

# Realistic Forecasting of Groundwater Level, Based on the Eigenstructure of Aquifer Dynamics

V. J. Bidwell

Lincoln Environmental, Lincoln Ventures Ltd, Lincoln, New Zealand (bidwellv@lincoln.ac.nz)

**Abstract:** Short-term management of groundwater resources, especially during droughts, can be assisted by forecasts of groundwater levels. Such forecasts need to account for the natural dynamic behaviour of the aquifer, likely recharge scenarios, and recent but unknown abstractions. These requirements mean that forecasts, at say monthly intervals, need to be updated with current observations on a real-time basis. One established procedure for this kind of problem is to fit autoregressive, moving-average, exogenous-variable (ARMAX) time-series models to the history of groundwater levels in response to estimates of land surface recharge. The ARMAX difference equations are then converted into forecast equations that allow real-time updating to include recent forecast errors as an additional source of information. Some disadvantages of this pure time-series analysis approach are the apparent lack of physical concepts in the model formulation and statistical aspects of model identification and calibration that are related to the inherent structure of ARMAX equations. This paper addresses these issues by describing a method for formulating ARMAX forecast equations from a linear system description based on the eigenvalues and eigenvectors (eigenstructure) of the dynamic behaviour of an aquifer. For the piezometric response of a heterogeneous aquifer to a fixed spatial distribution of land surface recharge, with time-varying magnitude, only a few eigenvalues are significant for describing the dynamics. The resulting model has a simple robust parameter structure, and is easily calibrated and implemented in spreadsheet form. The eigenstructure approach enables transfer of some parameter information from locations with good data records to those with sparse data. This modelling approach is demonstrated with monthly values of land surface recharge, estimated from a daily water balance model, and groundwater level data from an observation well in a 2000 km<sup>2</sup> alluvial aquifer in Canterbury, New Zealand.

**Keywords:** *Groundwater forecasting; Aquifer dynamics; Eigenvalue; ARMAX*

## 1 INTRODUCTION

Groundwater storage in aquifers may be considered, in the natural state, as the dynamic balance between recharge, driven by climatic processes, and discharge to surface waters. Most of the dynamic behaviour is caused by variations in recharge through the land surface rather than from rivers. This balance is modified by groundwater abstraction for human use. During periods of climatic drought, land surface recharge is zero, and groundwater levels decline at a rate determined by natural storage-discharge dynamics and the effect of, usually, increased abstractions for purposes such as irrigation. Management of the resource under these conditions can be assisted by short-term forecasts of groundwater levels (Ahn, 2000).

Forecasting of the uncertainties and dynamics of many economic and natural processes is based on the mathematics of time-series analysis (e.g., Box and Jenkins, 1970). The values of observed states

are considered either as univariate stochastic processes driven by random inputs, or as bivariate processes that include other observed inputs. This essentially “black box” approach has been applied to groundwater fluctuations by Tankersley and Graham (1993), in both univariate form and bivariate form with rainfall as the input variable. Ahn (2000) used Kalman filter updating of a time-series model for groundwater levels at multiple sites in a multi-layered aquifer system.

Time-series analysis can also be considered as a statistical approach to identifying linearised descriptions of physical processes (e.g., Young, 1984). Bidwell et al. (1991) applied the method of Young (1984) to a simplified conceptual model of groundwater level variations in response to recharge estimated as soil-water drainage from a water balance model.

Most mathematical descriptions of dynamic groundwater behaviour are in the form of linearised partial differential equations. This spatially-distributed linear process can also be

represented as a linear system of interconnected discrete-space components, such as finite difference or finite element schemes, or as a system of conceptual linear components based on the eigenvalue solution to the distributed process (Sahuquillo, 1983). Application of time-series analysis to this multi-component system description of groundwater dynamics offers the potential for completely linking aquifer properties, boundary conditions, recharge processes, and external stresses to the stochastic forecasting equations. This paper describes an approach that matches the stochastic difference equation models of time-series analysis to the physically-based, linear system, groundwater model of Sahuquillo (1983).

## 2 TIME-SERIES MODELS

The dynamic behaviour of many natural systems can be modelled in terms of stochastic linear difference equations, for situations where observations about the system are available at regular time intervals. The general structure of a difference equation that describes the dynamic relationship between an input time series  $X_n$  and an output series  $Y_n$  is:

$$Y_n = a_1 Y_{n-1} + a_2 Y_{n-2} + \Lambda + a_p Y_{n-p} + b_1 X_n + b_2 X_{n-1} + \Lambda + b_q X_{n-q+1} \quad (1)$$

Equation (1) relates the output series to past output values (autoregressive) and a “moving average” of present and past input values. This autoregressive, moving-average (ARMA) structure requires fewer model parameters than the equivalent purely AR or MA form. By means of the  $z$ -operator (discrete-time equivalent of the Laplace operator) in its time-shift form  $z^{-m}Y_n \equiv Y_{n-m}$ , (1) can be expressed as the transform:

$$Y_n = \frac{(b_1 + b_2 z^{-1} + \Lambda + b_q z^{-q+1})}{(1 - a_1 z^{-1} - a_2 z^{-2} - \Lambda - a_p z^{-p})} X_n = \frac{B(z^{-1})}{A(z^{-1})} X_n \quad (2)$$

Traditional time-series analysis (Box and Jenkins, 1970) considers a univariate stochastic series, after removal of trends and seasonal effects, as the output  $N_n$  from an ARMA model (2) with uncorrelated noise  $e_n$  as input.

### 2.1 ARMAX equations

When the stochastic difference equation includes the exogenous (X) input  $X_n$  as well as the stochastic term  $N_n$  then the model structure is

referred to as an ARMAX equation, with general form:

$$Y_n = \frac{B(z^{-1})}{A(z^{-1})} X_n + \frac{D(z^{-1})}{C(z^{-1})} e_n \quad (3)$$

This model (3) is also referred to as a transfer-function-noise (TFN) description, when the emphasis is more on identifying a system in the presence of noise rather than simulating the entire statistical structure.

### 2.2 Model identification and calibration

Identification of (3) refers to the number of terms in the polynomials  $A$ ,  $B$ ,  $C$ ,  $D$  in  $z^{-1}$ , and calibration refers to the values of their coefficients, such as the  $a_i$  and  $b_i$  in (1). The values of the observations  $X_n$ ,  $Y_n$  are assumed to be deviations from equilibrium values, such as averages or zero-input levels. However, in some situations these datum values may be unknown and therefore become part of the parameter set.

Superficially, these difference equations would appear to be amenable to least-squares regression, but this is precluded by the nature of the noise and interdependence among lagged variables, which cause bias in parameter estimation. Two approaches to identification and calibration are provided, respectively, by Box and Jenkins (1970) and Young (1984).

### 2.3 Forecast equation

The stochastic linear dynamic model (3) can be rewritten in the form:

$$A(z^{-1})C(z^{-1})Y_n = C(z^{-1})B(z^{-1})X_n + A(z^{-1})D(z^{-1})e_n \quad (4)$$

Equation (4) includes present and past values of  $e_n$ . These are also the forecast errors at one time step ahead, given that the input  $X_n$  is known. Inclusion of the past values of forecast error provides a self-correction component that can take account of the dynamic effect of unobserved influences.

## 3 GROUNDWATER MODEL

Time-series analysis is a “black box”, external approach to modelling an assumed linear dynamic system that links the observed input and output series. The following sections describe how the structure of a linear systems model of piezometric response to groundwater recharge can be synthesised from an understanding of groundwater flow and storage in aquifers. The

resulting model can then be expressed in forecast equation form (4), with some advantages in identification and calibration relative to the time-series approach.

### 3.1 Eigenstructure approach

Sahuquillo (1983) shows how the dynamic behaviour of piezometric head at any location in an aquifer can be modelled in terms of the eigenstructure of the solution to the groundwater flow equation. The eigenstructure, comprising the eigenvalues and eigenvectors, is a means of representing the distributed linear aquifer system as an infinite series of first-order water storages arranged in a parallel system structure. For the particular case of piezometric response to a fixed spatial pattern  $P(x,y)$  of land surface recharge, with time-varying magnitude  $R(t)$ , only a few of these water storage components are required for acceptable model accuracy. These water storages are conceptual and do not exist in any physically definable form.

### 3.2 Linear system model

Figure 1 shows the parallel system structure representing aquifer dynamics, and an additional component in series for modelling the dynamic effect of transport through the vadose zone and perching of groundwater above aquitards.

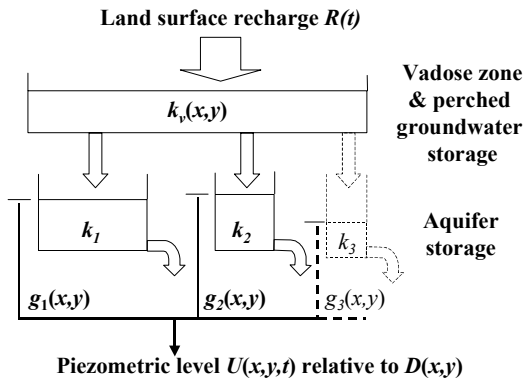


Figure 1. Eigenstructure conceptual model.

The eigenvalues  $k_i$  are the same at all locations, and can be considered as the discharge coefficients of the conceptual water storages, for which the response of the water content  $W_i(t)$  to recharge  $R(t)$  can be described by the solution of a first-order differential equation over a time interval  $\Delta t$ :

$$W_i(t) = W_i(t - \Delta t) \exp(-k_i \Delta t) + R(t) [1 - \exp(-k_i \Delta t)] \quad (5)$$

The eigenvalues are functions of transmissivity, storativity, aquifer geometry, and boundary conditions, but these need not be determined in the present application. The corresponding eigenvector coefficients  $g_i(x,y)$  define the contribution  $U_i(t)$  of each conceptual water storage to the dynamic piezometric effect  $U(x,y,t)$  relative to the steady piezometric effect  $D(x,y)$  at location  $(x,y)$ , so that:

$$U_i(t) = g_i(x,y) W_i(t) \\ U(x,y,t) = \sum_i U_i(t) \quad (6)$$

The transfer function of  $R(t)$  to  $U_i(t)$  by each water storage contribution, in terms of the discrete-time samples  $U_n^i$  and  $R_n$ , can be expressed in ARMA form (2) as:

$$U_n^i = \frac{\beta_i}{(1 - \alpha_i z^{-1})} R_n \quad (7)$$

for which:

$$\alpha_i = \exp(-k_i \Delta t) \\ \beta_i = g_i(x,y) (1 - \alpha_i) \quad (8)$$

The relationship (8) also holds for the vadose zone storage, but  $k_v(x,y)$  varies with location and the gain coefficient is unity, so that  $\beta_v = (1 - \alpha_v)$ .

The piezometric response  $H(x,y,t)$ , at a particular location, to the input series of land surface recharge  $R(t)$  over the whole aquifer is usually observed relative to an arbitrary datum. However, the dynamic response  $U(x,y,t)$  of the model in Figure 1 is superimposed on the assumed steady piezometric effect of other recharges (e.g., from rivers) and abstractions. This steady effect  $D(x,y)$  is an additional parameter at each location. For convenience, the observed piezometric time series  $H_n(x,y)$  at a particular location is transformed to:

$$U_n = H_n(x,y) - D(x,y) \quad (9)$$

### 3.3 ARMAX and forecast equations

The transfer-function-noise description of the system structure (Figure 1), with only the first two eigenvalues, is:

$$U_n = \frac{\beta_v}{(1 - \alpha_v z^{-1})} \times \left[ \frac{\beta_1}{(1 - \alpha_1 z^{-1})} + \frac{\beta_2}{(1 - \alpha_2 z^{-1})} \right] R_n + N_n \quad (10)$$

When (10) is multiplied out into the form of (2):

$$U_n = \frac{(b_1 + b_2 z^{-1})}{(1 - a_1 z^{-1} - a_2 z^{-2} - a_3 z^{-3})} R_n + N_n \quad (11)$$

The relationships between the coefficients of (10) and (11) are obtained by equating powers of  $z^{-1}$ .

The noise component  $N_n$  can be described adequately by a first-order structure, for many practical applications, and thus the complete ARMAX model becomes:

$$U_n = \frac{(b_1 + b_2 z^{-1})}{(1 - a_1 z^{-1} - a_2 z^{-2} - a_3 z^{-3})} R_n + \frac{1}{(1 - f_1 z^{-1})} e_n \quad (12)$$

When (12) is multiplied out to the form of (4) and transformed to the time domain, the resulting forecast equation is:

$$U_n = c_1 u_{n-1} + c_2 u_{n-2} + c_3 u_{n-3} + c_4 u_{n-4} + d_1 R_n + d_2 R_{n-1} + d_3 R_{n-2} - a_1 e_{n-1} - a_2 e_{n-2} - a_3 e_{n-3} + e_n \quad (13)$$

for which  $c_i, d_i$  are obtained from  $a_i, b_i, f_i$  in (12) by equating powers of  $z^{-1}$ .

The forecast equation (13) shows how, for this example, the three previous values of forecast error contribute to making a new forecast one time step ahead.

### 3.4 Model identification and calibration

Equations (9) – (13) are implemented in an Excel spreadsheet, as difference equation formulas, together with the relationships (not shown in this paper) between the coefficients of (10), (11), (12) and (13). Parameter estimation is conducted by means of the “solver” optimisation tool in Excel. The objective function is minimisation of the sums of squares of the forecast errors  $e_n$ .

The parameter search is conducted on  $D(x,y)$  and the coefficients  $\alpha_i, \beta_i, \alpha_i, \beta_i$  of (10), rather than the coefficients of (13). The reasons for this approach are, firstly, the former parameters are structurally independent and, secondly, the setting of initial values and constraints is easier. All the parameters are set to an initial value of zero, in the absence of any prior knowledge. The set of parameter constraints is:

$$0 \leq \alpha_v, \alpha_i < 1 \\ \beta_v, \beta_i \geq 0 \quad (14)$$

Because the eigenvalues are, theoretically, the same at all locations in the aquifer, values of  $\alpha_i$

calibrated at locations with good data may be transferred to locations with less data. In practice, this is especially true of  $\alpha_1$ , corresponding to the smallest eigenvalue  $k_1$ .

Optimisation is conducted on subsets of the parameters to facilitate stable convergence to final values, in the following order:

$$D(x,y) \\ D(x,y), \alpha_1, \beta_1 \\ D(x,y), \alpha_1, \beta_1, \alpha_2, \beta_2 \\ \Lambda \quad \Lambda \quad \Lambda \quad \Lambda \quad \Lambda$$

The best of the above, plus  $\alpha_v$ ,

Then add the noise parameter  $f_1$

## 4 APPLICATION

### 4.1 Aquifer description

The Central Canterbury Plains, in the South Island of New Zealand, overlay an interconnected system of unconfined and semi-confined aquifers in silt-sand-gravel alluvium up to about 500 m thick. This 2000 km<sup>2</sup> region is approximately 50 km long from mountains to the sea and 40 km wide between two large braided rivers that cross the plains from mountains to sea. These rivers are perched above the aquifers over much of their length and are a significant source of recharge.

The other significant source is land surface recharge under a mixture of land uses including irrigated and dryland pasture, crop farming, and forest.

Theoretical results from aquifer dynamics indicate that variations in piezometric head are caused primarily by land surface recharge, and that these are superimposed on a steady piezometric surface caused by river recharge and the effect of average groundwater abstraction. Therefore, land surface recharge is the only exogenous variable in the analysis.

### 4.2 Estimation of land surface recharge

Rainfall varies from about 1000 mm/y near the mountains to about 600 mm/y near the sea, and there is considerable variation in soil type and land use throughout the region. However, the eigenstructure approach requires only that land surface recharge be of the form  $P(x,y)R(t)$ . The fixed spatial pattern  $P(x,y)$  forms part of the linear system to be analysed, and only a single magnitude series  $R(t)$  is required as input to the system.

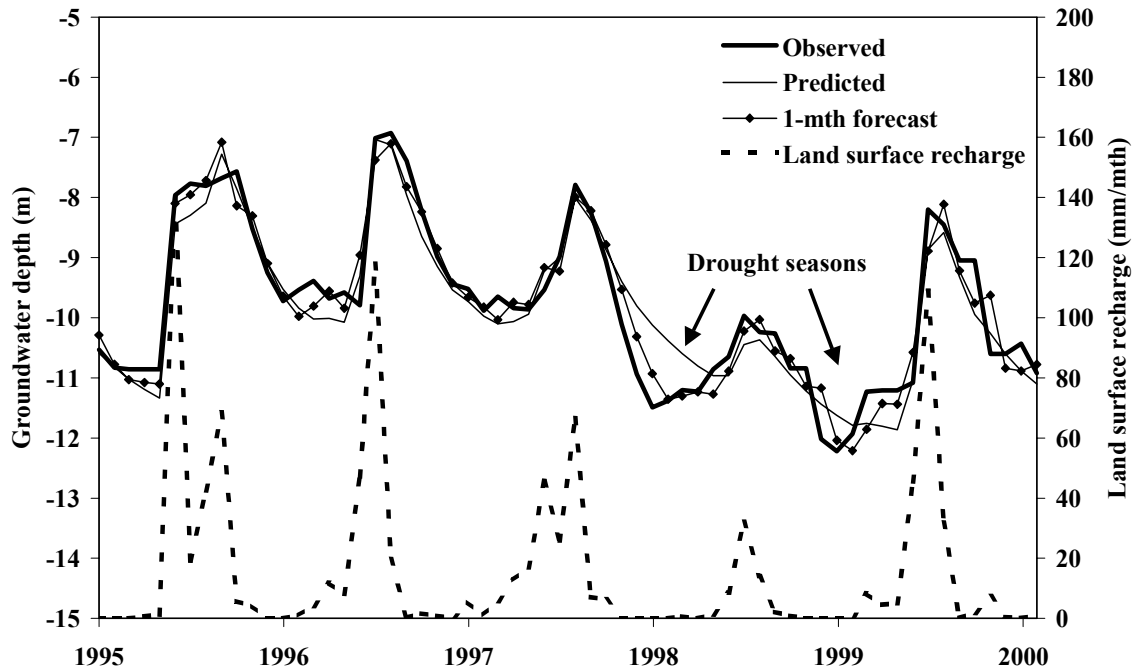


Figure 2. Comparison between the observed, predicted, and forecast piezometric levels at monthly intervals for Observation Well M36/0255, in response to land surface recharge.

A monthly time-series  $R_n$  of land surface recharge was estimated from a water balance model calculated on a daily basis. Rainfall and climatic data for Penman estimates of evaporation, at one location, were used for calculating water consumption and drainage  $R(t)$  from dryland pasture. The mean value of  $R(t)$  was 196 mm/y.

#### 4.3 Observation well

The eigenstructure modelling approach has been applied to several observation well records in the region (Bidwell and Morgan, 2002), of which one (M36/0255) was selected to illustrate this forecasting application. The piezometric observations at this well are a good indicator of low flow in the Halswell River, about 13 km distant, which is supplied by the aquifer. Therefore, the ability to forecast at this well enables early detection of likely environmental effects in the river caused by excessively low flow during droughts, because the likely cause is lowering of piezometric levels by additional groundwater abstraction in the region. A 17-year record of monthly observations (1983-2000) was used, with a few missing data that are easily handled by the method.

## 5 RESULTS

The significant parameter values of the eigenmodel system (Figure 1), steady local

piezometric datum  $D(x,y)$ , and noise parameter  $f_1$  are shown in Table 1.

Table 1. Eigenstructure-noise parameters for Observation Well M36/0255.

$k_v(x,y)$	( $\text{mth}^{-1}$ )	0.970
$k_1$	( $\text{mth}^{-1}$ )	0.046
$k_2$	( $\text{mth}^{-1}$ )	0.970
$g_1(x,y)$	( $\text{m mm}^{-1} \text{mth}$ )	0.279
$g_2(x,y)$	( $\text{m mm}^{-1} \text{mth}$ )	0.037
$D(x,y)$	(m)	-15.02
$f_1(x,y)$		0.711

The transfer function part of the model (2) accounted for 87% of the variance of the piezometric time-series  $U_n$ , and addition of the noise component  $N_n$  raised this to 94% for a one-step-ahead forecast. The remaining 6% was forecast error  $e_n$ , which had a standard deviation of 0.341 m. The randomness of the  $e_n$  series was demonstrated by means of the portmanteau lack of fit test (Box and Jenkins, 1970; 8.2.2) applied to the first 12 autocorrelation coefficients.

Figure 2 shows a comparison between observed piezometric head, prediction from the transfer function model, and one-month-ahead forecast from the complete transfer-function-noise model. The five-year record includes two drought seasons when groundwater abstractions caused

significant departure of piezometric head from that predicted on the basis of climatic data.

## 6 DISCUSSION AND CONCLUSIONS

Only the first two eigenvalues were required to account for the dynamics of the aquifer system at monthly time intervals. The dynamic behaviour is dominated by the first eigenvalue  $k_1$ , with its associated gain (0.279) equal to 88% of the total  $g_1(x,y)+g_2(x,y)$ . The value of  $k_1 = 0.046 \text{ mth}^{-1}$  corresponds to a mean storage time ( $1/k_1$ ) of 22 months. This value is similar to the first eigenvalue for other observation wells in this aquifer system (Bidwell and Morgan, 2002). The second eigenvalue  $k_2$  corresponds to a storage time of about one month, and accounts for minor transient behaviour.

The value (0.970) of the vadose zone coefficient  $k_v(x,y)$  also corresponds to a mean storage time of about one month for the system component that accounts for hydraulic transport lag at this location, as well as any effects of the one-month sampling interval. There is no apparent effect of any perched groundwater at this well, and groundwater level variations occur above the steady-state effect at 15 m below ground level (given by  $D(x,y)$ ).

The first-order noise model (with  $f_1 = 0.711$ ) was quite satisfactory for simulating the dependence structure, as shown by the test result for lack of dependence in the residuals  $e_n$ .

Figure 2 shows that the transfer function (first term of (12)) can be a good predictor of piezometric level during seasons of "normal" climatic stress, but not during the drought seasons of 1997/98 and 1998/99. During these periods the levels were significantly lower, probably due to increased groundwater abstraction for irrigation. However, the one-month-ahead forecasts from (13) effectively tracked the observed departure and would have provided indication of increased flow depletion in the Halswell River, on a real-time basis.

The forecast series (Figure 2) has been calculated as if the future input of land surface recharge is already known. For practical application during drought periods, the future input may be considered to be zero for worst-case management purposes, or likely non-zero scenario values may be used for a broader strategic view.

The eigenstructure approach provides a sound basis for identifying the structure of the ARMAX equation (12), because the coefficients in (12) are derived directly from the parallel system (10) based on theoretical solutions of the groundwater

flow equations. Decisions about the number of terms to be included in (12) are made in (10), for which the parameters are relatively independent under structural change. This approach contrasts with conventional time-series analysis in which the parameters of (12) all change with each structural addition.

The physical realism of the eigenstructure means that prior knowledge of the first eigenvalue (at least) can be transferred between observation sites, and sensible initial values and constraints can be set for parameter optimisation. These enable effective use of the optimisation (solver) function available in Microsoft Excel, rather than dedicated time-series-analysis software based on a statistical approach.

## 7 REFERENCES

- Ahn, H., Ground water drought management by a feedforward control method, *Journal of the American Water Resources Association*, 36(3), 501-510, 2000.
- Bidwell, V.J., P.F. Callander and C.R. Moore, An application of time-series analysis to groundwater investigation and management in Central Canterbury, New Zealand, *Journal of Hydrology (New Zealand)*, 30(1), 16-36, 1991.
- Bidwell, V.J. and M.J. Morgan, The eigenvalue approach to groundwater modelling for resource evaluation at regional scale, *Proceedings 4<sup>th</sup> International Conference on Calibration and Reliability in Groundwater Modelling, Prague*, 92-95, 2002.
- Box, G.E.P., and G.M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, 553 pp., San Francisco, 1970.
- Sahuquillo, A., An eigenvalue numerical technique for solving unsteady linear groundwater models continuously in time, *Water Resources Research*, 19(1), 87-93, 1983.
- Tankersley, C.D., W.D. Graham and K. Hatfield, Comparison of univariate and transfer function models of groundwater fluctuations, *Water Resources Research*, 29(10), 3517-3533, 1993.
- Young, P., *Recursive Estimation and Time-Series Analysis*, Springer-Verlag, 300 pp., Berlin, 1984.