WATER RESOURCES RESEARCH, VOL. 33, NO. 12, PAGES 2769–2780, DECEMBER 1997

WRE-353

Groundwater head sampling based on stochastic analysis

Hosung Ahn
Water Resources Evaluation Department, South Florida Water Management District, West Palm Beach

Jose D. Salas
Hydrologic Science and Engineering Program, Department of Civil Engineering, Colorado State University
Fort Collins

Abstract. The problem of determining a uniform sampling time interval for monitoring serially correlated groundwater heads is the main subject discussed here. The problem is approached by stochastic time series analysis and modeling. An autoregressive integrated moving average model is assumed to fit the underlying series. Given that groundwater head data can be sampled at different time intervals and that the same stochastic model must represent the time series regardless of the sampling timescale, the parameters of the underlying model for the series sampled at a given arbitrary time interval h are obtained as a function of h and as a function of the model parameters for the series sampled at a unit time interval. This is accomplished by linking the derived variances and autocovariances at the two sampling scales. The derived equations and the sampling design procedure are tested and illustrated using the groundwater head data of Collier County, Florida.

1. Introduction

Historical groundwater head data are valuable for mapping and mathematical modeling of groundwater systems. Many government agencies collect groundwater data within their jurisdiction boundaries for planning and management of water resources and environmental systems. However, in addition to the high cost of building monitoring wells, the cost of maintaining and operating a groundwater monitoring network is expensive. For example, in South Florida, which extends from the south of Lake Okeechobee to Key West, the groundwater monitoring network consists of ~1000 monitoring wells, and its annual operation cost reaches ~1.4×10^6 dollars (based on the 1994 budget figures). Thus the design of a groundwater monitoring system in the region becomes an important issue for water managers who are concerned with an efficient groundwater management program with a limited budget.

Two typical problems in groundwater monitoring design are selecting monitoring sites and choosing sampling time intervals. These problems have been commonly approached by statistical methods. The first problem related to the groundwater monitoring network design has been tackled by several approaches, including the variance-based approach [*Rouhani*, 1985; *Loaiciga*, 1989], optimization approaches [*Cieniawski et al.*, 1995; *Wagner*, 1995], and the heuristic approach based on facility location [*Hudak and Loaiciga*, 1992]. Likewise, the second problem of sampling time intervals, in which the underlying variable is serially correlated, has been studied especially in relation to water quality variables [*Lettenmaier*, 1976; *Sanders and Adrian*, 1978; *Loftis et al.*, 1991], where the sampling interval is determined as a function of the variance and autocorrelograms of the sample with the specified confidence interval for estimating the mean of the sequence. However, specific studies dealing with the groundwater head sampling interval, especially for serially correlated data, are lacking. Since the sampling time interval is directly related to the estimation accuracy and the annual operating cost of a monitoring program, the groundwater head sampling interval should be determined carefully.

It is generally known that groundwater head time series at monthly or smaller time intervals are serially correlated mainly because of the slow response time of groundwater flow as compared with the response time of surface runoff. To account for the serial correlation in determining the groundwater head sampling intervals, a stochastic time series model can be used. An example of choosing a sampling time interval for serially correlated data based on a simple stochastic time series model applied to a chemical manufacturing process for design of a discrete control scheme is given by *Box and Jenkins* [1976].

The main purpose of this paper is to develop a method for determining the sampling time interval of groundwater monitoring wells. Assuming a stochastic model to represent the groundwater head time series, the method finds a relationship between the model parameters at two different timescales and selects a sampling time interval for an allowable noise variance. The assumed stochastic model is the autoregressive integrated moving average (ARIMA) model. In section 2, the variance and the autocovariance equations for five low-order ARIMA models defined at arbitrary subsample intervals of h are derived as a function of the ARIMA model parameters at key sample intervals of H. Section 3 presents an application of the suggested method to determine the sampling time intervals of the selected groundwater wells in Collier County, Florida.

2. Methodology Based on the ARIMA Model

Measurements of groundwater head data are subject to several errors, including random errors, instrumental errors, data processing errors, and space-time interpolation errors. The magnitude of each error varies depending on many factors such as type of instruments, data collecting and processing

Copyright 1997 by the American Geophysical Union.

Paper number 97WR02187.
0043-1397/97/97WR-02187$09.00

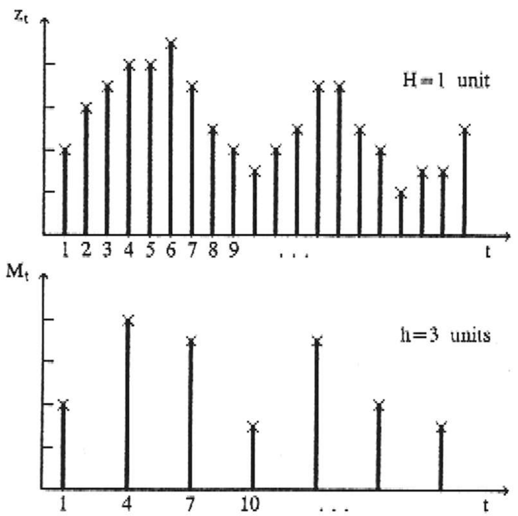

Figure 1. Schematic of a key sample series z_t (H = 1 unit) and a subsample series M_t (h = 3 units).

methods, and estimation methods. Among these errors, the errors caused by sampling a continuous data into discrete data points may be quite significant but inevitable because of the cost associated with groundwater monitoring. This type of error can be quantified by assuming a certain stochastic model and by estimating the corresponding noise variance.

One can assume that the underlying continuous process has an inherent noise variance. If the continuous process is discretized by a short (uniform) time interval, one would expect that the noise variance of the discretized process will be very close to that of the continuous process. However, as the sampling intervals increase, the degree of accuracy of the information decreases, and as a result, the noise variance increases. The concept that will be applied here is to find the noise variance of the underlying model for a series sampled at any arbitrary time intervals of h based on the model parameters for a series defined at another time interval H. This will be accomplished by linking the variance and the autocovariance properties of the process sampled at both time intervals. A relationship between the noise variance versus time interval will be established, which can be used for determining the sampling time interval given a predetermined allowable noise variance. This method is developed assuming that the groundwater head time series can be modeled by an ARIMA model.

Groundwater head time series for small time intervals such as several days are generally nonstationary due to long-term variations (seasonality and year-to-year changes) which can be eliminated by differencing the time series. Thus an ARIMA model is considered here. Let us suppose that z_t denotes a discrete time series at time t having uniform sampling intervals of H (key sample). Without loss of generality but for mathematical convenience, the key sample series is assumed to have a unit time interval ($H = 1$). In addition, let us define a backward shift operator B as $Bz_t = z_{t-1}$; hence $B^m z_t = z_{t-m}$ for any integer m. Also, let us define a backward difference operator ∇ as $\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$, so that $\nabla^d z_t = (1 - B)^d z_t$ for any integer d. Then, the ARIMA(p, d, q) process is given by [*Box and Jenkins*, 1976]

$$\phi(B)\nabla^d z_t = \theta(B)a_t \qquad (1)$$

in which d is the differencing order, $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ is the stationary autoregressive operator, $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \phi_q B^q)$ is the stationary moving average operator, p and q are the orders of autoregressive and moving average terms, respectively, and a_t is a white noise process having a mean of zero and a variance of $\sigma_a^2 = E[a_t^2]$, where $E[\ ]$ denotes expectation.

Low-order ARIMA models have been widely used in the field of water resources (see, for instance *Salas et al.* [1980] and *Stedinger and Vogel* [1984]). Thus five low-order models, namely ARIMA(0, 1, 1), ARIMA(0, 1, 2), ARIMA(1, 1, 0), ARIMA(2, 1, 0), and ARIMA(1, 1, 1) were chosen in this study.

2.1. ARIMA(0, 1, 1) Process

Let us assume that a time series z_t sampled at a unit time interval ($H = 1$) is available. The series is fitted by the ARIMA(0, 1, 1) process

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1} = a_t - \theta_1 a_{t-1} \qquad (2)$$

where the model parameters are θ_1 and σ_a^2. Suppose now that a subsample time series M_t is taken from the z_t series at a time interval $h > H$. (Figure 1 illustrates the case in which $H = 1$ and $h = 3$.) Then, the first-order backward difference of the M_t series can be expressed in terms of the z_t series as [*Box and Jenkins*, 1976]

$$\begin{aligned}\nabla M_t &= z_t - z_{t-h} = \nabla z_t + \nabla z_{t-1} + \cdots + \nabla z_{t-h+1}\\ &= (a_t + a_{t-1} + \cdots + a_{t-h+1})\\ &\quad - \theta_1(a_{t-1} + a_{t-2} + \cdots + a_{t-h})\\ &= A_t - \theta_1 A_{t-1} \end{aligned} \qquad (3)$$

$$\begin{aligned}\nabla M_{t-h} &= z_{t-h} - z_{t-2h} = \nabla z_{t-h} + \nabla z_{t-h-1} + \cdots + \nabla z_{t-2h+1}\\ &= (a_{t-h} + a_{t-h-1} + \cdots + a_{t-2h+1})\\ &\quad - \theta_1(a_{t-h-1} + a_{t-h-2} + \cdots + a_{t-2h})\\ &= A_{t-h} - \theta_1 A_{t-h-1}\end{aligned} \qquad (4)$$

where $A_{t-k} = (a_{t-k} + a_{t-k-1} + \cdots + a_{t-k-h+1})$ is the back-sum of h white noise terms for any time k. Now the variance and the lag-h autocovariance for the ∇M_t process may be obtained as a function of the parameters of the ∇z_t process as

$$\begin{aligned}\gamma_{\nabla M}(0) &= E[\nabla M_t \nabla M_t] = E[A_t A_t] - \theta_1 E[A_t A_{t-1}]\\ &\quad - \theta_1 E[A_{t-1} A_t] + \theta_1^2 E[A_{t-1} A_{t-1}]\\ &= h\sigma_a^2 - 2\theta_1(h-1)\sigma_a^2 + \theta_1^2 h\sigma_a^2\\ &= \{h(1 + \theta_1^2) - 2(h-1)\theta_1\}\sigma_a^2\end{aligned} \qquad (5)$$

$$\begin{aligned}\gamma_{\nabla M}(h) &= E[\nabla M_t \nabla M_{t-h}] = E[A_t A_{t-h}] - \theta_1 E[A_{t-1} A_{t-h}]\\ &\quad - \theta_1 E[A_t A_{t-h-1}] + \theta_1^2 E[A_{t-1} A_{t-h-1}]\\ &= 0 - \theta_1 E[a_{t-h} a_{t-h}] - 0 + 0 = -\theta_1 \sigma_a^2\end{aligned} \qquad (6)$$

and $\gamma_{\nabla M}(k)$ for $k > h$ ($k = 2h, 3h, 4h, \ldots$) is zero.

In addition, an ARIMA(0, 1, 1) model for the ∇M_t series can be written as

$$\nabla M_t = a_t - \theta_{1h} a_{t-h} \qquad (7)$$

where θ_{1h} and σ_{ah}^2 represent the model parameters of the series sampled at time interval h. That is, it is assumed that the sampling of an ARIMA(0, 1, 1) process defined at time interval H produces another ARIMA(0, 1, 1) process at time interval h. Thus the variance and the lag-h autocovariance for the ∇M_t process as a function of the parameters θ_{1h} and σ_{ah}^2 are given by [*Box and Jenkins*, 1976]

$$\gamma_{\nabla M}(0) = E[\nabla M_t \nabla M_t] = (1 + \theta_{1h}^2)\sigma_{ah}^2 \tag{8}$$

$$\gamma_{\nabla M}(h) = E[\nabla M_t \nabla M_{t-h}] = -\theta_{1h}\sigma_{ah}^2 \tag{9}$$

Now equating the variances of equations (5) and (8) and the lag-h autocovariances of equations (6) and (9), the following expressions are obtained:

$$\theta_{1h}\sigma_{ah}^2 = \theta_1\sigma_a^2 \tag{10}$$

$$\frac{(1 + \theta_{1h}^2)}{\theta_{1h}} = \frac{h(1 + \theta_1^2) - 2(h - 1)\theta_1}{\theta_1} \tag{11a}$$

These two expressions are useful for determining the noise variance σ_{ah}^2 at any arbitrary sampling interval h given the parameters of the model for the series sampled at time interval H. That is, for $h > H$, the above two equations can be solved for θ_{1h} and σ_{ah}^2 as a function of θ_1 and σ_a^2. On the other hand, for $h < H$, equations (10) and (11a) can still be used after switching θ_1, σ_a^2 and θ_{1h}, σ_{ah}^2, respectively, and replacing h by $1/h$. In this case, equation (11a) becomes

$$\frac{(1 + \theta_1^2)}{\theta_1} = \frac{(1/h)(1 + \theta_{1h}^2) - 2(1/h - 1)\theta_{1h}}{\theta_{1h}} \tag{11b}$$

For example, let us suppose that the variable z_t has been sampled at time interval $H = 1$ and fitted by an ARIMA(0, 1, 1) process with parameters $\theta_1 = 0.5$ and $\sigma_a^2 = 1.0$. Since the model order remains the same for the series sampled at time interval $h = 2$, equations (10) and (11a) give $\theta_{1h}\sigma_{ah}^2 = 0.5$ and $(1 + \theta_{1h}^2)/\theta_{1h} = 3.0$, which in turn give $\theta_{1h} = 0.382$ and $\sigma_{ah}^2 = 1.309$. Likewise, for $h = 1/2$, $\theta_1 = 0.5$, and $\sigma_a^2 = 1.0$, equations (10) and (11b) give $\theta_{1h} = 0.61$ and $\sigma_{ah}^2 = 0.82$. Figures 2 and 3 show θ_{1h} versus h and σ_{ah}^2/σ_h^2 versus h, respectively, in logarithmic scale, for $\sigma_a^2 = 1$, θ_1 ranging from 0.1 to 0.9 and h ranging from 0.025 to 60.

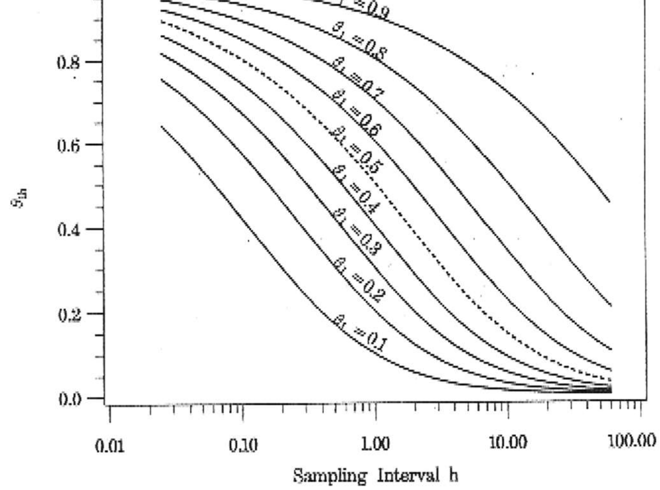

Figure 2. Parameter θ_{1h} versus sampling interval h for an ARIMA(0, 1, 1) model with parameters θ_1 and σ_a^2 defined at $H = 1$, in which $\sigma_a^2 = 1.0$ and θ_1 ranging from 0.1 to 0.9.

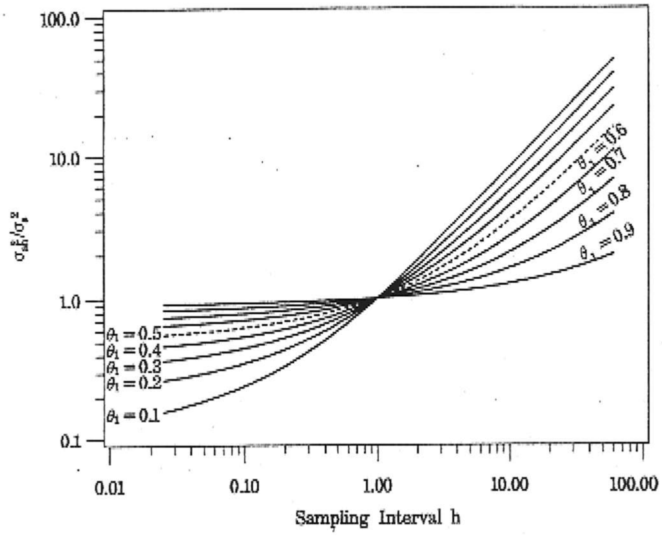

Figure 3. Scaled noise variance σ_{ah}^2/σ_a^2 versus sampling interval h for the Figure 2 conditions.

2.2. ARIMA(0, 1, 2) Process

The ARIMA(0, 1, 2) model of the series sampled at time interval $H = 1$ is given by

$$\nabla z_t = (1 - B)z_t = z_t - z_{t-1} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \tag{12}$$

where θ_1, θ_2, and σ_a^2 are the model parameters. Similar to the ARIMA(0, 1, 1) model, the first-order difference of the series M_t, which is sampled at time interval h, can be expressed in terms of the z_t process as

$$\begin{aligned}\nabla M_t = z_t - z_{t-h} &= \nabla z_t + \nabla z_{t-1} + \cdots + \nabla z_{t-h+1} \\ &= (a_t + a_{t-1} + \cdots + a_{t-h+1}) \\ &\quad - \theta_1(a_{t-1} + a_{t-2} + \cdots + a_{t-h}) \\ &\quad - \theta_2(a_{t-2} + a_{t-3} + \cdots + a_{t-h-1}) \\ &= A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2}\end{aligned} \tag{13}$$

$$\begin{aligned}\nabla M_{t-h} = z_{t-h} - z_{t-2h} &= (a_{t-h} + a_{t-h-1} + \cdots + a_{t-2h+1}) \\ &\quad - \theta_1(a_{t-h-1} + a_{t-h-2} + \cdots + a_{t-2h}) \\ &\quad - \theta_2(a_{t-h-2} + a_{t-h-3} + \cdots + a_{t-2h-1}) \\ &= A_{t-h} - \theta_1 A_{t-h-1} - \theta_2 A_{t-h-2}\end{aligned} \tag{14}$$

$$\begin{aligned}\nabla M_{t-2h} = z_{t-2h} - z_{t-3h} &= (a_{t-2h} + a_{t-2h-1} + \cdots + a_{t-3h+1}) \\ &\quad - \theta_1(a_{t-2h-1} + a_{t-2h-2} + \cdots + a_{t-3h}) \\ &\quad - \theta_2(a_{t-2h-2} + a_{t-2h-3} + \cdots + a_{t-3h-1}) \\ &= A_{t-2h} - \theta_1 A_{t-2h-1} - \theta_2 A_{t-2h-2}.\end{aligned} \tag{15}$$

Then, the variance and the lag-h autocovariance of the ∇M_t series are

$$\begin{aligned}\gamma_{\nabla M}(0) &= (1 + \theta_1^2 + \theta_2^2)E[A_t A_t] - 2\theta_1 E[A_t A_{t-1}] \\ &\quad - 2\theta_2 E[A_t A_{t-2}] + \theta_1\theta_2 E[A_{t-1} A_{t-2}] \\ &= \{h(1 + \theta_1^2 + \theta_2^2) - 2(h - 1)(\theta_1 + \theta_2 - \theta_1\theta_2) - 2\theta_2\}\sigma_a^2\end{aligned} \tag{16}$$

$$\gamma_{\nabla M}(h) = -\theta_1 E[A_{t-1}A_{t-h}] - \theta_2 E[A_{t-2}A_{t-h}] + \theta_1\theta_2 E[A_{t-2}A_{t-h-1}]$$
$$= -\theta_1\sigma_a^2 - (\theta_2\sigma_a^2 + \theta_2\sigma_a^2) + \theta_1\theta_2\sigma_a^2$$
$$= (\theta_1\theta_2 - \theta_1 - 2\theta_2)\sigma_a^2 \tag{17}$$

in which $E[A_{t-i}A_{t-j}] = (h + i - j)\sigma_a^2$. However, since $\gamma_{\nabla M}(k) = 0$ for $k > 1$, the above variance and autocovariance equations are not sufficient to estimate three parameters of the ARIMA(0, 1, 2) model. Thus one may use the second best model whenever the ARIMA(0, 1, 2) model is selected as the best model for a given key sample series.

Alternatively, the model parameters can be estimated by the variance, the lag-h autocovariance, and the second partial autocorrelation coefficient of the ARIMA(0, 1, 2) model. Denoting ϕ_{qk} ($k = 1, \cdots, q$) as the kth partial autocorrelation coefficient in an order q moving average model and $\rho_k = \gamma_{\nabla M}(k)/\gamma_{\nabla M}(0)$ as the kth autocorrelation coefficient, the first two Yule-Walker equations for the ARIMA(0, 1, 2) model are written as

$$\begin{aligned}\phi_{21} + \rho_1\phi_{22} &= \rho_1\\ \rho_1\phi_{21} + \phi_{22} &= \rho_2\end{aligned} \tag{18}$$

Since $\gamma_{\nabla M}(2h)$ of the ∇M_t process is zero, the lag-$2h$ partial autocorrelation coefficient of the ∇M_t series is given by

$$\phi_{22} = \frac{-\rho_1^2}{1-\rho_1^2} = \frac{-\gamma_{\nabla M}(h)^2}{\gamma_{\nabla M}(0)^2 - \gamma_{\nabla M}(h)^2} \tag{19}$$

where both $\gamma_{\nabla M}(0)$ and $\gamma_{\nabla M}(h)$ are obtained by equations (16) and (17), respectively.

Also, an ARIMA(0, 1, 2) model for the ∇M_t series may be written as

$$\nabla M_t = a_t - \theta_{1h}a_{t-h} - \theta_{2h}a_{t-2h} \tag{20}$$

where θ_{1h}, θ_{2h}, and σ_{ah}^2 are the parameters. The variance, the lag-h autocovariance, and the lag-$2h$ partial autocorrelation coefficient of the above ∇M_t series written in terms of the parameters θ_{1h}, θ_{2h}, and σ_{ah}^2 are given by [*Box and Jenkins*, 1976]

$$\gamma_{\nabla M}(0) = (1 + \theta_{1h}^2 + \theta_{2h}^2)\sigma_{ah}^2 \tag{21}$$

$$\gamma_{\nabla M}(h) = (\theta_{1h}\theta_{2h} - \theta_{1h})\sigma_{ah}^2 \tag{22}$$

$$\phi_{22} = \frac{(\theta_{1h}\theta_{2h} - \theta_{1h})^2}{(1 + \theta_{1h}^2 + \theta_{2h}^2)^2 - (\theta_{1h}\theta_{2h} - \theta_{1h})^2} \tag{23}$$

Then, the parameters θ_{1h}, θ_{2h}, and σ_{ah}^2 can be obtained as a function of the parameters θ_1, θ_2, and σ_a^2 by equating equations (16) and (21), (17) and (22), and (19) and (23), respectively.

2.3. ARIMA(1, 1, 1) Process

The ARIMA(1, 1, 1) process of the series defined at time interval $H = 1$ is given by

$$\nabla z_t = \phi_1\nabla z_{t-1} + a_t - \theta_1 a_{t-1} \tag{24a}$$

where ϕ_1, θ_1, and σ_a^2 are the model parameters. The corresponding ARIMA(1, 1, 1) model for the M_t series sampled at time interval h from the underlying series z_t is expressed as

$$\nabla M_t = \phi_{1h}\nabla M_{t-h} + a_t - \theta_{1h}a_{t-h} \tag{24b}$$

where ϕ_{1h}, θ_{1h}, and σ_{ah}^2 are the parameters. Following a similar approach as in the previous cases, the variance and the lag-h and lag-$2h$ autocovariances of ∇M_t series are obtained as a function of the ARIMA(1, 1, 1) model parameters of the ∇z_t series as

$$\gamma_{\nabla M}(0) = \left\{\frac{\phi_1^2}{1-\phi_1^2}\left[h(1 + \theta_1^2 - 2\phi_1\theta_1) + 2\psi(1 - \phi_1\theta_1)\frac{K_2}{\phi_1}\right] + [h(1 + \theta_1^2) + 2(h-1)\psi - 2h\phi_1\theta_1] + 2\psi(\xi K_1 - \theta_1 K_2)\right\}\sigma_a^2 \tag{25}$$

$$\gamma_{\nabla M}(h) = \left\{\frac{\beta^2\psi(\phi_1^h - 1)^2(1 - \phi_1\theta_1)}{1-\phi_1^2} + \psi + \beta(\phi_1^{h-1} - 1) \cdot [\phi_1 + \beta(\phi_1^h - 1)] - \theta_1\beta^2(\phi_1^h - 1)^2\right\}\sigma_a^2 \tag{26}$$

$$\gamma_{\nabla M}(2h) = \left\{\frac{\phi_1^h\psi(\phi_1^h - 1)^2(1 - \phi_1\theta_1)}{(\phi_1 - 1)^2(1 - \phi_1^2)} + \phi_1^{h-2}\psi\beta^2(\phi_1^h - 1)^2 \cdot (1 - \theta_1\phi_1)\right\}\sigma_a^2 \tag{27}$$

where $K_1 = \beta[\beta(\phi_1^{h-2} - 1) - h + 2]$, $K_2 = \beta[\beta(\phi_1^{h-1} - 1) - h + 1]$, $\psi = \phi_1 - \theta_1$, $\beta = \phi_1/(\phi_1 - 1)$, and $\xi = 0$ for $h = 2$ or $\xi = 1$ for $h > 2$. Full derivations of the above equations are shown in Appendix A.

In addition, the variance and the lag-h and lag-$2h$ autocovariances of ∇M_t in terms of the parameters ϕ_{1h}, θ_{1h}, and σ_{ah}^2 are [*Box and Jenkins*, 1976]

$$\gamma_{\nabla M}(0) = \frac{1 + \theta_{1h}^2 - 2\phi_{1h}\theta_{1h}}{1 - \phi_{1h}^2}\sigma_{ah}^2 \tag{28}$$

$$\gamma_{\nabla M}(h) = \frac{(1 - \phi_{1h}\theta_{1h})(\phi_{1h} - \theta_{1h})}{1 - \phi_{1h}^2}\sigma_{ah}^2 \tag{29}$$

$$\gamma_{\nabla M}(2h) = \phi_{1h}\gamma_{\nabla M}(h) \tag{30}$$

As before, equations (25)–(30) can be used to determine the parameters ϕ_{1h}, θ_{1h}, and σ_{ah}^2 as a function of the parameters ϕ_1, θ_1, and σ_a^2.

2.4. ARIMA(2, 1, 0) Process

The ARIMA(2, 1, 0) model for the series sampled at time interval $H = 1$ and the corresponding one sampled at time interval h are respectively given by

$$\nabla z_t = \phi_1\nabla z_{t-1} + \phi_2\nabla z_{t-2} + a_t \tag{31}$$

$$\nabla M_t = \phi_{1h}\nabla M_{t-h} + \phi_{2h}\nabla M_{t-2h} + a_t \tag{32}$$

where $\{\phi_1, \phi_2$, and $\sigma_a^2\}$ and $\{\phi_{1h}, \phi_{2h}$, and $\sigma_{ah}^2\}$ are the parameters for both time intervals H and h, respectively. Appendix B gives the derivation of the variance and lag-h and lag-$2h$ autocovariances of the ∇M_t series. They are summarized here as

$$\gamma_{\nabla M}(0) = h\sigma_a^2 + (\phi_1^2 + \phi_2^2)\sum_{i=1}^{h}\sum_{j=1-i}^{h-i}\gamma_{\nabla z}(j) + 2\phi_1\phi_2\sum_{i=1}^{h}\sum_{j=i-2}^{h-i-1}\gamma_{\nabla z}(j) + 2\phi_1\sum_{i=1}^{h-1}\sum_{j=1}^{h-i}\gamma_{\nabla za}(j-1)$$

$$+ 2\phi_2 \sum_{i=1}^{h-2} \sum_{j=1}^{h-i-1} \gamma_{\nabla za}(j-1) \quad (33)$$

$$\gamma_{\nabla M}(h) = (\phi_1^2 + \phi_2^2) \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j)$$

$$+ \phi_1\phi_2 \left\{ \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j-1) + \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j+1) \right\}$$

$$+ \phi_1 \sum_{i=1}^{h} \sum_{j=i}^{h-i-1} \gamma_{\nabla za}(j-1)$$

$$+ \phi_2 \left\{ \sum_{i=1}^{h-1} \gamma_{\nabla za}(i-1) + \sum_{i=1}^{h-1} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(j-1) \right\} \quad (34)$$

$$\gamma_{\nabla M}(2h) = (\phi_1^2 + \phi_2^2) \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j)$$

$$+ \phi_1\phi_2 \left\{ \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j-1) + \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j+1) \right\}$$

$$+ \phi_1 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(h+j-1) + \phi_2 \sum_{i=1}^{h-1} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(h+j-2) \quad (35)$$

where $\gamma_{\nabla z}(k)$ and $\gamma_{\nabla za}(k)$ are defined in Appendix B. In addition, the variance and autocovariances of the series ∇M_t in terms of the parameters ϕ_{1h}, ϕ_{2h}, and σ_{ah}^2 are [*Box and Jenkins*, 1976]

$$\gamma_{\nabla M}(0) = (1 - \phi_{2h})\sigma_{ah}^2/\lambda \quad (36)$$

$$\gamma_{\nabla M}(h) = \phi_{1h}\sigma_{ah}^2/\lambda \quad (37)$$

$$\gamma_{\nabla M}(2h) = (\phi_{1h}^2 + \phi_{2h} - \phi_{2h}^2)\sigma_{ah}^2/\lambda \quad (38)$$

with $\lambda = (1 + \phi_{2h})[(1 - \phi_{2h})^2 - \phi_{1h}^2]$. Equations (33)–(38) can be used to determine the parameters ϕ_{1h}, ϕ_{2h}, and σ_{ah}^2 as a function of ϕ_1, ϕ_2, and σ_a^2.

2.5. ARIMA(1, 1, 0) Process

The variance and the lag-h autocovariance of the sampled process which results from the ARIMA(1, 1, 0) process are

$$\gamma_{\nabla M}(0) = \phi_1^2 \left[h\gamma_{\nabla z}(0) + 2 \sum_{i=1}^{h-1} \sum_{j=1-i}^{h-i} \gamma_{\nabla z}(j) \right] + h\sigma_a^2$$

$$+ 2\phi_1\sigma_a^2 \left[h - 1 + \xi \sum_{i=1}^{h-2} \sum_{j=1}^{h-i-1} \phi_1^{j-1} \right] \quad (39)$$

$$\gamma_{\nabla M}(h) = \phi_1^2 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j) + \phi_1\sigma_a^2 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \phi_1^{j-1} \quad (40)$$

Since $\gamma_{\nabla z}(0) = \sigma_a^2/(1 - \phi_1^2)$ and $\gamma_{\nabla z}(j) = \phi_1^j \gamma_{\nabla z}(0)$ for $j \geq 0$, the above equations simplify to

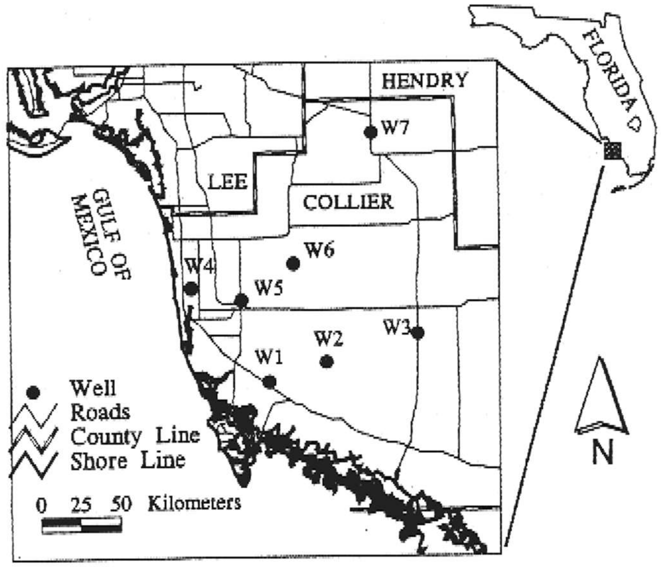

Figure 4. Location map showing the selected groundwater monitoring wells.

$$\gamma_{\nabla M}(0) = \left[\frac{\phi_1^2}{1 - \phi_1^2}(h + 2K_2) + h + 2(h-1)\phi_1 + 2\xi K_1 \right] \sigma_a^2 \quad (41)$$

$$\gamma_{\nabla M}(h) = \frac{\phi_1(\phi_1^h - 1)^2}{(1 - \phi_1^2)(\phi_1 - 1)^2} \sigma_a^2 \quad (42)$$

where K_1, K_2, and ξ are the same as for the ARIMA(1, 1, 1) model. Likewise, $\gamma_{\nabla M}(0)$ and $\gamma_{\nabla M}(h)$ written in terms of the parameters ϕ_{1h} and σ_{ah}^2 are [*Box and Jenkins*, 1976]

$$\gamma_{\nabla M}(0) = \sigma_{ah}^2/(1 - \phi_{1h}^2) \quad (43)$$

$$\gamma_{\nabla M}(h) = \phi_{1h}\gamma_{\nabla M}(0) \quad (44)$$

Finally, as in previous cases, one can find the parameters ϕ_{1h} and σ_{ah}^2 as a function of the parameters ϕ_1 and σ_a^2 by equating equations (41) and (43) and equations (42) and (44), respectively.

3. Application to Groundwater Head Data

This section illustrates an application of the modeling approach suggested herein to a groundwater head monitoring problem. For this study, seven sets of daily groundwater head time series (1985–1990) measured from the surficial aquifer monitoring wells located in Collier County, Florida, were used. Figure 4 shows the location of the selected wells. The area is characterized by moderately drained sandy soil with extensive agricultural and urban development so that the short-term groundwater fluctuation in the region is largely induced by local rainfall and groundwater pumping. Each well has a digital water level recorder which automatically logs groundwater head. The accuracy of the digital water level recorder is about ±3 mm (0.01 feet) [*Lietz et al.*, 1994].

3.1. Model Order Versus Sampling Time Interval

First of all, the key sample series of daily groundwater head in each well was fitted by ARIMA models. The daily time series of all seven sites (plots are not presented here) appear to be nonstationary so that differencing of the key sample series

Table 1. Some Statistics and the Estimated ARIMA Model Parameters for the Selected Groundwater Time Series

Statistics/Parameters	W1	W2	W3	W4	W5	W6	W7
Mean, for $d = 0$, m	1.788	1.489	1.760	3.323	2.021	2.126	6.908
Variance, for $d = 0$, m^2	0.268	0.187	0.274	0.085	0.064	0.203	0.480
Variance, for $d = 1$, m^2	0.00122	0.00328	0.00317	0.00631	0.00297	0.00132	0.02873
Selected ARIMA order	(1, 1, 0)	(0, 1, 2)	(1, 1, 0)	(1, 1, 1)	(2, 1, 0)	(1, 1, 1)	(0, 1, 2)
Parameters estimated by the maximum likelihood method							
ϕ_1	0.314	NA	0.237	0.826	0.137	0.499	NA
ϕ_2	NA	NA	NA	NA	−0.122	NA	NA
θ_1	NA	0.035	NA	0.907	NA	−0.080	−0.007
θ_2	NA	0.112	NA	NA	NA	NA	0.1626
σ_u^2 (m^2)	0.00110	0.00325	0.00299	0.00619	0.00289	0.00091	0.02817
AIC	−8696.7	−6329.9	−6510.1	−4916.5	−6587.9	−9109.9	−1599.4

H = 1 day; NA, nonapplicable; AIC, Akaike information criterion.

was necessary for making the data stationary. Ten low-order ARIMA models were fitted for each data, namely (0, 1, 1), (0, 1, 2), (1, 1, 1), (1, 1, 0), (2, 1, 0), (0, 2, 1), (0, 2, 2), (1, 2, 1), (1, 2, 0), and (2, 2, 0). Then, the best fitted ARIMA model in each well was selected based on the Akaike information criterion (AIC) [*Salas et al.*, 1980]. Also, the Bayesian information criterion and a criterion based on the minimum noise variance were investigated as alternative model selection criteria, but the results by both of these statistics were nearly identical to those obtained by the AIC statistic. Table 1 presents the best fitted ARIMA model in each well with its model parameters which were estimated by the least squares method.

To predict the noise variance of a subsample series defined at a uniform sampling time interval h given a key sample series at H, an appropriate ARIMA model structure needs to be assumed. The main assumption used in the derivation of the variance and the autocovariance equations of two different time intervals is that the ARIMA model order remains the same regardless of the time intervals of the series. While this assumption must hold based on mathematical arguments, one may question whether such an assumption is valid or not for the actual data. Thus a further stochastic analysis was made in order to verify to what degree the groundwater head data at various timescales can be fitted by the same order model. For this verification, several subsample series at arbitrary uniform sampling intervals were taken from each well's daily groundwater head data, and each of them were fitted by the ARIMA models. When subsample series at time intervals of h are taken from a key sample at $H = 1$, h different subsamples may arise. For instance, for $h = 2$, two possible subsamples from a key sample series are $x_1, x_3, x_5, \cdots$ and $x_2, x_4, x_6, \cdots$. Thus, for eight arbitrary timescales, namely, $h = 1, 2, 4, 5, 10, 20, 30$, and 60 days, a total of 132 subsamples of groundwater head series were available for analysis of each well. Then, for each well and each sampling interval h, the referred 10 ARIMA models were fitted from which the best model for each time series was selected based on the AIC statistics. Since the selected models for different subsamples may be different, the "best" model and the "second best" model selected in each case was that with the highest frequency.

Table 2 lists the best and the second best fitted ARIMA model orders for each time interval h. Table 2 shows that for all seven wells, the first-order differencing ARIMA models are always better than the second-order differencing models. The results of Table 2 may be analyzed in various ways. For instance, if the criteria are based on the frequency in which the best model (criterion a) or either the best or the second best model (criterion b) selected for the key sample series ($H = 1$) is also the best or the second best model for the other timescales h, the following values in Table 3 are obtained.

The results are quite good given that 10 different low-order models were considered. Furthermore, if the criteria are based on the frequency of the best model (given preference to the model having shorter time intervals) to be also the best (criterion a) or the best or the second best (criterion b) throughout all timescales, one gets the values shown in Table 4. The two sets of results in Table 4 confirm reasonably well the fact that the same model is valid for any sampling interval h. In addition, if one considers the likelihood that the best model for a given time interval is also the best or the second best for any adjacent time interval (for example, $h = 1$ and $h = 2$, or $h = 5$ and $h = 10$), the resulting frequencies are 45% (22 out of 49) and 69% (34 out of 49), respectively, both of which are significantly larger than 10% (that is, the theoretical likelihood when 10 low-order models are considered).

Table 2. Best (Second Best) Fitted ARIMA Models Based on AIC Statistics for Various Sampling Intervals

h, days	W1	W2	W3	W4	W5	W6	W7
1	A (E)	E (B)	A (B)	C (E)	B (E)	C (B)	E (C)
2	C (B)	B (E)	B (E)	C (E)	C (E)	C (E)	E (C)
4	C (B)	A (D)	D (A)	C (E)	C (E)	C (D)	B (D)
5	C (B)	E (B)	A (D)	E (C)	C (E)	A (E)	D (C)
10	A (D)	C (A)	A (D)	A (D)	D (B)	B (C)	D (B)
20	A (E)	E (B)	A (D)	C (D)	C (D)	D (A)	A (D)
30	E (B)	E (B)	E (B)	E (D)	C (E)	E (D)	E (B)
60	E (B)	E (B)	E (B)	E (D)	C (E)	E (D)	E (B)

A, ARIMA(1, 1, 0) model; B, ARIMA(2, 1, 0); C, ARIMA(1, 1, 1); D, ARIMA(0, 1, 1); E, ARIMA(0, 1, 2).

Table 3. Frequency in Which the Order of Subsample Model is the Same as That of Key Sample Model

	Frequency, %						
Criteria	W1	W2	W3	W4	W5	W6	W7
a	29	57	43	43	0	29	43
b	57	71	100	86	86	43	57

Criterion a, best model; criterion b, best or second best model.

Table 4. Frequency of Being the Selected Best Model Order

Criteria	Frequency, %						
	W1	W2	W3	W4	W5	W6	W7
Best model order	(1, 1, 1)	(0, 1, 2)	(1, 1, 0)	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(0, 1, 2)
a	38	63	50	50	75	38	50
b	38	75	63	63	75	50	50

3.2. Noise Variance Versus Sampling Time Interval

On the basis of the ARIMA model parameters determined for the key sample series at time intervals of $H = 1$ day, the "derived" noise variances σ_{ah}^2 at arbitrary sampling intervals ranging from 2 to 60 days were computed using the derived variance and autocovariance equations. The implicit systems of nonlinear equations that must be solved for unknown parameters of the subsampled process were solved by a modified Powell algorithm with the parameters of key sample series as the initial input [*International Mathematics and Statistics Libraries*, 1991]. The modified Powell algorithm provided good convergent solutions in this case without experiencing any particular numerical problems. Although the best fitted ARIMA model order at each well was identified in subsection 3.1, the noise variances σ_{ah}^2 for the five low-order ARIMA models discussed in section 2 were estimated for comparison. The results of σ_{ah}^2 versus h for wells W2, W3, and W4 are plotted in Figures 5, 6, and 7, respectively (the results of the other four wells are similar but not presented here). Figures 5–7 also include the "estimated" noise variances σ_{ah}^2 versus h which were obtained by fitting the best ARIMA model in each well (listed in Table 2) with the h interval subsample series taken from the daily key sample series. Since the variation of noise variances for all different h interval series taken from a daily series was nearly negligible compared to the variation of the noise variances resulting from different ARIMA models, the estimated noise variance of each time interval h here is that obtained from an arbitrary selected series among h different series.

For wells W2 and W3, there are good agreements between the derived and the corresponding estimated noise variances. For well W4 (Figure 7), some differences between the derived and the estimated noise variances are observed especially for large values of h even though the results are good for $h < 10$ days. For well W4, one may be able to get better results for large values of h if the key sample time interval H is greater than 1 day. Therefore the relation σ_{ah}^2 versus h shown in Figure 8 was computed considering the key sample series for $H = 5$ days taken from the daily time series. The result shown in Figure 8 indicates that except for $h = 1$ and $h = 60$ days, reasonably good correspondence exists between the derived and the estimated noise variances throughout the rest of the time intervals.

Also, it should be noted that the alternative solution approach for the ARIMA(0, 1, 2) model given by equations (19) and (23) gives reasonable estimations of noise variances in most cases; Especially, Figures 5 and 8 clearly demonstrate that for $h \gg H$ case, the result based on the alternative solution approach for the ARIMA(0, 1, 2) model is better than the approach based on the second best ARIMA model.

The foregoing analyses indicate that the derived variance and autocovariance equations in section 2 may provide useful estimates of noise variances over a wide range of sampling timescales. The relation σ_{ah}^2 versus h (shown graphically, for instance, in Figure 8) also demonstrates that the noise variance σ_{ah}^2 of subsample series for both $h > H$ and $h < H$ cases may

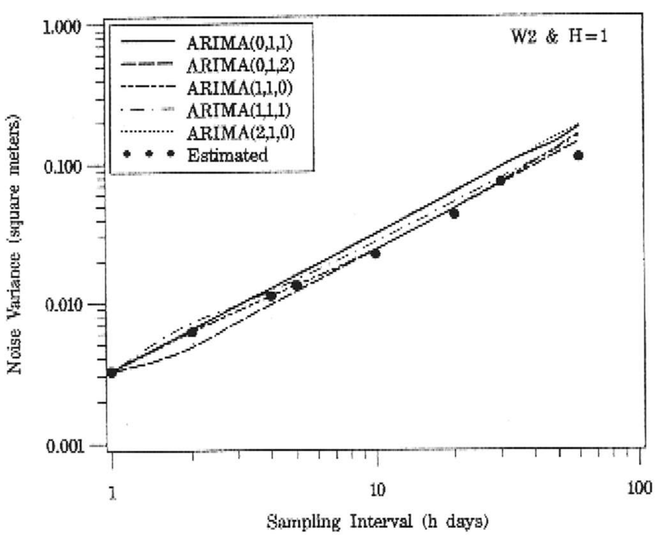

Figure 5. Noise variance σ_{ah}^2 versus sampling interval h obtained from the derived variance and autocovariance equations with the groundwater head data at well W2, in which $H = 1$ day was used. In addition, the estimated noise variance (dots) were obtained from the best fitted ARIMA model to the groundwater head series sampled at each time interval h.

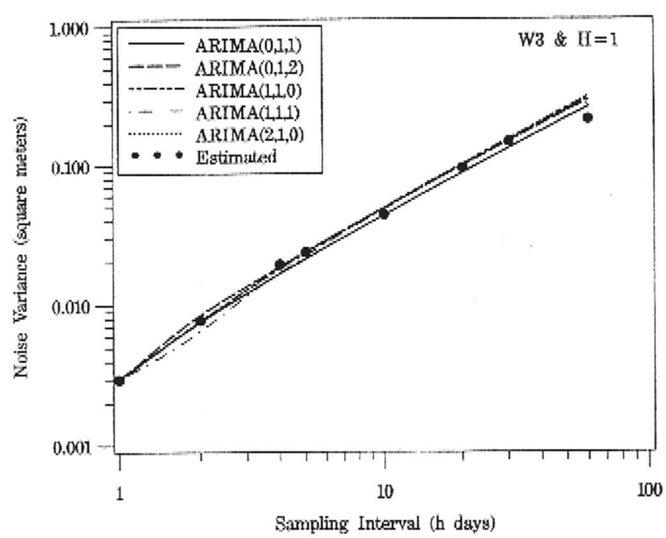

Figure 6. Same as Figure 3, but well W3 and $H = 1$ day.

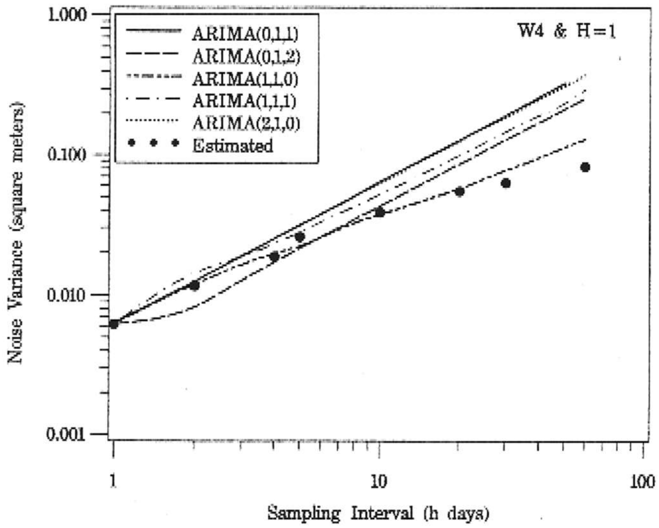

Figure 7. Same as Figure 3, but well W4 and H = 1 day.

be easily obtained on the basis of model parameters determined for a key sample series at H interval. The case $h < H$ is especially important when the subsample series for time interval h is not available.

The relations σ_{ah}^2 versus h as in Figures 5–8 can be applied for determining the groundwater head sampling interval at each well. That is, under a specified allowable noise variance, the corresponding sampling interval at each well can be read from the σ_{ah}^2 versus h curves. The allowable noise variance might be either a constant throughout a region or a variable depending on the regional condition. For example, the sampling intervals which are obtained from the σ_{ah}^2 versus h curves (marked by dots in Figures 5–8) for two arbitrary allowable noise variances are shown in Table 5.

As one may expect, the well having the highest sample variance (W7) needs to be sampled more frequently, while the one with the smallest variation (W5) needs less frequent sampling. Note that even though the variances of two wells are nearly identical (i.e., W1 and W3 for d = 0, or W2 and W3 for d = 1), the sampling time intervals obtained by the proposed method are different as a result of the effect of serial correlation of the data.

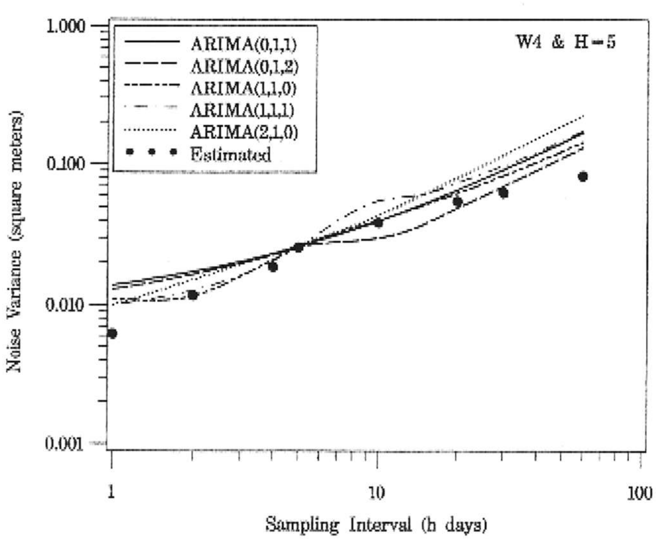

Figure 8. Same as Figure 3, but well W4 and H = 5 days.

Table 5. Suggested Groundwater Head Sampling Intervals

Allowable Noise Variance σ_{ah}^2	Sampling Interval h, Days						
	W1	W2	W3	W4	W5	W6	W7
0.04 m^2	15	16	9	10	30	10	1
0.10 m^2	30	45	20	90	100	25	5

4. Concluding Remarks

An approach for determining a uniform sampling time interval for serially correlated groundwater heads using stochastic time series analysis has been presented herein. The proposed approach involves fitting an ARIMA model to the available time series sampled at a given time interval H (key sample series), finding the parameters of the model for the series sampled at any arbitrary time interval h (subsample series), and establishing a relationship between the noise variance σ_{ah}^2 and time interval h, from which one can determine the sampling time interval h given a predetermined allowable noise variance for the well at hand. The main contribution of this paper has been linking the properties of the stochastic models of series sampled at two different time intervals and establishing a sampling criterion based on the noise variance.

The derived equations and overall sampling design procedure have been tested for the groundwater head data collected from Collier County, Florida. In general, the noise variances computed from the derived equations match reasonably well the corresponding estimated variances obtained from the actual data. In cases where some departures are observed, the approach is still useful by selecting a key sample series at different time intervals. The suggested approach can be implemented for determining a groundwater monitoring program throughout a region. For instance, a sampling criteria may be specified such that the same error variance is maintained throughout the region. This case was illustrated in subsection 3.2 where the allowable error variance σ_{ah}^2 was set and the uniform sampling time interval h at each well was determined from the derived σ_{ah}^2 versus h relations. In addition, the proposed method may be attractive for multilayered aquifer systems where the serial correlations of groundwater heads may be different from one layer to another. In such cases, it may be possible to design a sampling scheme for obtaining uniformity of measurement errors in groundwater monitoring throughout layers based solely upon the temporal correlations of the groundwater head data.

Appendix A: Variance and Autocovariances for the ARIMA(1, 1, 1) Model With Sampling Interval h

For the ARIMA(1, 1, 1) model of equation (24a) the cross covariance of ∇z_t and a_t is given by $\gamma_{\nabla za}(k)$ = Cov (∇z_t, a_{t-k}) = $E[\nabla z_t a_{t-k}]$ since $E[\nabla z_t] = E[a_t] = 0$. Such cross covariances are given by $\gamma_{\nabla za}(0) = E[\nabla z_t a_t] = E[(\phi_1 \nabla z_{t-1} + a_t - \theta_1 a_{t-1})a_t] = \sigma_a^2$ and $\gamma_{\nabla za}(k) = E[\nabla z_t a_{t-k}] = \phi_1 E[\nabla z_{t-1} a_{t-k}] = \phi_1^{k-1}\psi\sigma_a^2$ for $k \geq 1$, with $\psi = \phi_1 - \theta_1$ and $\gamma_{\nabla za}(k) = 0$ for $k < 0$.

From equations (24a) and (24b) the ∇M_t series at sampling time interval h can be expressed in terms of the z_t series at sampling interval $H = 1$ as

$$\nabla M_t = z_t - z_{t-h} = \nabla z_t + \nabla z_{t-1} + \cdots + \nabla z_{t-h+1}$$

$$= \phi_1(\nabla z_{t-1} + \nabla z_{t-2} + \cdots + \nabla z_{t-h})$$

$$+ (a_t + a_{t-1} + \cdots + a_{t-h+1}) - \theta_1(a_{t-1} + a_{t-2} + \cdots + a_{t-h})$$

$$= \phi_1 \nabla Z_{t-1} + A_t - \theta_1 A_{t-1} \qquad \text{(A1)}$$

$$\nabla M_{t-h} = z_{t-h} - z_{t-2h} = \phi_1(\nabla z_{t-h-1} + \nabla z_{t-h-2} + \cdots + \nabla z_{t-2h})$$

$$+ (a_{t-h} + a_{t-h-1} + \cdots + a_{t-2h+1})$$

$$- \theta_1(a_{t-h-1} + a_{t-h-2} + \cdots + a_{t-2h})$$

$$= \phi_1 \nabla Z_{t-h-1} + A_{t-h} - \theta_1 A_{t-h-1} \qquad \text{(A2)}$$

$$\nabla M_{t-2h} = z_{t-2h} - z_{t-3h} = \phi_1(\nabla z_{t-2h-1} + \nabla z_{t-2h-2} + \cdots + \nabla z_{t-3h})$$

$$+ (a_{t-2h} + a_{t-2h-1} + \cdots + a_{t-3h+1})$$

$$- \theta_1(a_{t-2h-1} + a_{t-2h-2} + \cdots + a_{t-3h})$$

$$= \phi_1 \nabla Z_{t-2h-1} + A_{t-2h} - \theta_1 A_{t-2h-1} \qquad \text{(A3)}$$

where $\nabla Z_{t-k} = (z_{t-k} + z_{t-k-1} + \cdots + z_{t-k-h+1})$ and $A_{t-k} = (a_{t-k} + a_{t-k-1} + \cdots + a_{t-k-h+1})$ have h back-sum terms.

Then, the variance of the series ∇M_t expressed as a function of the parameters of the ∇z_t series is given by

$$\gamma_{\nabla M}(0) = E[\nabla M_t \nabla M_t] = \phi_1^2 E[\nabla Z_{t-1} \nabla Z_{t-1}]$$

$$+ 2\phi_1 E[\nabla Z_{t-1} A_t] - 2\phi_1\theta_1 E[\nabla Z_{t-1} A_{t-1}]$$

$$- 2\theta_1 E[A_t A_{t-1}] + (1 + \theta_1^2) E[A_{t-1} A_{t-1}] \qquad \text{(A4)}$$

Since the variance and autocovariances of ∇z_t are given by $\gamma_{\nabla z}(0) = (1 + \theta_1^2 - 2\phi_1\theta_1)\sigma_a^2/(1 - \phi_1^2)$, $\gamma_{\nabla z}(1) = (1 - \phi_1\theta_1)\psi\sigma_a^2/(1 - \phi_1^2)$, and $\gamma_{\nabla z}(k) = \phi_1^{k-1}\gamma_{\nabla z}(1)$ for $k > 1$ [*Box and Jenkins*, 1976], the first expectation term in equation (A4) is determined by

$$E[\nabla Z_{t-1} \nabla Z_{t-1}]$$

$$= \left\{ \begin{array}{llll} \gamma_{\nabla z}(0) & +\gamma_{\nabla z}(1) & +\cdots & +\gamma_{\nabla z}(h-1) \\ +\gamma_{\nabla z}(1) & +\gamma_{\nabla z}(0) & +\cdots & +\gamma_{\nabla z}(h-2) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(h-1) & +\gamma_{\nabla z}(h-2) & +\cdots & +\gamma_{\nabla z}(0) \end{array} \right\}$$

$$= \sum_{i=1}^{h} \sum_{j=1-i}^{h-i} \gamma_{\nabla z}(j)$$

$$= \frac{\sigma_a^2}{1 - \phi_1} \left\{ h(1 + \theta_1^2 - 2\phi_1\theta_1) + 2\psi(1 - \phi_1\theta_1) \right.$$

$$\left. \cdot \frac{\beta}{\phi_1} [\beta(\phi_1^{h-1} - 1) - (h - 1)] \right\}.$$

The second expectation in equation (A4) is given by

$$E[\nabla Z_{t-1} A_t] = \left\{ \begin{array}{llll} (0 + 1 + \psi & +\phi_1\psi & +\cdots & +\phi_1^{h-3}\psi)\sigma_a^2 \\ +(0 + 0 + 1 & +\psi & +\cdots & +\phi_1^{h-4}\psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(0 + 0 + 0 & +0 & +\cdots & +0)\sigma_a^2 \end{array} \right\}$$

$$= \sigma_a^2 \left(h - 1 + \xi\psi \sum_{i=1}^{h-2} \sum_{j=1}^{h-i-1} \phi_1^{j-1} \right)$$

$$= \sigma_a^2 \left\{ h - 1 + \xi\psi \frac{\beta}{\phi_1} [\beta(\phi_1^{h-2} - 1) - (h - 2)] \right\}$$

in which $\xi = 0$ for $h = 2$ and $\xi = 1$ for $h > 2$, and $\beta = \phi_1/(\phi_1 - 1)$. Likewise, the other expectation terms are given by

$$E[\nabla Z_{t-1} A_{t-1}] = \left\{ \begin{array}{llll} (1 + \psi & +\phi_1\psi & +\cdots & +\phi_1^{h-2}\psi)\sigma_a^2 \\ +(0 + 1 & +\psi & +\cdots & +\phi_1^{h-3}\psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(0 + 0 & +0 & +\cdots & +1)\sigma_a^2 \end{array} \right\}$$

$$= \sigma_a^2 \left(h + \psi \sum_{i=1}^{h-1} \sum_{j=1}^{h-i} \phi_1^{j-1} \right)$$

$$= \sigma_a^2 \left\{ h + \psi \frac{\beta}{\phi_1} [\beta(\phi_1^{h-1} - 1) - (h - 1)] \right\}$$

$$E[A_t A_t] = E[A_{t-1} A_{t-1}] = h\sigma_a^2$$

$$E[A_t A_{t-1}] = (h - 1)\sigma_a^2$$

Rearranging equation (A4) with the above expectation terms results in equation (25). Also, the lag-h autocovariance of ∇M_t as a function of the parameters of ∇z_t becomes

$$\gamma_{\nabla M}(h) = E[\nabla M_t \nabla M_{t-h}] = \phi_1^2 E[\nabla Z_{t-1} \nabla Z_{t-h-1}]$$

$$+ \phi_1 E[\nabla Z_{t-1} A_{t-h}] - \phi_1\theta_1 E[\nabla Z_{t-1} A_{t-h-1}]$$

$$- \theta_1 E[A_{t-1} A_{t-h}] \qquad \text{(A5)}$$

in which the expectation terms are computed by

$$E[\nabla Z_{t-1} \nabla Z_{t-h-1}]$$

$$= \left\{ \begin{array}{llll} \gamma_{\nabla z}(h) & +\gamma_{\nabla z}(h-1) & +\cdots & +\gamma_{\nabla z}(1) \\ +\gamma_{\nabla z}(h+1) & +\gamma_{\nabla z}(h) & +\cdots & +\gamma_{\nabla z}(2) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(2h-1) & +\gamma_{\nabla z}(2h-2) & +\cdots & +\gamma_{\nabla z}(h) \end{array} \right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j) = \beta^2\psi\sigma_a^2 \frac{(\phi_1^h - 1)^2(1 - \phi_1\theta_1)}{\phi_1^2(1 - \phi_1^2)}$$

$$E[\nabla Z_{t-1} A_{t-h}]$$

$$= \left\{ \begin{array}{llll} (\phi_1^{h-2}\psi & +\phi_1^{h-3}\psi & +\cdots & +\psi + 1)\sigma_a^2 \\ +(\phi_1^{h-1}\psi & +\phi_1^{h-2}\psi & +\cdots & +\phi_1\psi + \psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(\phi_1^{2h-3}\psi & +\phi_1^{2h-4}\psi & +\cdots & +\phi_1^{h-2}\psi)\sigma_a^2 \end{array} \right\}$$

$$= \sigma_a^2 \left[1 + \psi \left(\sum_{i=1}^{h-1} \phi_1^{i-1} + \sum_{i=1}^{h-1} \sum_{j=i}^{h+i-1} \phi_1^{j-1} \right) \right]$$

$$= \sigma_a^2 \left\{ 1 + \frac{\psi\beta}{\phi_1} (\phi_1^{h-1} - 1) \left[1 + \frac{\beta}{\phi_1} (\phi_1^h - 1) \right] \right\}$$

$$E[A_{t-1}A_{t-h}] = \sigma_a^2$$

$E[\nabla Z_{t-1}A_{t-h-1}]$

$$= \left\{\begin{array}{llll} (\phi_1^{h-1}\psi & +\phi_1^{h-2}\psi & +\cdots & +\psi)\sigma_a^2 \\ +(\phi_1^{h}\psi & +\phi_1^{h-1}\psi & +\cdots & +\phi_1\psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(\phi_1^{2h-2}\psi & +\phi_1^{2h-3}\psi & +\cdots & +\phi_1^{h-1}\psi)\sigma_a^2 \end{array}\right\}$$

$$= \psi\sigma_a^2 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \phi_1^{j-1} = \psi\sigma_a^2\beta^2(\phi_1^h - 1)^2/\phi_1^2$$

Substituting the foregoing expectations into equation (A5) yields equation (26).

In addition, the lag-$2h$ autocovariance of ∇M_t is given by

$$\gamma_{\nabla M}(2h) = E[\nabla M_t \nabla M_{t-2h}] = \phi_1^2 E[\nabla Z_{t-1}\nabla Z_{t-2h-1}]$$
$$+ \phi_1 E[\nabla Z_{t-1}A_{t-2h}] - \phi_1\theta_1 E[\nabla Z_{t-1}A_{t-2h-1}] \quad \text{(A6)}$$

where the expectation terms are computed by

$E[\nabla Z_{t-1}\nabla Z_{t-2h-1}]$

$$= \left\{\begin{array}{llll} \gamma_{\nabla z}(2h) & +\gamma_{\nabla z}(2h-1) & +\cdots & +\gamma_{\nabla z}(h+1) \\ +\gamma_{\nabla z}(2h+1) & +\gamma_{\nabla z}(2h) & \cdots & +\gamma_{\nabla z}(h+2) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(3h-1) & +\gamma_{\nabla z}(3h-2) & \cdots & +\gamma_{\nabla z}(2h) \end{array}\right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j) = \phi_1^h\psi\sigma_a^2 \frac{(\phi_1^h - 1)^2(1 - \phi_1\theta_1)}{(\phi_1 - 1)^2(1 - \phi_1^2)}$$

$E[\nabla Z_{t-1}A_{t-2h-1}]$

$$= \left\{\begin{array}{llll} (\phi_1^{2h-1}\psi & +\phi_1^{2h-2}\psi & +\cdots & +\phi_1^h\psi)\sigma_a^2 \\ +(\phi_1^{2h}\psi & +\phi_1^{2h-1}\psi & +\cdots & +\phi_1^{h+1}\psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(\phi_1^{3h-2}\psi & +\phi_1^{3h-3}\psi & +\cdots & +\phi_1^{2h-1}\psi)\sigma_a^2 \end{array}\right\}$$

$$= \psi\sigma_a^2 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \phi_1^{h+j-1} = \psi\sigma_a^2\beta^2\phi_1^{h-2}(\phi_1^h - 1)^2$$

$E[\nabla Z_{t-1}A_{t-2h}]$

$$= \left\{\begin{array}{llll} (\phi_1^{2h-2}\psi & +\phi_1^{2h-3}\psi & +\cdots & +\phi_1^{h-1}\psi)\sigma_a^2 \\ +(\phi_1^{2h-1}\psi & +\phi_1^{2h-2}\psi & +\cdots & +\phi_1^{h}\psi)\sigma_a^2 \\ \vdots & \vdots & \cdots & \vdots \\ +(\phi_1^{3h-3}\psi & +\phi_1^{3h-4}\psi & +\cdots & +\phi_1^{2h-2}\psi)\sigma_a^2 \end{array}\right\}$$

$$= \psi\sigma_a^2 \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \phi_1^{h+j-2} = \psi\sigma_a^2\beta^2\phi_1^{h-3}(\phi_1^h - 1)^2$$

Replacing the above expectation terms into equation (A6) results in equation (27).

Appendix B: Variance and Autocovariances of the ARIMA(2, 1, 0) Model With Sampling Interval h

For the ARIMA(2, 1, 0) model of equation (31), the cross covariances of ∇z_t and a_t can be expressed as $\gamma_{\nabla za}(0) = E[\nabla z_t a_t] = E[(\phi_1\nabla z_{t-1} + \phi_2\nabla z_{t-2} + a_t)a_t] = \sigma_a^2$, $\gamma_{\nabla za}(1) = E[\nabla z_t a_{t-1}] = E[(\phi_1\nabla z_{t-1} + \phi_2\nabla z_{t-2} + a_t)a_{t-1}] = \phi_1 E[\nabla z_{t-1}a_{t-1}] = \phi_1\sigma_a^2$, and $\gamma_{\nabla za}(k) = E[\nabla z_t a_{t-k}] = \phi_1\gamma_{\nabla za}(k-1) + \phi_2\gamma_{\nabla za}(k-2)$ for $k > 1$.

From equations (31) and (32), the ∇M_t series can be expressed in terms of the z_t series as

$$\nabla M_t = \nabla z_t - \nabla z_{t-h} = \nabla z_t + \nabla z_{t-1} + \cdots + \nabla z_{t-h+1}$$
$$= \phi_1(\nabla z_{t-1} + \nabla z_{t-2} + \cdots + \nabla z_{t-h})$$
$$+ \phi_2(\nabla z_{t-2} + \nabla z_{t-3} + \cdots + \nabla z_{t-h-1})$$
$$+ (a_t + a_{t-1} + \cdots + a_{t-h+1})$$
$$= \phi_1\nabla Z_{t-1} + \phi_2\nabla Z_{t-2} + A_t \quad \text{(B1)}$$

$$\nabla M_{t-h} = \nabla z_{t-h} - \nabla z_{t-2h}$$
$$= \phi_1(\nabla z_{t-h-1} + \nabla z_{t-h-2} + \cdots + \nabla z_{t-2h})$$
$$+ \phi_2(\nabla z_{t-h-2} + \nabla z_{t-h-2} + \cdots + \nabla z_{t-2h-1})$$
$$+ (a_{t-h} + a_{t-h-1} + \cdots + a_{t-2h+1})$$
$$= \phi_1\nabla Z_{t-h-1} + \phi_2\nabla Z_{t-h-2} + A_{t-h} \quad \text{(B2)}$$

$$\nabla M_{t-2h} = \nabla z_{t-2h} - \nabla z_{t-3h}$$
$$= \phi_1(\nabla z_{t-2h} + \nabla z_{t-2h-1} + \cdots + \nabla z_{t-3h})$$
$$+ \phi_2(\nabla z_{t-2h-2} + \nabla z_{t-2h-3} + \cdots + \nabla z_{t-3h-1})$$
$$+ (a_{t-2h} + a_{t-2h-1} + \cdots + a_{t-3h+1})$$
$$= \phi_1\nabla Z_{t-2h-1} + \phi_2\nabla Z_{t-2h-2} + A_{t-2h}. \quad \text{(B3)}$$

Then, the variance of the ∇M_t series written in terms of the parameters of the ∇z_t series becomes

$$\gamma_{\nabla M}(0) = E[\nabla M_t\nabla M_t] = (\phi_1^2 + \phi_2^2)E[\nabla Z_{t-1}\nabla Z_{t-1}]$$
$$+ E[A_tA_t] + 2\phi_1 E[\nabla Z_{t-1}A_t]$$
$$+ 2\phi_1\phi_2 E[\nabla Z_{t-1}\nabla Z_{t-2}] + 2\theta_2 E[\nabla Z_{t-2}A_t] \quad \text{(B4)}$$

where $E[\nabla Z_{t-1}\nabla Z_{t-1}]$ and $E[A_tA_t]$ are the same as for the ARIMA(1, 1, 1) model, and the remaining expectation terms are determined by

$E[\nabla Z_{t-2}A_t]$

$$= \left\{\begin{array}{llll} 0 + 0 + \gamma_{\nabla za}(0) & +\gamma_{\nabla za}(1) & +\cdots & +\gamma_{\nabla za}(h-3) \\ +0 + 0 + 0 & +\gamma_{\nabla za}(0) & +\cdots & +\gamma_{\nabla za}(h-4) \\ \vdots & \vdots & \cdots & \vdots \\ +0 + 0 + 0 & +0 & +\cdots & +0 \end{array}\right\}$$

$$= \sum_{i=1}^{h-2} \sum_{j=1}^{h-i-1} \gamma_{\nabla za}(j-1)$$

$E[\nabla Z_{t-1}\nabla Z_{t-2}]$

$$= \left\{\begin{array}{llll} \gamma_{\nabla z}(1) & +\gamma_{\nabla z}(0) & +\cdots & +\gamma_{\nabla z}(h-2) \\ +\gamma_{\nabla z}(0) & +\gamma_{\nabla z}(1) & +\cdots & +\gamma_{\nabla z}(h-3) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(h-2) & +\gamma_{\nabla z}(h-3) & +\cdots & +\gamma_{\nabla z}(1) \end{array}\right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i-2}^{h-i-1} \gamma_{\nabla z}(j)$$

$E[\nabla Z_{t-1} A_t]$

$$= \left\{ \begin{array}{cccc} 0 + \gamma_{\nabla za}(0) & +\gamma_{\nabla za}(1) & +\cdots & +\gamma_{\nabla za}(h-2) \\ +0+0 & +\gamma_{\nabla za}(0) & +\cdots & +\gamma_{\nabla za}(h-3) \\ \vdots & \vdots & \cdots & \vdots \\ +0+0 & +0 & +\cdots & +0 \end{array} \right\}$$

$$= \sum_{i=1}^{h-1} \sum_{j=1}^{h-i} \gamma_{\nabla za}(j-1).$$

Then, equation (33) results by replacing the above expectation terms into equation (B4).

Also, the lag-h autocovariance of the ∇M_t series is obtained by

$$\gamma_{\nabla M}(h) = E[\nabla M_t \nabla M_{t-h}] = (\phi_1^2 + \phi_2^2) E[\nabla Z_{t-1} \nabla Z_{t-h-1}]$$
$$+ \phi_1\phi_2 E[\nabla Z_{t-2} \nabla Z_{t-h-1}] + \phi_1 E[\nabla Z_{t-1} \nabla Z_{t-h-2}]$$
$$+ \phi_1 E[\nabla Z_{t-1} A_{t-h}] + \phi_2 E[\nabla Z_{t-2} A_{t-h}] \qquad \text{(B5)}$$

where the term $E[\nabla Z_{t-1} \nabla Z_{t-h-1}]$ is the same as that for the ARIMA(1, 1, 1) model, and the other expectation terms are computed by

$E[\nabla Z_{t-2} \nabla Z_{t-h-1}]$

$$= \left\{ \begin{array}{cccc} \gamma_{\nabla z}(h-1) & +\gamma_{\nabla z}(h-2) & +\cdots & +\gamma_{\nabla z}(0) \\ +\gamma_{\nabla z}(h) & +\gamma_{\nabla z}(h-1) & +\cdots & +\gamma_{\nabla z}(1) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(2h-2) & +\gamma_{\nabla z}(2h-3) & +\cdots & +\gamma_{\nabla z}(h-1) \end{array} \right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i}^{h-i-1} \gamma_{\nabla z}(j-1)$$

$E[\nabla Z_{t-1} \nabla Z_{t-h-2}]$

$$= \left\{ \begin{array}{cccc} \gamma_{\nabla z}(h+1) & +\gamma_{\nabla z}(h) & +\cdots & +\gamma_{\nabla z}(2) \\ +\gamma_{\nabla z}(h+2) & +\gamma_{\nabla z}(h+1) & +\cdots & +\gamma_{\nabla z}(3) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla z}(2h) & +\gamma_{\nabla z}(2h-1) & +\cdots & +\gamma_{\nabla z}(h+1) \end{array} \right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(j+1)$$

$E[\nabla Z_{t-1} A_{t-h}]$

$$= \left\{ \begin{array}{cccc} \gamma_{\nabla za}(h-1) & +\gamma_{\nabla za}(h-2) & +\cdots & +\gamma_{\nabla za}(0) \\ +\gamma_{\nabla za}(h) & +\gamma_{\nabla za}(h-1) & +\cdots & +\gamma_{\nabla za}(1) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla za}(2h-2) & +\gamma_{\nabla za}(2h-3) & +\cdots & +\gamma_{\nabla za}(h-1) \end{array} \right\}$$

$$= \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(j-1)$$

$E[\nabla Z_{t-2} A_{t-h}]$

$$= \left\{ \begin{array}{cccc} \gamma_{\nabla za}(h-2) & +\gamma_{\nabla za}(h-3) & +\cdots & +0 \\ +\gamma_{\nabla za}(h-1) & +\gamma_{\nabla za}(h-2) & +\cdots & +\gamma_{\nabla za}(0) \\ \vdots & \vdots & \cdots & \vdots \\ +\gamma_{\nabla za}(2h-3) & +\gamma_{\nabla za}(2h-2) & +\cdots & +\gamma_{\nabla za}(h-2) \end{array} \right\}$$

$$= \sum_{i=1}^{h-1} \gamma_{\nabla za}(i-1) + \sum_{i=1}^{h-1} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(j-1).$$

In addition, the lag-$2h$ autocovariance of the ∇M_t series is given by

$$\gamma_{\nabla M}(2h) = E[\nabla M_t \nabla M_{t-2h}] = (\phi_1^2 + \phi_2^2) E[\nabla Z_{t-1} \nabla Z_{t-2h-1}]$$
$$+ \phi_1\phi_2 E[\nabla Z_{t-2} \nabla Z_{t-2h-1}] + \phi_1\phi_2 E[\nabla Z_{t-1} \nabla Z_{t-2h-2}]$$
$$+ \phi_1 E[\nabla Z_{t-1} A_{t-2h}] + \phi_2 E[\nabla Z_{t-2} A_{t-2h}] \qquad \text{(B6)}$$

in which the expectation terms are

$$E[\nabla Z_{t-1} \nabla Z_{t-2h-1}] = \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j)$$

$$E[\nabla Z_{t-2} \nabla Z_{t-2h-1}] = \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j-1)$$

$$E[\nabla Z_{t-1} \nabla Z_{t-2h-2}] = \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla z}(h+j+1)$$

$$E[\nabla Z_{t-1} A_{t-2h}] = \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(h+j-1)$$

$$E[\nabla Z_{t-2} A_{t-2h}] = \sum_{i=1}^{h} \sum_{j=i}^{h+i-1} \gamma_{\nabla za}(h+j-2)$$

Acknowledgments. The research leading to this paper has been supported by the South Florida Water Management District project "Assessment of groundwater monitoring network in southwest Florida region" (H.A.) and the USDA Agricultural Research Service project "Field to farm to ecosystem scale decision support models" (J.D.S.). The authors acknowledge the comments and suggestions to the original manuscript by Renzo Ross and two anonymous reviewers.

References

Box, G. E. P., and G. W. Jenkins, *Time Series Analysis: Forecasting and Control*, rev. ed., Holden-Day, Merrifield, Va., 1976.

Cieniawski, S. E., J. W. Eheart, and S. Ranjithan, Using genetic algorithms to solve multi-objective groundwater monitoring problem, *Water Resour. Res.*, *31*(2), 399–409, 1995.

Hudak, P. F., and H. A. Loaiciga, A location modeling approach for groundwater monitoring network augmentation, *Water Resour. Res.*, *28*(3), 643–649, 1992.

International Mathematics and Statistics Libraries, *User's Manual: FORTRAN Subroutines for Mathematical Applications*, Houston, Tex., 1991.

Lettenmaier, D. P., Detection of trends in water quality data from records with dependent observations, *Water Resour. Res.*, *12*(5), 1037–1046, 1976.

Lietz, C., H. LaRose, and T. Richards, Water resources data Florida, Water Year 1993, Vol. 2B, South Florida ground water, *U.S. Geol. Surv. Water Data Rep. FL-93-2B*, 1994.

Loaiciga, H. A., An optimization approach for groundwater quality monitoring network design, *Water Resour. Res.*, *25*(8), 1771–1783, 1989.

Loftis, J. C., G. B. McBride, and J. C. Ellis, Considerations of scale in water quality monitoring and data analysis, *Water Resour. Bull.*, *27*(2), 255–264, 1991.

Rouhani, S., Variance reduction analysis, *Water Resour. Res.*, *21*(6), 837–846, 1985.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resour. Publ., Fort Collins, Colo., 1980.

Sanders, T. G., and D. D. Adrian, Sampling frequency for river quality monitoring, *Water Resour. Res.*, *14*(4), 569–576, 1978.

Stedinger, J. R., and R. M. Vogel, Disaggregation procedures for generating serially correlated flow vectors, *Water Resour. Res.*, *20*(1), 47–56, 1984.

Wagner, B. J., Sampling design methods for groundwater modeling under uncertainty, *Water Resour. Res.*, *31*(10), 2581–2591, 1995.

H. Ahn, Water Resources Evaluation Department, South Florida Water Management District, West Palm Beach, FL 33406. (e-mail: hosung.ahn@sfwmd.gov)

J. D. Salas, Hydrologic Science and Engineering Program, Department of Civil Engineering, Colorado State University, Fort Collins, CO 80523. (e-mail: jsalas@lance.colostate.edu)

(Received January 3, 1997; revised July 22, 1997;
accepted July 30, 1997.)