Second Conference on
Control Engineering, 1982,
Newcastle, 25-27 August

Control of Water Usage: Evaluation of Strategies

M.N. VISWANATHAN
Engineer, Hunter District Water Board, Newcastle
and
B.D.O. ANDERSON
Professor, Department of Systems Engineering, Australian National University, ACT

SUMMARY A methodology is given for modelling water consumptions using temperature and rainfall data as inputs. A model is developed for consumption in the Hunter Valley and is used to evaluate the effectiveness of usage restrictions imposed during a prolonged drought.

1 INTRODUCTION

Following a severe drought period which commenced in mid 1979, storage levels in the Hunter Valley water supply system fell to such a level that the Hunter District Water Board imposed usage restrictions. The restrictions have now lasted several years, and from time to time their nature has been varied.

The Board wanted to estimate the effectiveness of the restrictions i.e., the amount of water saved. In control system terms, the Board wished to know the response resulting from the controls, so that on future occasions when control was deemed necessary, there would be information allowing the appropriate selection of controls. Because there are very major day to day fluctuations in water usage with temperature and rainfall, and underlying seasonal fluctuations together with changes caused by a varying population, some form of data processing, as opposed to, for example, a single averaging over many days of consumption in non-restriction and restriction periods, was necessary to estimate the savings. This paper reports the development of this processing.

To determine the effectiveness of restrictions on water consumption levels, a model is first developed that predicts water consumption under no restriction conditions. (The form of this model is sufficiently general that the model would certainly serve as a basis for a study of many other water supply modelling problems.) In relation to the Hunter Valley problem, the model is used to predict the consumption levels which would have resulted had no restriction been imposed. This is done for an eighteen month period over which time restrictions were in force. By comparing the predicted consumptions under the no restriction assumption with the actual consumptions, one can then measure the effect of the restrictions on consumption levels.

The development of water consumption models is not new, see example (Weeks and McMahon 1973 a, b, 1974 and Danielson, 1979). These references predict either monthly or daily consumption levels for individual households using a simple linear regression analysis relating consumption with climatic and economic variables. Reports of studies of a situation like that in the Hunter Valley could however not be found.

The major determinants of water consumption are well known to be rainfall and temperature. Therefore, a model is sought with these quantities as the main inputs, which are also readily available. As described in more detail below, a recursive least squares method can be used to estimate the coefficients in a linear model; using data from the prerestriction period. The model is also expanded, as further described below, to reflect seasonal variations and the underlying population growth trend.

2 LINEAR MODELS

Let x_i, T_i and R_i denote the consumption, maximum temperature and rainfall on day i. With $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ as polynomials and Z^{-1} denoting a backward shift operator i.e., $Z^{-1}x_i = x_{-1}$, one might postulate a model of the form.

$$A(Z^{-1})\ x_i = B(Z^{-1})\ T_i + C(Z^{-1})\ R_i \qquad (1)$$

At once, several variations to this model need to be made. First, since the model will eventually be used to predict nature of x_i when the true values of past x_i are not available (i.e., to estimate what consumption would have been had there been no restriction in force), it is desirable that $A(Z^{-1})$ just be a constant - otherwise x_i would depend on x_{i-1}, x_{i-2}, ... Second, one must postulate the occurance of errors, stemming from the fact that the model can never provide a perfect fit with reality. Third, even if on one day the rainfall were zero and temperature were zero, some water would be used; thus a base level value of x_i should be postulated.

These three observations imply that we should consider a model of the form.

$$x_i = d_0 + B(Z^{-1})\ T_i + C(Z^{-1})\ R_i + e_i \qquad (2)$$

Where d_0 denotes the base level consumption and e_i is a zero mean random sequence which, because of lack of further knowledge, is assumed white:

$$E\ [e_i\ e_j] = \sigma^2\ \delta_{ij} \qquad (3)$$

The coefficients in $B(\cdot)$ and $C(\cdot)$ and the value of σ^2 all need to be estimated.

3 PARAMETER ESTIMATION

The estimation of the coefficients $B(\cdot)$ and $C(\cdot)$ is straight forward. A least squares estimate

(Isermann, 1974; Box and Jenkins, 1970) is sought postulate that $B(-)$ and $C(-)$ have degrees n and m respectively.

Define

$$z_i = [T_i, T_{i-1} \dots T_{i-n}, R_i \, R_{i-1} \dots R_{i-n}]^T \tag{4}$$

and $a = [b_0 \, b_1 \dots b_n, c_0 \, c_1 \dots c_m]$

$$\text{where } B(z^{-1}) = \sum_{i=0}^{n} b_i z^{-i} \tag{5}$$

$$C(z^{-1}) = \sum_{i=0}^{m} c_i z^{-i} \tag{6}$$

Assuming that values of T_i, R_i and x_i are known for $i = 1, 2 \dots K$, we estimate "a" by

$$\hat{a} = \arg_a \text{ in } \sum_{1}^{K} [z_i^T a - x_i]^2 \tag{7}$$

This gives rise to

$$\hat{a} = [\sum_{1}^{K} z_i z_i^T]^{-1} [\sum_{1}^{K} z_i x_i] \tag{8}$$

This quantity can be obtained recursively, as is well known; define

$$P_k^{-1} = P_{k-1}^{-1} + z_k z_k^T$$

$$P_o^{-1} = 0 \tag{9a}$$

$$b_k = b_{k-1} + z_k x_k \tag{9b}$$

Then

$$\hat{a} = P_K b_K \tag{9c}$$

In order to ensure that P_k^{-1} is nonsingular, and thus P_k exists for all k, one can take P_o^{-1} as a small positive matrix. But even if P_o^{-1} is zero, in a properly posed problem, P_k^{-1} will be nonsingular for all sufficiently large k. Then the following alternative recursive equations can be used.

$$P_k = P_{k-1} - P_{k-1} z_k [1 + z_k^T P_{k-1} z_k] z_k^T P_{k-1} \tag{10a}$$

$$\hat{a}^k = \hat{a}^{k-1} - P_k [z_k z_k^T \hat{a}^{k-1} - z_k x_k] \tag{10b}$$

$$\text{with } \hat{a} = \hat{a}^k \tag{10c}$$

(The superscript is used on $\hat{a}$ to denote a time index, while the subscript is reserved to denote a particular entry of the $\hat{a}$ vector.)

4 MODELLING HUNTER VALLEY USAGE

As indicated in Section 3, integer values for the degrees of $B(-)$ and $C(-)$ must be assumed. Generally speaking, one wants as few coefficients as possible, consistent with having a good fit for the model, goodness of fit being measured by:

$$J = \sum_{1}^{k} [z_i^T a - x_i]^2 \tag{11}$$

After several trials, it was found that degree of 1 and 0 for $B(-)$ and $C(-)$ sufficient, so that the model took the form

$$x_i = d_0 + b_0 T_i + b_1 T_{i-1} + C_0 R_i + e_i \tag{12}$$

However an examination of the daily prediction errors $z_i^T \hat{a} - x_i$ resulting with this equation showed a pronounced sinusoidal variation of typical errors over the year. Consequently (12) was further modified to take into account the seasonal effects and also an annual growth rate of 2 per cent in consumption. This figure represents a long term trend in yearly consumption figures that has been observed over many years and can be attributed to population trend. The modified equation is

$$x_i = \left\{ [a_1 + a_2 \cos \frac{2\pi D}{365}] T_{i-1} + [a_3 + a_4 \cos \frac{2\pi D}{365}] T_i + [a_5 + a_6 \cos \frac{2\pi D}{365}] R_i + [a_7 + a_8 \cos \frac{2\pi D}{365}] \right\} [1 + \frac{0.02D}{365}] + e_i \tag{13}$$

Here D denotes the day of the year, i.e., D = 1 corresponds to January 1 (By contrast i = 0 corresponds to the first day for which data is used to estimate the coefficients of the model, and this may not be January 1). One could criticise the above arrangement on the grounds that it implies seasonal peaks must occur on January 1 and July 1, which is obviously only approximately true. The additional complexity in the model which would be needed to allow seasonal peaks at other times was not however deemed warranted.

Least squares estimates of the "a" were obtained by using consumptions, maximum temperature and rainfall data for almost all of 1979 (during December, restrictions were imposed). The resulting estimates were

$$\hat{a}_1 = 0.00 \qquad \hat{a}_5 = -1.08$$

$$\hat{a}_2 = 1.50 \qquad \hat{a}_6 = -0.51$$

$$\hat{a}_3 = 3.01 \qquad \hat{a}_7 = 179.5$$

$$\hat{a}_4 = 0.88 \qquad \hat{a}_8 = -14.15 \tag{14}$$

As one would expect, the signs of the $\hat{a}_i$ indicate that consumption increases with higher maximum temperatures and lower rainfall.

Figure 1 shows temperature and rainfall levels for 1979 and Figure 2 shows actual consumption and predicted consumption (i.e., right hand side of (13) with a_i replaced by $\hat{a}_i$ and $e_i = 0$).

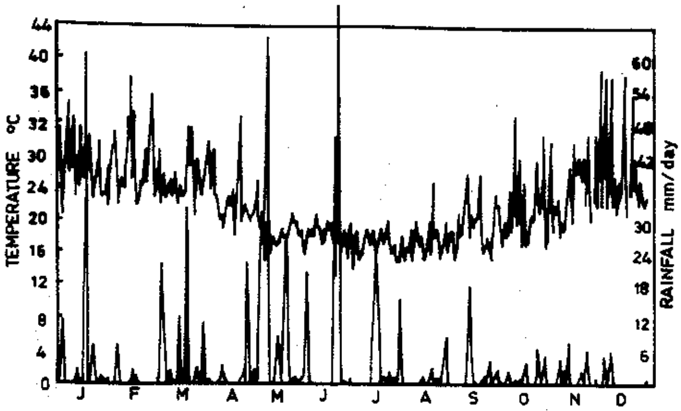

Figure 1. Rainfall & Temperature Variation (1979)

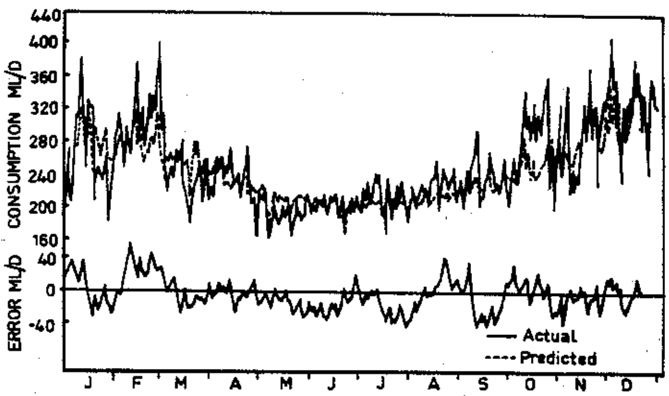

Figure 2. Actual and Predicted Consumption (1979)

5 QUALITY OF THE MODEL

One of the assumptions made in the analysis is that the errors are uncorrelated. This assumption can be examined by determining the auto-correlation function of the errors, which is defined as

Auto Correlation at lag "l":

$$\rho_l = \frac{E[(e_i - \mu)(e_{i+l} - \mu)]}{\sqrt{E[(e_i - \mu)^2]\ E\ [(e_{i+l} - \mu)^2]}} \qquad (15)$$

The sample auto-correlation function of errors as a function of the lag "l" is shown in Figure 3.

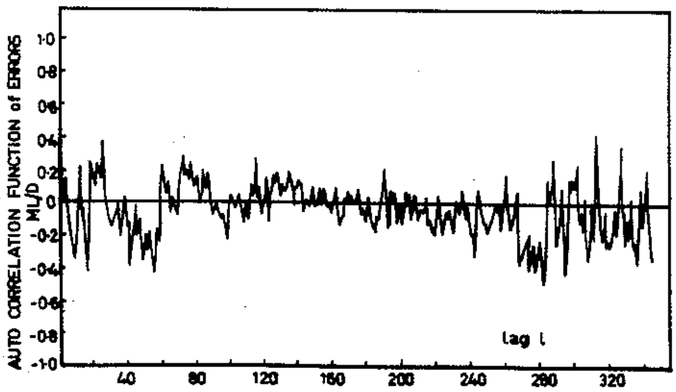

Figure 3. Auto Correlation Function of Errors (1979)

The distribution of the auto-correlation function of the errors shows that the errors are reasonably uncorrelated. As is also reasonable the auto-correlation for small lags tend to be bigger in magnitude than for intermediate lags, and for very large lags (in excess of 11 months) it tends to fluctuate substantially, presumably because of the small sample size available for computing the coefficients.

The prediction of savings in times of water restrictions depends on an important assumption: that a model of usage derived using the data for the year 1979 when no restrictions existed would have correctly predicted consumption in 1980 and 1981 if no restrictions had been in force. This assumption was checked by predicting consumptions for the years 1977 and 1978 using 1979 parameters. The predicted consumptions for the years 1977 and 1978 were in reasonable agreement with that of the actual consumptions.

A more sophisticated check was also carried out. A single model was found for the two years 1977 and 1979, and separate models for each of the years were also obtained; then the squares of the residuals value of J, see (11) were found for the three cases. The ideal situation would be one where the values of $J_{1977\,+\,1979}$ was identical with $J_{1977} + J_{1979}$; in general however, the first quantity will be larger. If it is not much larger, it is reasonable to postulate that a model for 1977 works for 1979. This type of qualitative reasoning can be made more quantitative using F distributions, as described in Viswanathan, 1982 . The conclusion, also borne out by examining parameter estimates, is that the yearly models are different but not very much so. Accordingly the model derived using 1979 could be used to predict consumption during the post 1979 restriction period.

7 ESTIMATION OF CONFIDENCE INTERVALS

Differences between predicted consumption and actual consumption arise for two reasons: in general, $\hat{a}_i \neq a_i$ and $e_i \neq 0$. The variance of the prediction on day k, assuming the coefficients have been estimated using data over day 1 through day K is given by (see Appendix 1)

$$\sigma_k^{\;2} = \sigma^2\ [1 - z_k^{\;T}(\sum_1^K z_i\ z_i^{\;T})\ z_K] \qquad (16)$$

with σ^2 being estimated by

$$\hat{\sigma}^2 = \frac{1}{K - p} \sum_1^K \text{(Prediction errors)}^2 \qquad (17)$$

Here, P is the number of parameters, Viz 8, in the model.

Using the variance $\sigma_k^{\;2}$ a confidence interval for the prediction can be determined. Thus if the errors can be assumed Gaussian (a major assumption in itself), the confidence interval for a 95% confidence level is

$\hat{x}_k \pm 1.96\ \sigma_k$

Where $\hat{x}_k$ is the prediction replace x_i by $\hat{x}_i$, a_i by $\hat{a}_i$ and e_i by zero in (13)

The standard deviation σ_k of the prediction error obtain on a daily basis for a confidence level of 95 per cent is of the order of 25 per cent of the prediction. However, in the present analysis it is more relevant to consider the standard deviation of the prediction error over a longer period, typically that during which the various restrictions were in force.

For example the standard deviation of the prediction error over a 60 day period:

$$\sigma\ (60\ \text{day}) = 1/60\sqrt{\sigma_1^{\,2} + \ldots + \sigma_{60}^{\,2}} \qquad (18)$$

Approximately $\sigma(60)$ is less than σ obtained on a daily basis by approximately $1/\sqrt{60}$.

The confidence interval for the predicted consumption levels for the year 1979 for a 60 day period and for a confidence level of 95 per cent is 14,400 ± 490 megalitres, representing about 3.5% deviation around the estimates.

8 EFFECT OF RESTRICTIONS

Using coefficient estimates obtained using 1979 no restriction data and allowing for 2% per annum growth, the water consumption under no restriction conditions was predicted for the years 1980 and 1981. *Figures 4(a) and 4(b) show* the temperature and rainfall variations during the year 1980 and part of 1981. Figure 5 shows the predicted and actual consumptions for these periods. Table I summarises the effect of various restrictions on water consumption levels.

The following observations are made with regard to the restrictions and their effect on the consumption levels.

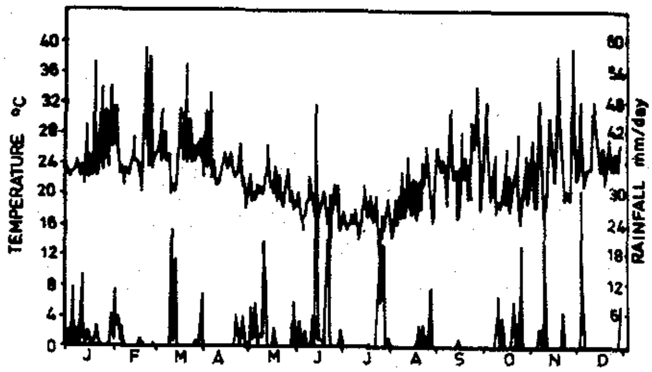

Figure 4a. Temperature and Rainfall (1980)

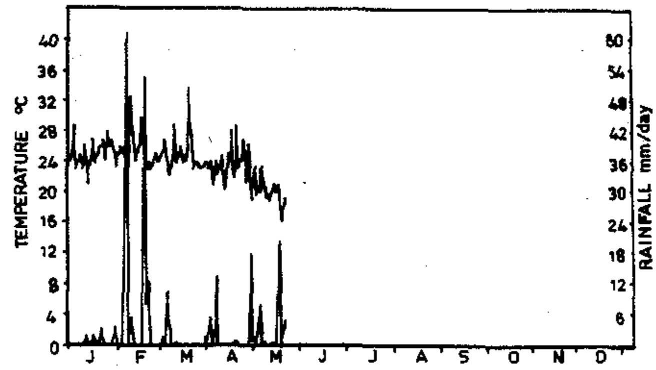

Figure 4b. Temperature and Rainfall (1981)

TABLE I

EFFECT OF WATER RESTRICTIONS

Restrictions	Date Commenced	Nature of Restrictions	Actual Consumption ML	Predicted Consumption ML	% Saving
R1	6-12-1979	Fixed sprinklers on alternate days only	24,713	25,388	2.7
R2	29-2-1980	Total ban on fixed sprinklers	16,770	17,824	5.9
R3	9-5-1980	R2 and hand held hoses 5 hours a day	27,501	29,232	5.9
R4	19-9-1980	R2 and hand held hoses 3 hours a day	23,372	30,767	24.0
R5	5-1-1980	R2 and hand held hoses 2 hours a day - 4 days a week	7,293	10,586	31.1
R6	9-2-1981	R2 and buckets only	-	-	-
R7	16-2-1981	R2 and hand held hoses 2 hours a day - 4 days a week	6,603	10,333	36.1
R8	17-3-1981 (upto 22nd May 1981)	R2 and hand held hoses 12 hours a day - alternate days	12,373	16,292	24.1

i. During the period of 18 months, 6 different types of restrictions were imposed. Restrictions R1 to R3 ran from 6 December 1979 to 19 September 1980 and were least effective. The percentage savings varied between 2.7 and 5.9. Restrictions R4 to R8 were in force between 19 September 1980 and 22 May 1981, and these were most effective, with savings varying between 24.0 per cent and 36.1 per cent.

ii. During the periods of restrictions R4 to R8 although the actual mean consumption levels were considerably less than the predicted nil restriction consumption levels, the daily fluctuations in consumption followed the temperature variations. In fact this trend is evident even during the winter months i.e., during the period of restriction R3.

iii. During the winter months, the difference between the actual and estimated predicted consumption is minimal. This is an expected trend as all the restrictions imposed have aimed at restricting the water used in gardens, and this usage is minimal during the winter periods. This trend is also evident during the period of R8 where the difference between the predicted and actual consumption is closing towards the month of May 1981.

iv. The Hunter District Water Board undertook an extensive campaign towards January 1981, urging the public to reduce water consumption. The effect of this campaign became evident during the period of restrictions R7 and R8 which had the actual mean consumption of 198 and 187 ML, the lowest levels to be achieved during the entire period during which the restrictions were in force.

9 CONCLUSIONS

1. A model has been developed to predict the water consumption levels in the absence of restrictions for the Hunter Valley area. The modelling methodology is however general, and given the past and present temperatures and rainfalls, allowing for seasonal fluctuation. The model is effective in making long term predictions.

2. The model developed was based on the method called "Recursive Least Squares". The model's ability to predict current consumptions was checked using 1977,78 and 79 data of temperatures and rainfalls. Using 1979 data, when compared with actual consumption it was observed that at 95 per cent confidence level the maximum error was of the order of 3.5 per cent for a 60 day period.

3. Using the year 1979 data i.e., temperature and rainfall, the model parameters were determined. Assuming that the model and its parameters are invariant between the years 1979,1980 and 1981 except for an annual growth of 2 per cent in consumption, the consumptions for the year 1980 and 1981 (upto May 22nd) under no restriction conditions were obtained. These compared with the actual consumptions indicate the effect of restrictions on consumption levels.

4. Between December 1979 and May 1981 a total of 6 different types of restrictions R1...R8 were imposed differing in various levels of severity. The restrictions R1,R2 and R3 had very little effect on the consumption levels whereas restrictions R4...R8 had significant effect. The saving in consumption was of the order of 5 per cent for the period R1...R3 and during R4...R8 was approximately 27 per cent. Knowledge of these savings provides a basis for implementing strategies to control water usage.

5. In winter months, the restrictions had very little effect on consumption levels. The restrictions were most effective during summer months. In spite of restrictions the daily temperature variations still influenced the daily consumption levels.

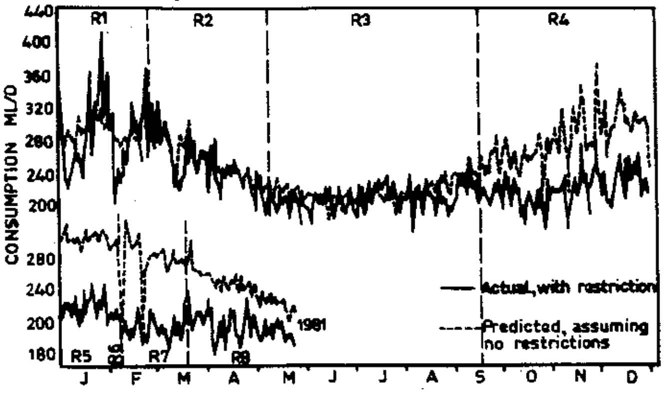

Figure 5. Actual and Predicted Consumption (1980-1981)

10 REFERENCES

BOX, G.E.P. and JENKINS, G.M. (1970). Time Series Analysis, Forecasting and Control, San Franciso, Holden Day.

DANIELSON, L.E. (1979). An analysis of residential demand for water using micro time series data. Water Resources Research, Vol. 15, No. 5, pp 763-767.

ISERMANN, R. (1974). Prozess-identifikation, Berlin, Springer Verlag.

VISWANATHAN, M.N. (1982). Determination of Effectiveness of Water Consumption Estimations: Methodology and Application. Thesis (M.E.) University of Newcastle.

WEEKS, C.R. and McMAHON, T.A. (1973a). A comparison of urban water use in Australia and the U.S. Journal AWWA, Vol. XXII, pp 232-236.

WEEKS, C.R. and McMAHON, T.A. (1973b). Climate and water use in Australian cities. Aust. Geographical Stud., Vol. 11, pp 99-104.

WEEKS, C.R. and McMAHON, T.A. (1974). Urban water use in Australia. Civ. Engg. Trans., I.E. Aust., Vol. CE 16, No. 1, pp 58-63.

APPENDIX I: VARIANCE OF PREDICTION ERROR

Let $\Delta a = a - \hat{a}$ and $\Delta x_k = x_k - \hat{x}_k$. Then

$$\Delta x_k = z_k^T a + e_k - z_k^T \hat{a}_k$$

$$= z_k^T \Delta a + e_k \qquad \text{(A1)}$$

Also, $\hat{a} = [\sum_{k=1}^{K} z_k z_k^T]^{-1} [\sum_{k=1}^{K} z_k x_k]$

$$= \sum_{k=1}^{K} [z_k z_k^T]^{-1} [\sum_{k=1}^{K} z_k z_k^T a + \sum_{k=1}^{K} z_k e_k]$$

so that

$$\Delta a = -[\sum_{k=1}^{K} z_k z_k^T]^{-1} [\sum_{k=1}^{K} z_k e_k] \qquad \text{(A2)}$$

Using (A1),(A2) and the whiteness of e_k, one obtains

$$E[\Delta x_k^2] = \sigma^2 [1 - z_k^T \sum_{1}^{K} (z_k z_k^T)^{-1} z_k] \qquad \text{(A3)}$$