

## A PRACTICAL ESTIMATION METHOD FOR GROUNDWATER LEVEL ELEVATIONS

CHOON KYUNG CHO AND SUNGKWON KANG

**ABSTRACT.** A practical estimation method for groundwater level elevations is introduced. Using geostatistical techniques with drift, averaging process and ratio, experimental variograms show significant improved correlation compared with those from conventional techniques. The estimation method is applied to a field experimental data set.

### 1. Introduction

Groundwater and groundwater contamination have been important issues in environmental concerns. Since movement of groundwater contaminants follows groundwater flow direction, estimation of groundwater level elevation is one of key factors for determining the fate of the contaminants and their spatial distributions. Often, in practice, due to technical and economical reasons, groundwater level elevations are measured at irregularly spaced locations. However, most groundwater numerical simulation algorithms are developed under the assumption that input data on uniform or regular grids are given. Also, very often, hydrologists need information at unmeasured locations. Therefore, based on an available data set, a systematic method for estimating groundwater level elevations at unsampled locations is needed.

The purpose of this paper is to address a geostatistical method, called kriging, and introduce some of new concepts such as averaging process and ratio which produce significant improvement for obtaining correlation between the geological structure of the site and the variance of groundwater level elevations. The method produces more

---

Received May 12, 1997.

1991 Mathematics Subject Classification: 86A32, 65U05.

Key words and phrases: geostatistics, kriging, groundwater level, optimization.

The authors thank Dr. Stauffer for providing field data. This research was supported by the 1996 Chosun University Research Funds.

accurate continuous information than the one obtained by the conventional method. As an interpolation method, kriging is one of optimization techniques, and it accounts for underlying geological structures. Therefore, statistical, geological, and mathematical aspects are involved in the development of estimation methods. Our analyses and statements are by no means mathematically or statistically rigorous due to uncertainties of underground geological structures and their complex interactions with groundwater in real fields. Instead, we focus on methodologies and main ideas arising in estimation of groundwater level elevations so that our methods are able to apply for practical simulations. Rigorous analyses will be reported in a forthcoming paper.

Geostatistics can be considered as a collection of techniques for solving the estimation problems involving spatial variables. In recent years, geostatistical methods have been applied to various problems in hydrology and other sciences (see, for example, [2,3,7,14,18]). Among them, the punctual kriging and the universal kriging have been most commonly used in soil and water sciences to estimate a spatially distributed random variable at unsampled locations. Both methods are based on the second-order stationarity or the geostatistical intrinsic hypothesis: the first and second moments of the random variables or regionalized variables are stationary (see Section 2). The two estimation methods are "optimal" in the sense that they produce the exactly same estimated values as the field data at measured locations (unbiasedness) and minimize the variance of the estimated expected values at unmeasured locations. But, since most field data do not satisfy the intrinsic assumption, there are needs to account for apparent anisotropic aspects resided in the field data.

In this paper, groundwater level elevations at unmeasured locations are estimated based on an available field samples using a modified punctual kriging method. Applying drifts and the new proposed concepts, averaging process and ratio, the intrinsic hypothesis will be overcome. General kriging procedure for groundwater level elevations is following.

- 1) Based on a given sample data set collected in a site, find an experimental semivariogram representing correlation between geological structure of the site and the variance of data distributions. This is the key step in kriging, and it is difficult. All available information such as geological structure, dynamical or geophysical aspects of the site are

incorporated. The proposed concepts such as averaging process and ratio are considered in this step.

2) Find a mathematical model representing the continuous information resided in the experimental semivariogram. This step requires development of a numerical algorithm for fitting the mathematical model to the experimental semivariogram.

3) Solve the kriging system to obtain the optimal weights on the selected samples for kriging. This solution will produce the best expected estimation groundwater level elevation at each desired location. In this step, an efficient numerical solver for the kriging system is needed.

The contents of this paper are following. In Section 2, regionalized variables, random functions, theoretical and experimental variograms, and mathematical models needed in kriging procedure are explained. To account for spatial anisotropic aspects of the geological structure, a drift and two new proposed concepts, averaging process and ratio, are also introduced. The punctual kriging is described in Section 3. In Section 4, the 26 water level elevation samples collected at an experimental site are chosen for kriging. The semivariograms obtained by considering averaging and ratio show significant improvement compared to those by conventional methods.

## 2. Regionalized variables and Semivariogram

### 2.1. Regionalized variables and semivariogram

A *regionalized variable* is a numerical function with a spatial distribution which varies from one place to another with apparent continuity, but the changes of which cannot be represented by any workable function [3]. It contains two apparently contradictory characteristics: (i) a local, random, erratic aspect which calls to mind the notion of random variable and (ii) a general (or average) structured aspect which requires a certain functional representation.

Typical regionalized variables are functions describing natural phenomena that have geographic distributions such as the elevation of the ground surface, the amount of precipitation in an area, porosity in a reservoir, ore content in a mineralized body, etc. Some of characteristics of regionalized variables are *localization*, *anisotropy* and *continuity*. Localization is associated with geometric features such as specific size,

shape, and orientation of the sample. Since we are interested in estimating groundwater level elevations, localization will be restricted to point supports in this paper. Changes in values of an anisotropic regionalized variable are gradual in one direction and rapid or irregular in another. The spatial variation of a regionalized variable may be extremely large to very small, depending on the phenomenon. Despite the complexity of the fluctuations, an average continuity is generally present [18]. A proper formulation must take this double aspect of randomness and structure into account in such a way as to provide a simple representation of the spatial variability and lead to a consistent and operational approach of the solution of problems. One such formulation is the probabilistic interpretation as provided by *random functions*.

A *random variable*  $Z(x)$  is a variable which takes a certain number of numerical values according to a certain probability distribution defined in a certain spatial domain  $\Omega$ , and the regionalized variable  $z(x)$  can be considered as a particular realization of the set of random variables  $\{Z(x) : x \in \Omega\}$ . This set of random variables is called a *random function* and will be written  $Z(x)$ . Thus, a random function  $Z(x)$  expresses the random and structured aspects of the regionalized variable  $z(x)$ : (i) locally, at a point  $x_1$ ,  $Z(x_1)$  is a random variable, and (ii)  $Z(x)$  is also a random function in the sense that for each pair of points  $x_1$  and  $x_1 + h$ , the corresponding random variables  $Z(x_1)$  and  $Z(x_1 + h)$  are not, in general, independent but are related by a correlation expressing the spatial structure of the initial regionalized variable  $z(x)$ . The probabilistic interpretation of a regionalized variable  $z(x)$  as a particular realization of a certain random function  $Z(x)$  has an operative sense only when it is possible to infer all or part of the probability law which defined this random function in its entirety. Obviously, it is not rigorously possible to infer the probability law of a random function  $Z(x)$  from a single realization  $z(x)$  which is, in addition, limited to a finite number of sample points  $x_i$ . Many realizations  $z_1(x), \dots, z_k(x)$  of the random function  $Z(x)$  are required in order to infer the probability law of  $Z(x)$ . Since, in practice, we shall be limited to a single realization  $\{z(x_i)\}$  of  $Z(x)$  at the positions  $x_i$ , certain assumptions are necessary. These assumptions involve various degrees of spatial homogeneity and are introduced under the general heading

of the stationary hypothesis. In practice, even if only over a certain region, the phenomenon under study can very often be considered as homogeneous, the regionalized variable repeating itself in space. This homogeneity or repetition provides the equivalent of many realizations of the same random function  $Z(x)$  and permits a certain amount of statistical inference. Two experimental values  $z(x_0)$  and  $z(x_0 + h)$  at two different points  $x_0$  and  $x_0 + h$  can, thus, be considered as two different realizations of the same random variable  $Z(x_0)$ . This type of approach is not peculiar to geostatistics, it is used to infer the distribution law of the random variable  $Z(x)$  from a histogram of data values  $\{z(x_i)\}$ , or more simply, to infer the mathematical expectation  $E\{Z(x)\}$  from the arithmetic mean of the data.

Consider a random function  $Z(x)$  defined on a certain spatial domain  $\Omega \subset \mathbf{R}^n$ . For every set of  $k$  points,  $x_1, \dots, x_k \in \Omega$ , called *support points*, there corresponds a  $k$ -component vectorial random variable

$$\{Z(x_1), Z(x_2), \dots, Z(x_k)\}.$$

This vectorial random variable is characterized by the  $k$ -variable distribution function

$$F_{x_1 x_2 \dots x_k}(z_1, z_2, \dots, z_k) = \text{Prob}\{Z(x_1) < z_1, \dots, Z(x_k) < z_k\}.$$

The set of all these distribution functions, for all positive integers  $k$  and for every possible choice of support points in  $\Omega$ , constitutes the "spatial law" of the random function  $Z(x)$ .

A random function  $Z(x)$  defined on  $\Omega$  is *stationary*, in the strict sense, if its spatial law is invariant under translation. More precisely, any two  $k$ -component vectorial random variables  $\{Z(x_1), \dots, Z(x_k)\}$  and  $\{Z(x_1 + h), \dots, Z(x_k + h)\}$  are identical in law (have the same  $k$ -variable distribution law) regardless of the translation vector  $h$  for all  $k = 1, 2, 3, \dots$ . However, in linear geostatistics, as we are only interested in the first two moments, it will be enough to assume first that these moments exist, and then to limit the stationarity assumptions to them.

A random function  $Z(x)$  defined on  $\Omega$  is *stationary of order 2* if (i) the mathematical expectation  $E\{Z(x)\}$  exists and does not depend on the support point  $x$ , i.e., there is a constant  $m$  such that

$$(2.1) \quad E\{Z(x)\} = m \quad \text{for all } x \in \Omega,$$

and (ii) for each pair of random variables  $\{Z(x), Z(x+h)\}$  the covariance  $C$  exists and depends on the separation vector  $h$  with  $x+h \in \Omega$ , i.e.,

$$(2.2) \quad C(h) = E\{Z(x+h)Z(x)\} - m^2$$

for all  $x \in \Omega$ .

The stationarity of the covariance implies the stationarity of the variance and the variogram. The following relations are immediately evident:

$$(2.3) \quad \begin{aligned} \text{Var}\{Z(x)\} &= E\{[Z(x) - m]^2\} = C(0), \\ \gamma(h) &= \frac{1}{2}E\{[Z(x+h) - Z(x)]^2\} = C(0) - C(h) \end{aligned}$$

for all  $x \in \Omega$ , where  $\text{Var}$  is the variance of  $Z(x)$  and  $\gamma(h)$  is called the *semivariogram* or the *intrinsic function*. Thus, under the hypothesis of second-order stationarity, the covariance and the variogram are two equivalent tools for  $Z(x)$  separated by a vector  $h$ . We can also define a third tool, the *correlogram*:

$$(2.4) \quad \rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}.$$

Then the covariance  $C(h)$  satisfies [13] that

$$(2.5) \quad \begin{aligned} \text{(i)} \quad & C(0) = \text{Var}\{Z(x)\} \geq 0, \\ \text{(ii)} \quad & C(h) = C(-h), \\ \text{(iii)} \quad & |C(h)| \leq C(0) \quad (\text{Schwartz inequality}). \end{aligned}$$

On the other hand, a random function  $Z(x)$  satisfies the (*geostatistical intrinsic hypothesis* [17] if (i) the mathematical expectation exists and does not depend on the support point  $x$ , and (ii') for all vectors  $h$  with  $x+h \in \Omega$  the increment  $[Z(x+h) - Z(x)]^2$  has a finite variance which does not depend on  $x$ , i.e.,

$$(2.6) \quad \text{Var}\{Z(x+h) - Z(x)\} = E\{[Z(x+h) - Z(x)]^2\} = 2\gamma(h)$$

for all  $x \in \Omega$ .

Note that the second-order stationarity assumes the existence of a covariance and, thus, of a finite *a priori* variance,  $Var\{Z(x)\} = C(0)$ . Thus, the existence of the semivariogram function  $\gamma(h)$  represents a weaker hypothesis than the existence of the covariance, i.e., the second-order stationary implies the intrinsic hypothesis but the converse is not true. The intrinsic hypothesis can also be seen as the limitation of the second-order stationarity to the increments of the random function  $Z(x)$ . A random function satisfying the intrinsic hypothesis has the first and second moments of the difference  $Z(x+h) - Z(x)$  which depend only on the separation vector  $h$  of the two points  $x+h$  and  $x$ , and not on their individual locations.

## 2.2. Drift

In intrinsic theory, we assume that random functions satisfy the intrinsic hypothesis. But, in practice, many regionalized variables or random variables show anisotropic behaviors due to the underlying geological nature. To overcome this difficulty several methods have been proposed. Among them, the most common way is to consider the drift. For a random function  $Z(x)$  the *drift*  $M(x)$  is defined by

$$(2.7) \quad M(x) = E\{Z(x)\}$$

and the *residual*  $Y(x)$  is defined by

$$(2.8) \quad Y(x) = Z(x) - M(x).$$

Thus, the drift at a point  $x$  is the expected value of the random variable  $Z$  at point  $x$ . Then it is easy to see that the residual  $Y(x)$  satisfies that

$$(2.9) \quad E\{Y(x)\} = 0.$$

The form of the drift  $M(x)$  depends on spatial data distribution. Often it is expressed as a linear combination of polynomials in the spatial coordinates. Also,  $M(x)$  may be a linear combination of any functions  $\{f_\ell : \ell = 1 \text{ to } k\}$  with the coefficients  $a_\ell$  of which are unknown, so that the drift  $M(x) = \sum_{\ell=1}^k a_\ell f_\ell(x)$  remains unknown.

### 2.3. Averaging process and ratio

The residuals obtained by subtracting the drift from the original data show more randomness compared with the original one. But, still, most residuals obtained from field data do not satisfy the intrinsic hypothesis. Thus, the experimental semivariograms applied to the residuals show erratic behaviors and fail to obtain correlation between spatial structure and variation of the residual distributions. These erratic behaviors are mainly due to the heterogeneity of the aquifer material and the anisotropic aspects remained in the residuals. To overcome this difficulty, we propose the concepts: *averaging process* and *ratio*.

Averaging process is to account for heterogeneity of the aquifer material of the given site and, thus, to obtain continuous information, in averaged sense, from the erratic data distribution. Two kinds of averaging process can be implemented: averaging over a certain size of spatial domain and averaging with a certain fixed number of samples. If samples are collected uniformly in space, these two methods are equivalent. But, since most sampling networks consist of irregularly spaced data, the averaging over a certain size of spatial domain would be better for many cases. On the other hand, ratio is to account for anisotropy effects due to different range of influence on each direction in data distributions. This ratio is a realization of the advection and dispersion or diffusion processes of groundwater in conjunction with the aquifer materials. To obtain the ratio between, for example, two different directions, choose the data range representing the main advection and dispersion or diffusion profiles showed in the data distributions in the considered region, and compare them. In Section 4, we show how these two concepts contribute to obtaining improved semivariograms.

### 2.4. Experimental semivariogram, mathematical model

The semivariograms defined by (2.3) or (2.6) are theoretical ones. Under the second order stationary assumption, (2.3) and (2.6) are equivalent. In this case,  $C(0) = \lim_{|h| \rightarrow \infty} \gamma(h)$ . But, without any assumption on existence of the covariance  $C(0)$ , i.e., under the intrinsic assumption, only (2.6) is well-defined. In practice, only a finite number of samples are given from the investigation area. Thus, we



consider a discrete version of the semivariogram, called an *experimental semivariogram*.

Suppose that we are given a set of  $n$  data values  $\{Z(x_i) : x_i \in \Omega, 1 \leq i \leq n\}$ . The (*classical*) *experimental semivariogram*  $\gamma(h)$  is defined by the arithmetic mean of the squared difference  $[Z(x_i) - Z(x_i + h)]$  [1,14], that is,

$$(2.10) \quad \gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} |Z(x_i) - Z(x_i + h)|^2,$$

where  $N(h)$  is the number of data pairs separated by the vector  $h$ . Note that the classical estimator (2.10) is not robust since the influence of outliers on the semivariogram increases as  $|Z(x) - Z(x+h)|^2$  increases. To treat the data with outliers many robust estimators have been developed (see, for example, [9,11,15]).

The experimental semivariogram is known only at discrete points. The discrete experimental semivariogram should be modelled by a continuous function, a *mathematical model*, that can be evaluated for any desired vector  $h$  because the experimental variogram cannot provide variance for every *lag*  $h$ . Since any anisotropic aspects of variance distributions can be treated as drift or other concepts such as ratio introduced in the previous section, most mathematical variogram models are described as functions of variable  $h = |h|$ , the radius of  $h$ , not the vector  $h$  itself. This implies that the mathematical variogram models depend only on the distance between sample locations not on individual sample location. The *spherical model*

$$\gamma(h) = \begin{cases} c_0 + \beta(\frac{3h}{2\alpha} - \frac{h^3}{2\alpha^3}), & h \leq \alpha, \\ c_0 + \beta, & h > \alpha, \end{cases}$$

the *exponential model*

$$\gamma(h) = c_0 + \beta(1 - \exp(-\frac{h}{\alpha}))$$

are commonly used [8,13], where the parameters  $c_0$  is the *nugget effect*,  $c_0 + \beta$  is the *sill* value, and  $\alpha$  is the distance  $h$  at which variogram

reaches the sill value. Recall that the theoretical or experimental semivariograms defined by (2.3), (2.6) or (2.10) satisfy that  $\gamma(0) = 0$ . But, in practice, most experimental semivariograms show the nugget effects due to the complexity and heterogeneity of the field aquifer materials, and, thus naturally, the mathematical models fitted to those semivariograms have the nugget effects. However, these nugget effects do not violated the nonsingular and symmetric nature of the corresponding kriging system (see Section 3). The sill value comes from the second order stationary assumption, i.e., the existence of the covariance  $C(0) = \lim_{h \rightarrow \infty} \gamma(h)$ . The *polynomial model*

$$(2.11) \quad \gamma(h) = \sum_{i=0}^m a_i h^i$$

is also commonly used. Note that, in this model, we do not assume that the existence of the limit  $\lim_{h \rightarrow \infty} \gamma(h)$ . This case may satisfy the intrinsic assumption but not the second-order stationary one. The choice of mathematical model depends strongly on data distribution represented by the experimental semivariogram.

### 3. Kriging

Kriging is an interpolation method to estimate values at unmeasured locations. It uses information from the mathematical model fitted to the experimental semivariogram to find an optimal set of weights. The kriging is based on the intrinsic assumption.

Consider the problem of estimating the unknown  $Z(x_o)$  at  $x_o \in \Omega$  from a set of  $n$  data values  $\{Z(x_i) : x_i \in \Omega, 1 \leq i \leq n\}$ . Suppose that the estimator  $Z^*(x_o)$  is a function of the available data:

$$Z^*(x_o) = f(Z(x_1), \dots, Z(x_n)).$$

Then it must be a function such that

- (i) it satisfies the non-bias condition,

$$E\{Z(x_o) - Z^*(x_o)\} = 0;$$

(ii) it is reasonably simple, so as to permit the calculation of the estimation variance

$$\begin{aligned} \sigma_E^2 &= E\{[Z(x_o) - Z^*(x_o)]^2\} \\ &= E\{(Z(x_o))^2\} + E\{(Z^*(x_o))^2\} - 2E\{Z(x_o)Z^*(x_o)\}. \end{aligned}$$

For any function  $f$ , the calculation of the first-order moment  $E\{Z(x_o) - Z^*(x_o)\}$  and the various second-order moments of the expression of  $\sigma_E^2$  requires that the  $n$ -variable distribution  $\{Z(x_1), \dots, Z(x_n)\}$  be known. However, since it is generally not possible to infer this distribution from a unique realization of the random variable  $Z(x_o)$ , we will consider, in this paper, the following form of estimator, a linear combination of the available data,

$$(3.1) \quad Z^*(x_o) = \sum_{i=1}^n w_i Z(x_i), \quad w_i \in \mathbf{R}^+,$$

where  $w_i$ 's are called the *weights* on the data and  $\mathbf{R}^+$  is the set of all positive real numbers.

Optimization of the statistic  $Z^*(x_o)$  can be performed by imposing the following two constraints

$$(3.2) \quad E\{Z^*(x_o) - Z(x_o)\} = 0,$$

$$(3.3) \quad E\{[Z^*(x_o) - Z(x_o)]^2\} \text{ is a minimum with respect to } w_i.$$

These restrictions imply that the difference  $Z^*(x_o) - Z(x_o)$  is unbiased and the variance of this difference is a minimum. Using (3.1) and (3.2), it can be written

$$(3.4) \quad E\left\{\sum_{i=1}^n w_i Z(x_i) - Z(x_o)\right\} = 0.$$

Taking the expectation of each value and equating it to the mean,  $m$ , which is assumed to be constant, yield

$$(3.5) \quad \sum_{i=1}^n w_i E\{Z(x_i)\} - E\{Z(x_o)\} = \sum_{i=1}^n w_i m - m = 0$$

resulting in

$$(3.6) \quad \sum_{i=1}^n w_i = 1.$$

The *punctual kriging* is most commonly used in practice. It produces the estimated values which are exactly same as sample values at measured locations, and, also, minimize the variance of the estimated values at unmeasured locations. Specifically, it is a minimization procedure for (3.3) which is carried out over  $(w_1, w_2, \dots, w_n)$  subject to  $\sum_{i=1}^n w_i = 1$ . That is, minimize

$$(3.7) \quad E\left\{ \left[ \sum_{i=1}^n w_i Z(x_i) - Z(x_o) \right]^2 \right\} - 2\lambda \left( \sum_{i=1}^n w_i - 1 \right)$$

with respect to  $w_1, w_2, \dots, w_n$  and  $\lambda$  (the parameter  $\lambda$  is a *Lagrange multiplier* [10,12] that ensures  $\sum_{i=1}^n w_i = 1$ ). Now the condition  $\sum_{i=1}^n w_i = 1$  implies that

$$(3.8) \quad \left[ \sum_{i=1}^n w_i Z(x_i) - Z(x_o) \right]^2 \\ = - \sum_{i=1}^n \sum_{j=1}^n w_i w_j (Z(x_i) - Z(x_j))^2 / 2 + 2 \sum_{i=1}^n w_i (Z(x_o) - Z(x_i))^2 / 2.$$

So that (3.7) becomes

$$(3.9) \quad - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma(x_i - x_j) + 2 \sum_{i=1}^n w_i \gamma(x_o - x_i) - 2\lambda \left( \sum_{i=1}^n w_i - 1 \right),$$

where  $\gamma(h)$  is the value obtained from the fitted mathematical model function for the experimental semivariogram. By differentiating (3.9) with respect to  $w_1, w_2, \dots, w_n, \lambda$ , we have the following optimality conditions

$$(3.10) \quad -2 \sum_{j=1}^n w_j \gamma(x_i - x_j) + 2\gamma(x_o - x_i) - 2\lambda = 0, \quad i = 1, 2, \dots, n, \\ \sum_{i=1}^n w_i = 1.$$

Thus, the punctual kriging system is represented by the following matrix form

$$(3.11) \quad AW = b,$$

where

$$(3.12) \quad A = \begin{bmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \gamma(\mathbf{x}_1 - \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \gamma(\mathbf{x}_2 - \mathbf{x}_1) & \gamma(\mathbf{x}_2 - \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_2 - \mathbf{x}_n) & 1 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \gamma(\mathbf{x}_n - \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix},$$

$$W = [w_1, w_2, \dots, w_n, \lambda]^T,$$

$$b = [\gamma(\mathbf{x}_o - \mathbf{x}_1), \gamma(\mathbf{x}_o - \mathbf{x}_2), \dots, \gamma(\mathbf{x}_o - \mathbf{x}_n), 1]^T.$$

Note that  $A$  is an  $(n + 1) \times (n + 1)$  nonsingular symmetric matrix. The optimal weights  $w_1, w_2, \dots, w_n$  can be obtained from  $W = A^{-1}b$ . The resulting estimation variance of punctual kriging becomes

$$(3.13) \quad \sigma^2 = \sum_{i=1}^n w_i \gamma(\mathbf{x}_o - \mathbf{x}_i) + \lambda - \gamma(\mathbf{x}_o - \mathbf{x}_o).$$

### 4. Application

In this section, we apply the kriging method to a sample data set collected during a field experiment [5,16]. All computations were performed on a SUN SPARC Station 20 and the algorithms needed for our analyses were developed under the MATLAB environment. The data set selected for our analyses consists of 26 groundwater level elevations measured in the deep level of the test site aquifer. Since the size of data set was not large, we did not try to develop efficient numerical algorithms needed in optimization procedures and solving kriging system. Instead, we developed the numerical codes based on well-known algorithms such as the Conjugate Gradient Method or Levenberg-Marquardt method [10,12].

The aquifer at the test site consists of a shallow alluvial terrace deposit averaging approximately  $11m$  in thickness. The aquifer is composed of poorly-sorted to well-sorted sandy gravel and gravelly sand with minor amounts of silt and clay. Sediments are generally unconsolidated, and occur as irregular horizontal or nearly horizontal lenses and layers. Marine sediments belong to the Eutaw Formation and consisting of clays, silts, and fine-grained sands form an aquitard beneath this alluvial aquifer. More details on the test site and other related previous experiments, see [1,4,6,19].

Table 1 shows the 26 samples with locations and water level elevations in meter scale. The spatial ranges of the samples were  $-86.10m \leq x \leq 101.69m$  and  $-49.01m \leq y \leq 263.96m$ . The averaged screening depths of the wells were ranged from  $55.51m$  to  $57.93m$  above the mean sea level. The screening depths of the wells W6, W11 and W15 (see Table 1) were not reported, but they were assumed to be within the similar ranges as other wells since the wells were installed to measure water level elevations of the deep area of the same aquifer. The water level elevation values were ranged from  $63.58m$  to  $65.03m$  above the mean sea level. The mean and variance of the water levels were  $64.5858$  and  $0.1040$ , respectively (see Table 2).

First, we need to analyze whether the 26 samples satisfy the geostatistical intrinsic assumption described in Section 2 before applying kriging. Figure 1 shows the water level distributions along each direction. It is easy to observe from Figure 1 that the water level elevations have apparent correlation between  $x$  and  $y$  coordinates and the heights: approximately linear relation in both directions. Thus, the samples do not satisfy geostatistical assumption and, hence, we cannot apply punctual kriging directly to our samples. Based on the observation of water level distribution profiles in Figure 1, the linear polynomial  $a + bx + cy$  was chosen for estimating the drift. The parameters  $a, b, c$  in the linear drift were estimated by solving the following minimization problem :

$$\text{Minimize}_{a,b,c} \sum_{i=1}^{26} [a + bx_i + cy_i - Z(x_i)]^2$$

using the Conjugate Gradient Method (see, for example, [10,12]), where each  $Z(x_i)$  is the measured water level elevation at the location  $x_i =$

$(x_i, y_i)$ . The estimated parameters were  $a = 64.7993$ ,  $b = 0.0013$  and  $c = -0.0025$ .

The residuals were obtained by subtracting the drift from the original data. Figure 2 shows the field data (+) the estimated drift(o) in each direction. We observed that the estimated drifts were affected by the two samples, W9 and W15 (see Table 1), whose values are much lower than others. Figure 3 shows the distribution of the residuals at measured locations. It is easy to see that randomness of the residual distributions has increased significantly compared with that showed in the original data (compare with Figure 1). More precisely, the mean and variance of the residuals were 0 and 0.0565, respectively. Thus, by considering the drift, the variance of the residuals was reduced approximately 46% compared to that of the original data.

Kriging was applied for the residuals to estimate expected residual values at unmeasured locations. Experimental semivariograms were estimated from the residuals by using the classical estimator given by (2.10). However, they did not show any apparent correlation between lags  $h$  and the corresponding variogram values due to heterogeneity of the aquifer (see Figure 4(a)). To obtain a reasonable correlation, the concepts of averaging process and ratio introduced in Section 2 were implemented (Figure 4(b)).

Several averaging procedures with different distances were applied for  $\gamma(h)$ . The averaged semivariograms were more continuous than that of discrete one. Also, we observed that the averaged experimental variograms showed significantly improved correlation between variances and lags  $h$  compared to those of without any averaging processes. However, there was a difficulty in fitting a mathematical model to the averaged experimental variogram due to the anisotropy effects on data distributions. To account for the anisotropy effects on correlation, we applied several different ratios and found out that the ratio  $x : y = 3 : 1$  and the averaged distance  $d = 5m$  produced a reasonable correlation. Figure 4(b) shows the improved semivariogram with the averaged distance  $d = 5$  and the ratio  $x : y = 3 : 1$  together with the fitted mathematical model. The linear model  $\gamma(h) = c_0 + \alpha h$  was chosen for the mathematical model since the profile showed by the semivariogram was approximately linear (see Figure 4(b)). The estimated parameters were  $c_0 = 0.0337$ , the nugget effect, and  $\alpha = 2.0569 \times 10^{-4}$ . The sill

was not observed.

We estimated the residuals at the lattice points separated by  $5m$  in the range of  $-100m \leq x \leq 150m$  and  $-100m \leq y \leq 300m$  by kriging. The final estimated water level elevations at those points were obtained by summing the estimated residuals and the drifts. Also, we observed that the estimation variances were higher in near boundary than those inside of the domain due to boundary effects. This observation suggests an investigation for developing estimation methods which can reduce the boundary effects. Figures 5 and 6 show the contour and the three-dimensional plots of the estimated water level elevations, respectively. It is easy to see, from Figures 5 and 6, that the groundwater of the studied area flows from right to left in  $x$ -direction and from near to far field in  $y$ -direction. Recall that the measured groundwater level at the sample W9 located at  $(x, y) = (20.84, 6.80)$  was much lower than nearby samples. This shows that there may be a possible sink source around the location.

Table 1. Water level

sample	$x(m)$	$y(m)$	$z(m)$	water level
W1	-86.10	2.56	57.73	64.89
W2	-11.41	83.28	55.77	64.53
W3	1.62	56.17	55.92	64.68
W4	-30.84	10.38	56.25	64.71
W5	-85.85	190.51	55.53	64.13
W6	95.13	123.74	-	64.75
W7	101.69	6.15	55.58	64.70
W8	-20.20	-49.01	55.98	64.82
W9	20.84	6.80	56.61	63.84
W10	-24.76	66.37	57.69	64.76
W11	3.14	147.64	-	64.54
W12	34.16	-24.73	55.83	65.03
W13	6.16	70.25	55.51	64.73
W14	44.65	93.42	56.71	64.74
W15	-59.79	67.92	-	64.62
W16	50.81	49.03	55.74	64.88
W17	-5.65	-8.82	57.93	64.94



W18	-18.59	35.34	56.01	64.72
W19	29.03	24.44	56.22	64.82
W20	-31.39	111.92	55.56	64.59
W21	30.07	125.92	55.87	64.57
W22	-67.73	140.19	55.77	64.41
W23	-24.52	178.32	55.83	64.35
W24	26.25	226.15	55.99	64.33
W25	-47.80	263.96	56.04	63.58
W26	51.93	173.77	55.98	64.57

Table 2. Mean and variance

mean (given samples)	64.5858
$\sigma^2$ (given samples)	0.1040
mean (residuals)	0
$\sigma^2$ (residuals)	0.0565

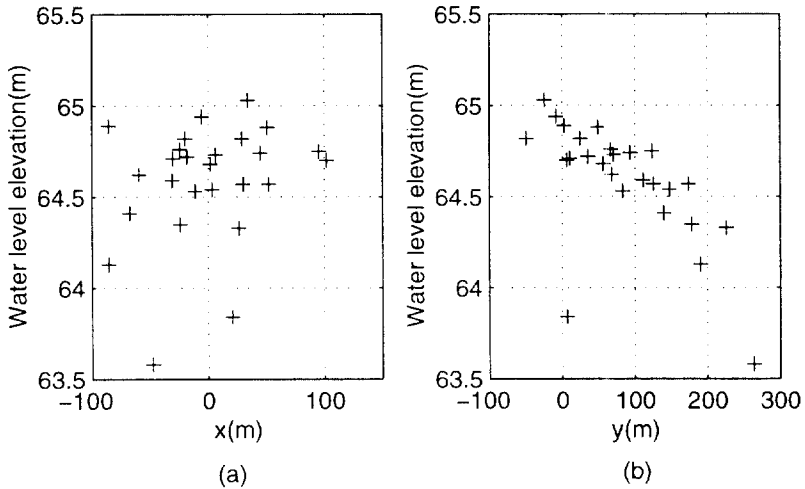


Fig. 1. Distribution of water level.

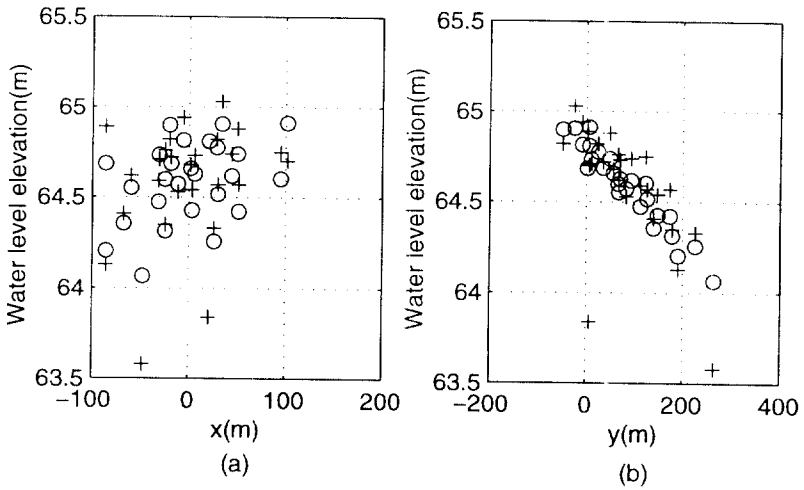


Fig. 2. Estimated drifts (field data(+); drift(o))  
(linear drift= $64.7993 + 0.0013x - 0.0025y$ )

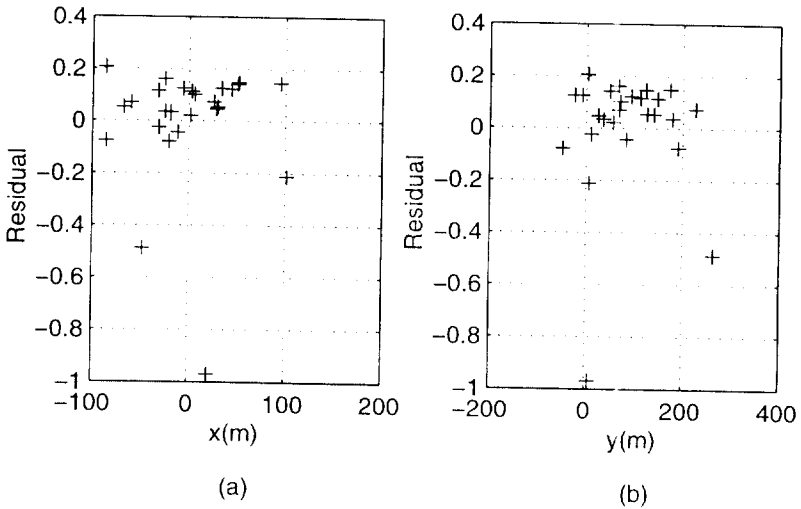


Fig. 3. Residuals.

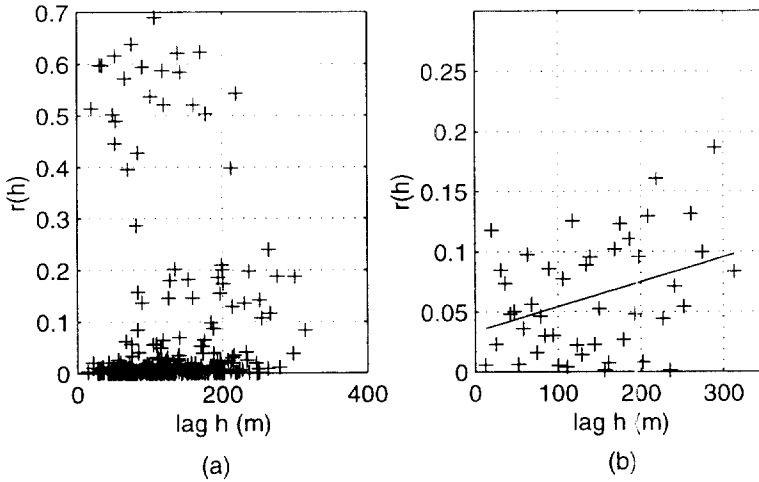


Fig. 4. Semivariograms and the fitted mathematical model for the residuals ((a) by classical estimator; (b) with averaging and ratio:  $d = 5$ ,  $x : y = 3 : 1$ , fitted linear model(—)= $0.0337 + 0.00020569h$ ).

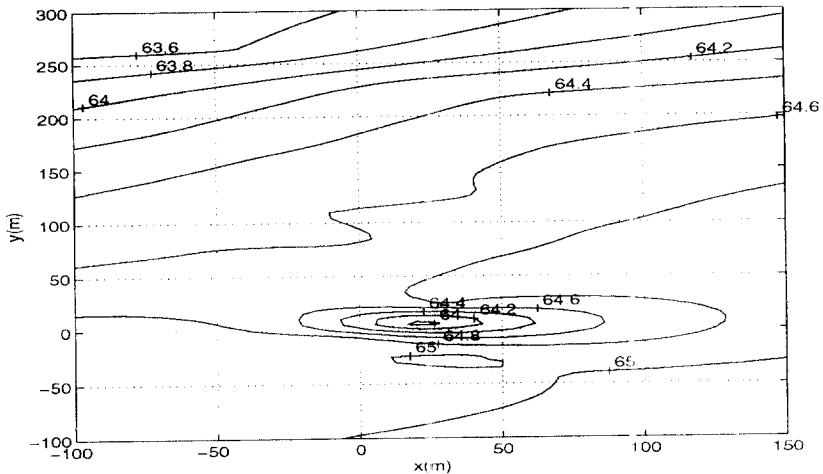


Fig. 5. Contour of the estimated water levels.

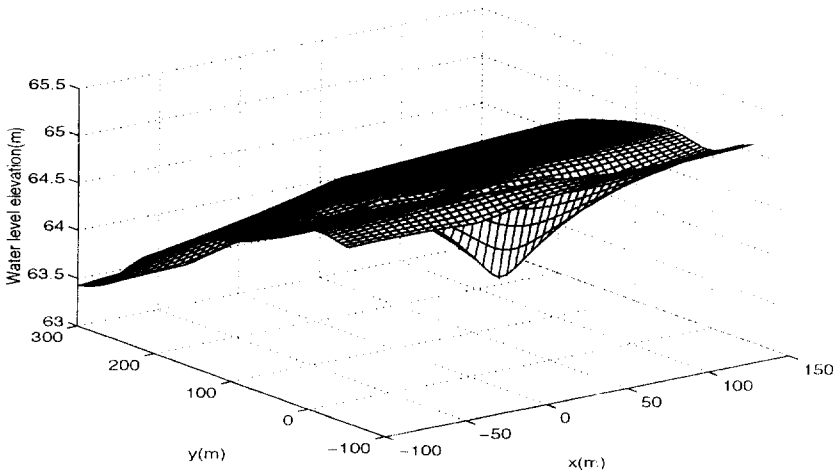


Fig. 6. Three-dimensional plot of the estimated water levels.

## References

- [1] E. E. Adams and L. W. Gelhar, *Field study of dispersion in a heterogeneous aquifer, 2. Spatial moments analysis*, Water Resour. Res. **28**(12) (1992), 3293-3307.
- [2] ASCE Task Committee on Geostatistical Techniques in Geohydrology of the Ground Water Hydrology Committee of the ASCE Hydraulics Division, *Review of Geostatistics in Geohydrology. I. Basic concepts; II. Applications*, Hydraulic Engineering **116**(5) (1990), 612-658; 633-658.
- [3] R. A. Blais and P. A. Carlier, *Applications of geostatistics in ore evaluation*, Canadian Inst. of Mining and Metallurgy, Ore Reserve Estimation and Grade Controls **9** (1967), 41-68.
- [4] J. M. Boggs and E. E. Adams, *Field study of dispersion in a heterogeneous aquifer, 4. Investigation of adsorption and sampling bias*, Water Resour. Res. **28**(12) (1992), 3325-3336.
- [5] J. M. Boggs, L. M. Beard, W. R. Waldrop, T. B. Stauffer, W. G. MacIntyre and C. P. Antworth, *Transport of tritium and four organic compounds during a natural gradient experiment (MADE-2)*, EPRI Report TR-101998, Electric Power Research Institute, Palo Alto, CA 94304, 1993.
- [6] J. M. Boggs, S. C. Young, L. M. Beard, L. W. Gelhar, K. R. Rehfeldt and E. E. Adams, *Field study of dispersion in a heterogeneous aquifer, 1. Overview*

- and site description*, Water Resour. Res. **28**(12) (1992), 3281-3291.
- [7] R. M. Cooper and J. D. Istok, *Geostatistics applied to groundwater contamination. I. Methodology; II. Application; III. Global estimates*, J. Environmental Engineering **114** (1988), 270-286; 287-299; 915-928.
- [8] N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley and Sons, Inc., New York, 1991.
- [9] N. Cressie and D. M. Hawkins, *Robust estimation of the variogram: I*, Mathematical Geology **12**(2) (1980), 115-125.
- [10] J. E. Jr. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
- [11] P. A. Dowd, *The variogram and kriging: robust and resistant estimators*, NATO ASI Ser., Ser. C: Math. Phy. Sci. **122** (1984), 91-106.
- [12] R. Fletcher, *Practical Methods of Optimization*, 2nd Ed., John Wiley and Sons, Chichester, 1993.
- [13] A. G. Journel and J. C. Huijbregts, *Mining Geostatistics*, Academic Press, San Diego, 1993.
- [14] S. Kang, T. B. Stauffer and K. Hatfield, *A practical three-dimensional estimation method for in situ groundwater contaminants: Part I. Methodology; Part II. Application*, Preprints, 1994.
- [15] D. Li and L. W. Lake, *A moving semivariance estimator*, Water Resour. Res. **30**(5) (1994), 1479-1489.
- [16] W. G. MacIntyre, J. M. Boggs, C. P. Antworth and T. B. Stauffer, *Degradation kinetics of aromatic organic solutes introduced into a heterogeneous aquifer*, Water Resources Research **29**(12) (1993), 4045-4051.
- [17] G. Matheron, *The intrinsic random functions and their applications*, Adv. Appl. Prob. **5** (1973), 439-468.
- [18] R. A. Olea, *Optimum Mapping Techniques Using Regionalized Variable Theory*, Kansas Geological Survey Series on Spatial Analysis, No.2, 1975.
- [19] K. R. Rehfeldt, J. M. Boggs and L. W. Gelhar, *Field study of dispersion in a heterogeneous aquifer, 3. Geostatistical analysis of hydraulic conductivity*, Water Resour. Res. **28**(12) (1992), 3309-3324.

Department of Mathematics  
College of Natural Sciences  
Chosun University  
Kwangju 501-759, Korea  
E-mail: sungkang@chollian.dacom.co.kr