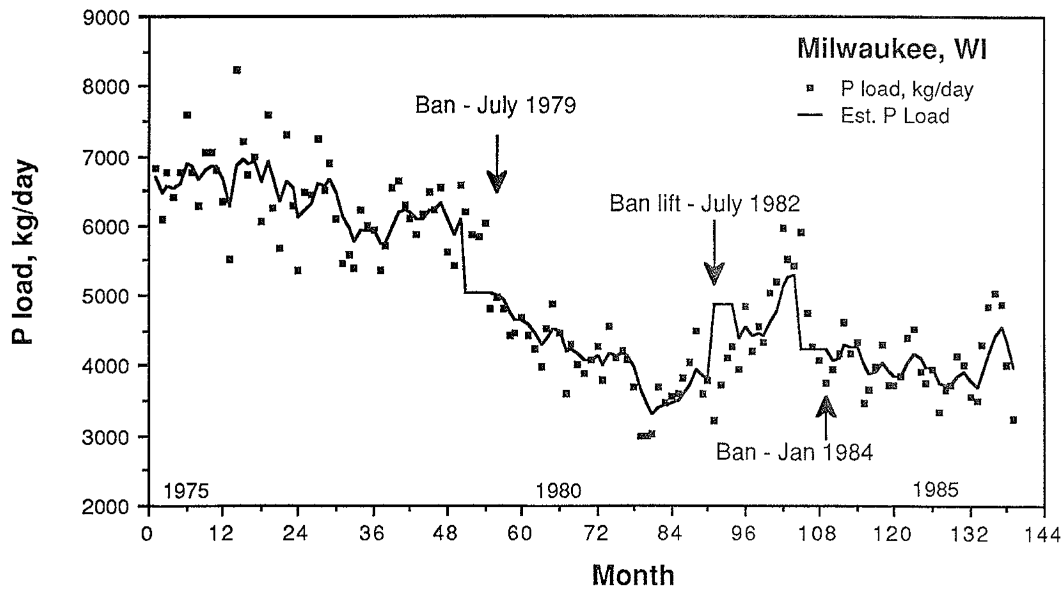

Figure 1. Phosphorus Loading and Forecast for Milwaukee

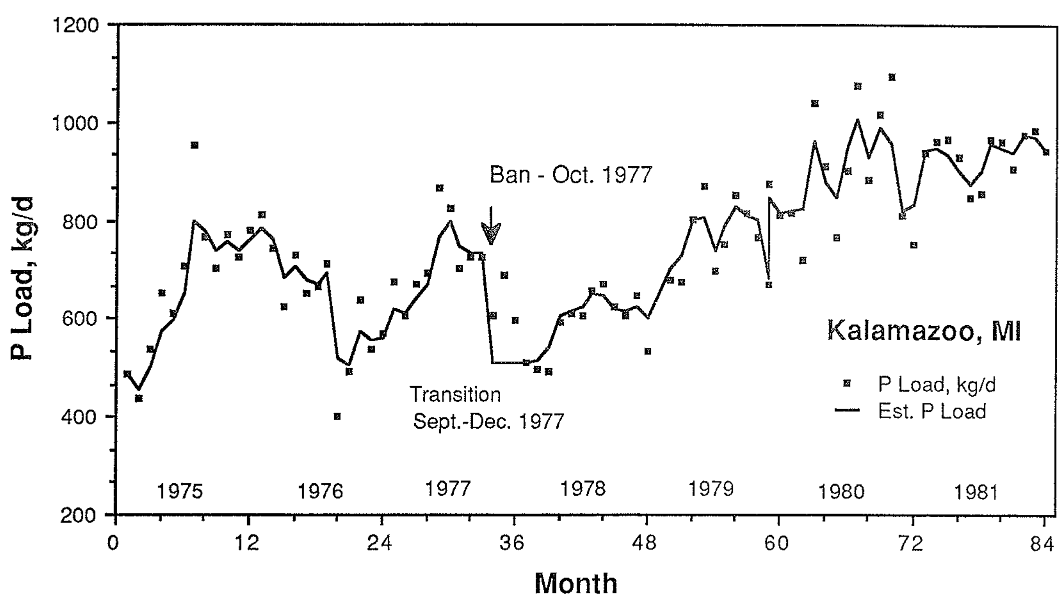

Figure 2. Phosphorus Loading and Forecast for Kalamazoo

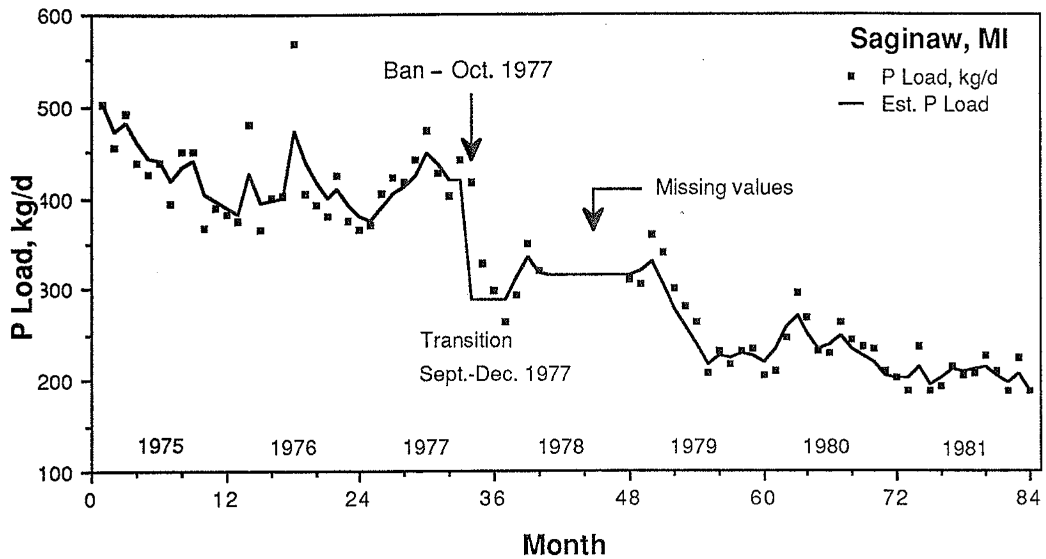

Figure 3. Phosphorus Loading and Forecast for Saginaw

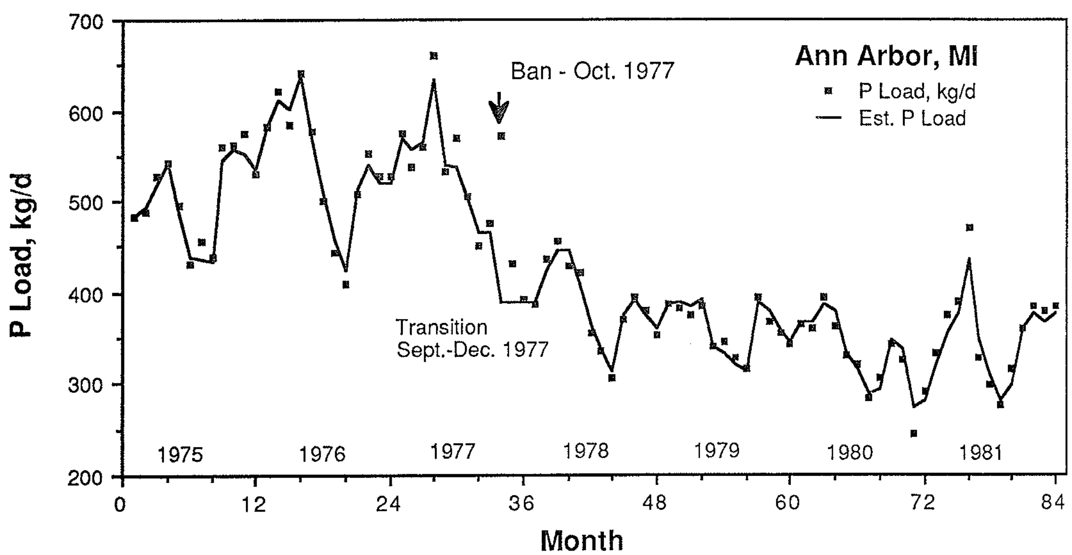

Figure 4. Phosphorus Loading and Forecast for Ann Arbor

intervention is very likely to be wrong since the variation is unlikely to be random about a fixed mean level. It will be biased if there is any trend (and often there is, even if only from random drift). Even if the series happened to be stationary, autocorrelation in the data would distort all probability statements (confidence interval estimates, for example) derived from simple averages.

Data which occur in the form of a time series need to be modeled with an appropriate time series model. Our experience shows that this model does not need to be very complicated.

In the case of the phosphate detergent ban, there was a transition period to fully implement the ban. It was feasible to model the transition itself, so it was treated as a gap in the data record. For a moving average model, which has been used in these examples, the forecasts are level. The uncertainty associated with the estimated effect (the shift in level) increases as the length of the transition gap increases.

A simple ARIMA (0,1,1) time series model, incorporating random walk and instantaneous error effects, has been adequate for more than half of all data sets examined to date. Adding a seasonal component has allowed a number of other data series to be fitted and analyzed.

REFERENCES

Berthouex, P.M., Pallesen, L., Booman, K., and Sedlak, R. (1983). Discussion of: a preliminary assessment of Michigan's phosphorus detergent ban. J. Water Poll. Cont. Fed., 55, 323-326. Errata, 55, 1026.

Booman, K.A. and Sedlak, R.I. (1986). Phosphate detergents - a closer look. J. Water Poll. Cont. Fed., 58, 1092-1100.

Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis, forecasting, and control. Holden-Day, San Francisco.

Box, G.E.P. and Tiao, G.C. (1975). Intervention analysis with applications to economic and environmental problems. J. Amer. Stat. Assoc., 70, 70-79.

Hartig, J.H. and Horvath, F.J. (1982). A preliminary assessment of Michigan's phosphorus detergent ban. J. Water Poll. Cont. Fed., 54, 193-197.

Jones, Edgar R. and Hubbard, Susan D. (1986). Maryland's phosphate ban - history and early results. J. Water Poll. Cont. Fed., 58, 816-822.

Pallesen, L., Berthouex, P.M., and Booman, K.A. (1985). Environmental intervention Analysis: Wisconsin's ban on phosphate detergents. Water Research, 19, 353-362.