HYDROLOGICAL PROCESSES
Hydrol. Process. **19**, 2023–2034 (2005)
Published online 22 December 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5664

Quantifying uncertainty of the semivariogram of transmissivity of an existing groundwater monitoring network

Gwo-Fong Lin* and Guo-Rong Chen
Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan

Abstract:

A method is presented for quantifying the uncertainty of the semivariogram of transmissivity and determining the required number of measurements. In this method, the estimated semivariogram and its 95% confidence limits are first determined from a finite number of measurements. The uncertainty of the estimated semivariogram is then quantified using the random field simulation technique. For a given value of the quantitative index of uncertainty, the required number of measured data can finally be obtained. Actual transmissivity data of an existing groundwater monitoring network are used in the application of the proposed method. The required numbers of measurements of transmissivity for four different values of the quantitative index of uncertainty are provided, from which reliable semivariograms of the transmissivity can be obtained. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS semivariogram; transmissivity; uncertainty quantification; random field simulation; groundwater monitoring network

INTRODUCTION

Natural permeable earth materials are highly heterogeneous in terms of their hydraulic properties, such as the transmissivity of aquifers (Gelhar, 1993). The spatial distribution of the transmissivity is often obtained based on sampled data using kriging techniques (Delhomme, 1979). In order to estimate the transmissivity by kriging, it is necessary to determine the semivariogram that describes the spatial variability structure of the transmissivity. The semivariogram has great influence on the estimation of transmissivity. In practice, the semivariogram is usually determined based on limited measured data at observation wells. Several experimental suggestions about the number of measured data have been made in the literature. For example, Journel and Huijbregts (1978) suggested at least 30 pairs of measurements for each lag distance are required, and Webster and Oliver (1992) suggested at least 150–200 total measurements are needed in order to ensure a reliable semivariogram.

There is still a need to quantify the uncertainty of the semivariogram. Russo and Jury (1987) presented an analysis of the uncertainty of the estimated semivariogram using Monte Carlo simulation. Shafer and Varljen (1990) applied the jackknife method to obtain the approximation of confidence limits of the estimated semivariogram.

In addition to the quantification of the uncertainty of the semivariogram, the size of data set needed to reduce the uncertainty is another important issue. To reduce the uncertainty of the semivariogram needs more measured data, and that is usually limited by the cost. The objective of this paper is to present a method for quantifying the uncertainty of the semivariogram and determining the required number of measurements for an existing groundwater monitoring network. First, a method is proposed which is capable of quantifying the uncertainty of the semivariogram using a random field generator. Then, the random field generator used herein

* Correspondence to: Gwo-Fong Lin, Dept of Civil Eng, National Taiwan University, Taipei, 10617 Taiwan. E-mail: gflin@ntu.edu.tw

Received 5 May 2003
Accepted 9 June 2004

is validated. Furthermore, application of the proposed method to the Pingtung Plain in southern Taiwan is performed, and the corresponding results are presented. Finally, conclusions are drawn for future consideration of the number of measurements of transmissivity.

DESCRIPTION OF THE QUANTITATIVE METHOD

Basic concepts and assumptions

In this paper, the spatial distribution of the hydraulic properties is regarded as a random field with particular statistics. The measured data at the given location $z(x)$, where x denotes the spatial coordinate, is a realization of the random field $Z(x)$. Random fields can be statistically characterized by two assumptions. One is the second-order stationarity and the other is the intrinsic hypothesis. The second-order stationarity satisfies

$$E[Z(x)] = m_Z \tag{1}$$

$$E[(Z(x) - m_Z)(Z(x+h) - m_Z)] = C_Z(h) \tag{2}$$

where E is the expectation operator, m_Z is the mean of the random field Z, and $C_Z(h)$ is the covariance function which depends on the separation h. It should be noted that h is a vector.

Semivariogram analysis

In order to estimate the spatial distribution of hydraulic properties, it is necessary to know the semivariogram. The semivariogram is defined as

$$\gamma_Z(h) = \frac{1}{2}E[Z(x+h) - Z(h)]^2 \tag{3}$$

Since only limited realizations can be obtained in practice, the ergodic hypothesis is usually taken. Thus the experimental semivariogram is calculated as

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[Z(x_i) - Z(x_i + h)]^2 \tag{4}$$

where $N(h)$ is the number of data pairs.

The experimental semivariogram should be fitted to a positive definite model. Several theoretical models have been proposed. In this paper, the experimental semivariogram is fitted to three commonly used models: Gaussian, exponential and spherical. These models are defined with two parameters, the correlation length a and the sill w, as follows.

1. Gaussian model

$$\gamma(h) = w\left[1 - \exp\left(-\frac{h^2}{a^2}\right)\right] \tag{5}$$

2. Exponential model

$$\gamma(h) = w\left[1 - \exp\left(-\frac{h}{a}\right)\right] \tag{6}$$

3. Spherical model

$$\gamma(h) = \begin{cases} w\left(\dfrac{3h}{2\alpha} - \dfrac{1h^3}{2\alpha^3}\right), & 0 \le h \le a \\ w, & h > a \end{cases} \tag{7}$$

Jackknife method

The jackknife method is a parameter estimation technique developed by Quenouille (1956). Shafer and Varljen (1990) used the jackknife method to estimate confidence limits of semivariograms from only one realization of a spatially correlated random field, and verified the accuracy and validity of this method. The entire sample data set is partitioned into g subgroups. Partition-dependent estimates of the semivariogram, $J_j[\hat{\gamma}(h)]$, are obtained as

$$J_j[\hat{\gamma}(h)] = g \cdot \hat{\gamma}_{\text{all}}(h) - (g-1) \cdot \hat{\gamma}_j(h), \quad j = 1, 2, 3, \ldots, g \tag{8}$$

where $\hat{\gamma}_{\text{all}}$ and $\hat{\gamma}_j$ are the semivariograms estimated by using all the data and all the data remaining after removing the jth partition, respectively.

The jackknife estimate of the semivariogram, $J[\hat{\gamma}(h)]$, is

$$J[\hat{\gamma}(h)] = \frac{1}{g} \sum_{j=1}^{g} J_j[\hat{\gamma}(h)] \tag{9}$$

The jackknife variance, $\sigma_J^2(h)$, is estimated as

$$\sigma_J^2(h) = \frac{1}{g(g-1)} \sum_{j=1}^{g} \{J_j[\hat{\gamma}(h)] - J[\hat{\gamma}(h)]\}^2 \tag{10}$$

Jackknife bands of $\pm 2\sigma_J$ approximate the 95% confidence limit of the estimated semivariogram (Shafer and Varljen, 1990).

Random field generation

In practice, the parameters of the semivariogram are usually determined based on finite measured data. Hence, to quantify the uncertainty of the semivariogram is an important task. Using simulation techniques, a large number of realizations with prescribed statistics can be generated. The generated realizations at given locations serve as the measured data at the observation wells.

Regarding the spatially distributed random field, several methods have been developed, for example, the turning band method (Mantoglou and Wilson, 1982; Bras and Rodriguez-Iturbe, 1985), the fast Fourier transform (Bartlett, 1975; Bastin *et al*., 1984) and the nearest neighbourhood method (Smith and Schwartz, 1980; King and Smith, 1988). Based on the sequential Gaussian simulation, Bellin and Rubin (1996) presented an algorithm, named HYDRO_GEN, which is capable of generating realizations with the prescribed statistics at given locations. Because the HYDRO_GEN algorithm is found to be accurate and extremely fast (Bellin and Rubin, 1996), it is used herein to generate realizations.

A quantitative index of uncertainty of the estimated semivariogram

The variance of the estimated semivariogram, $\sigma^2_{\hat{\gamma}(h)}$, can be written as

$$\sigma^2_{\hat{\gamma}(h)} = \frac{1}{N_s} \sum_{i=1}^{N_s} [\hat{\gamma}_i(h) - \gamma(h)]^2 \tag{11}$$

where N_s is the number of simulations, and $\gamma(h)$ and $\hat{\gamma}_i(h)$ are semivariograms estimated by using all the data from simulations and those only from the ith simulation, respectively.

In order to quantify the uncertainty of the estimated semivariogram, there is a need to define a quantitative index. For such a purpose, the relative standard error (SE) of the semivariogram is used herein. The SE is defined as

$$\mathrm{SE} = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{\sigma_{\hat{\gamma}(h)}}{\gamma(h)} \right) \tag{12}$$

where N_p is the total number of data pairs. The SE is calculated based on the experimental semivariogram and represents uncertainty in the semivariogram.

APPLICATION AND DISCUSSION

For applications of the proposed method, actual data of transmissivity in the Pingtung Plain in southern Taiwan are used. The Pingtung Plain, with an area of 1210 km^2, is the most southern groundwater region in Taiwan. There are 36 groundwater observation wells in the Pingtung Plain. Locations of groundwater observation wells are shown in Figure 1.

The transmissivity was measured only once at each observation well in 2000. To reduce the uncertainty of the estimated semivariogram, more measured data are needed. Under the present condition, the only way to

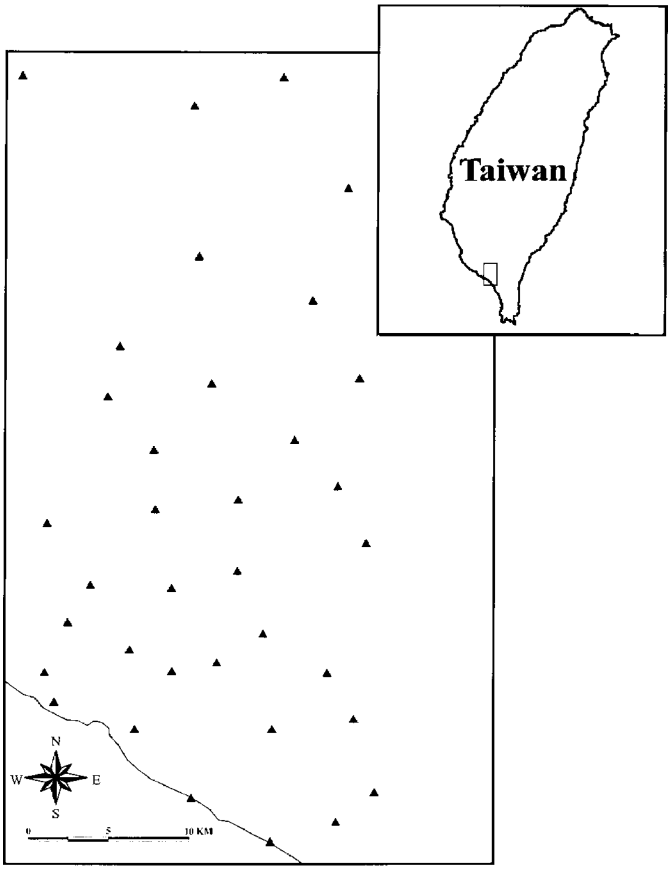

Figure 1. The study area and the locations of wells

reduce the uncertainty is to increase the number of measurements at each existing groundwater observation well. Determining the required number of measurements under a particular quantitative uncertainty criterion is an important task.

In this section, the proposed method is applied to quantify the uncertainty of the semivariogram and to determine the required numbers of measurements at each observation well for various relative standard errors. The results are useful in the sampling design of the groundwater monitoring network in the Pingtung Plain.

Statistical results of the measured data in the Pingtung Plain

The transmissivity data of the aquifer (which is about 25 m below the ground surface on average) in the Pingtung Plain are obtained based on pumping test data at these observation wells (Water Resources Bureau, 2000). Because only one datum at each groundwater observation well was measured, the ergodic hypothesis must be adopted. According to Delhomme (1979), the logarithms of point values of transmissivity are normally distributed. Thus the logarithm of transmissivity is used herein. The statistics of the measured data are listed in Table I. The measured log T ranges from 0·606 to 4·337. The chi-squared goodness-of-fit test is used to check whether log T is normally distributed at the 5% level of significance. The chi-squared test indicates that log T follows the normal distribution.

Experimental and fitted semivariograms

The experimental semivariogram is calculated and three theoretical models, Gaussian, exponential and spherical, are fitted to it. Table II summarizes the performance of the three fitted semivariograms. It shows that the exponential model has the best performance among the three theoretical semivariogram models. The fitted exponential model has a correlation length a of 7·8 km and a sill w of 0·66. The experimental and fitted semivariograms are presented in Figure 2.

Validation of the random field generator

For validation of the random field generator, second-order stationary random fields are generated and the simulated results are presented and discussed herein. The random field $Z(x)$ is characterized by its mean m_Z, variance σ_Z^2 and semivariogram $\gamma_Z(h)$. The exponential model (with $m_Z = 0$, $\sigma_Z^2 = 1$, $\omega = 1$ and $\alpha = 2$) is chosen as the semivariogram model. Two kinds of random fields are considered, namely 20×20 and 40×40 node. For each kind, 100 random fields are generated using the HYDRO_GEN algorithm. Then,

Table I. The statistics of the measured transmissivity T

Statistic	T (m^2/day)	Log T (log m^2/day)
Mean	2309·58	2·883
Standard deviation	4076·331	0·759
Minimum	4·032	0·606
Maximum	21754·08	4·337

Table II. The performance of the three fitted semivariograms

Type of model	Fitted correlation length a	Fitted sill w	RMSE
Gaussian	6·5	0·54	0·089
Exponential	7·8	0·66	0·055
Spherical	7·1	0·59	0·063

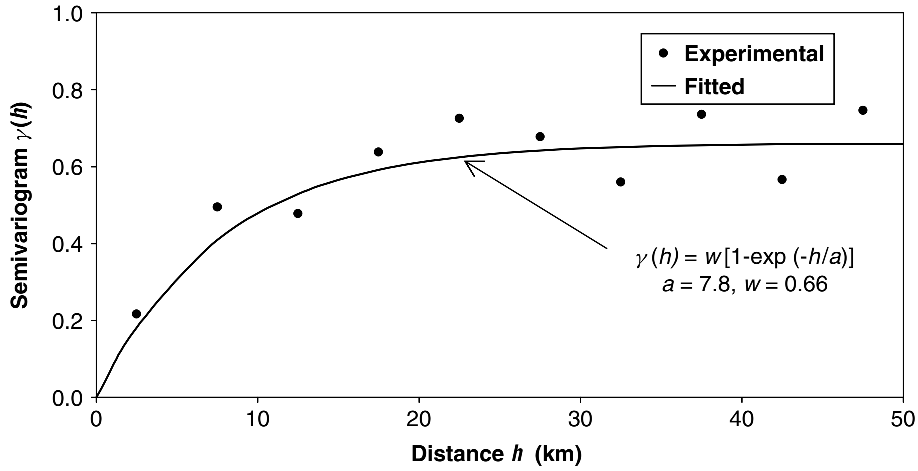

Figure 2. The experimental and fitted semivariograms

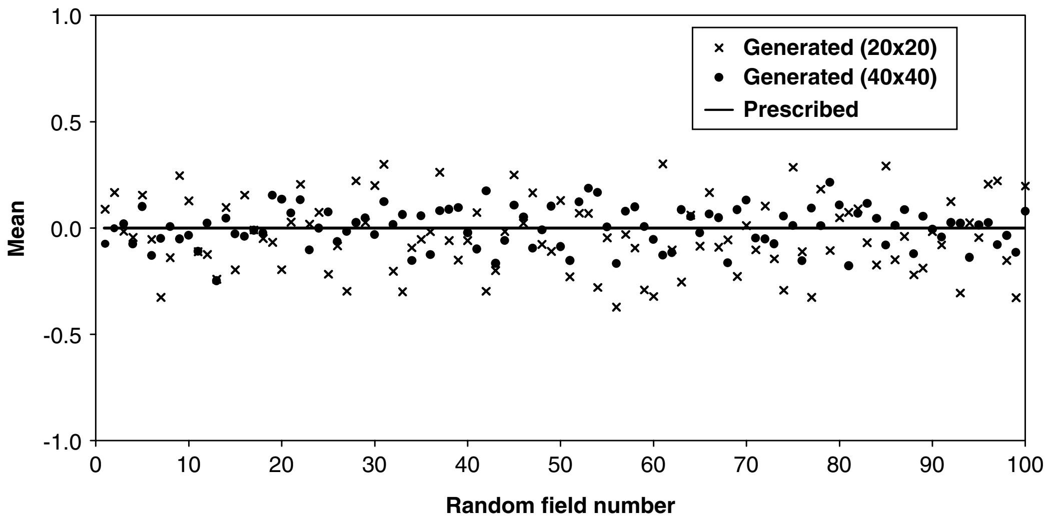

Figure 3. Comparison of the generated and prescribed means

the mean, variance and semivariogram of these two kinds of random fields are calculated and compared with the prescribed statistics. Figures 3, 4 and 5 compare the generated and prescribed means, variances and semivariograms, respectively. They show that the statistics of generated random fields are in good agreement with the prescribed statistics. The generated 40×40 node random fields are more capable of reproducing the prescribed statistics than the generated 20×20 node random fields due to the larger sample size.

Sensitivity analysis

Five cases with different combinations of parameter values are simulated in order to investigate the influence of the correlation length and the sill on the uncertainty of the estimated semivariogram. The first case uses a correlation length of 7·8 and a sill of 0·66, both of which are obtained from the measured data. The simulated results of Case 1 serve as a basis for comparison with those of the other cases. In Cases 2 and 3, the values of the correlation length are varied around the initial value of 7·8, using multiples of 50 and 150% of the initial

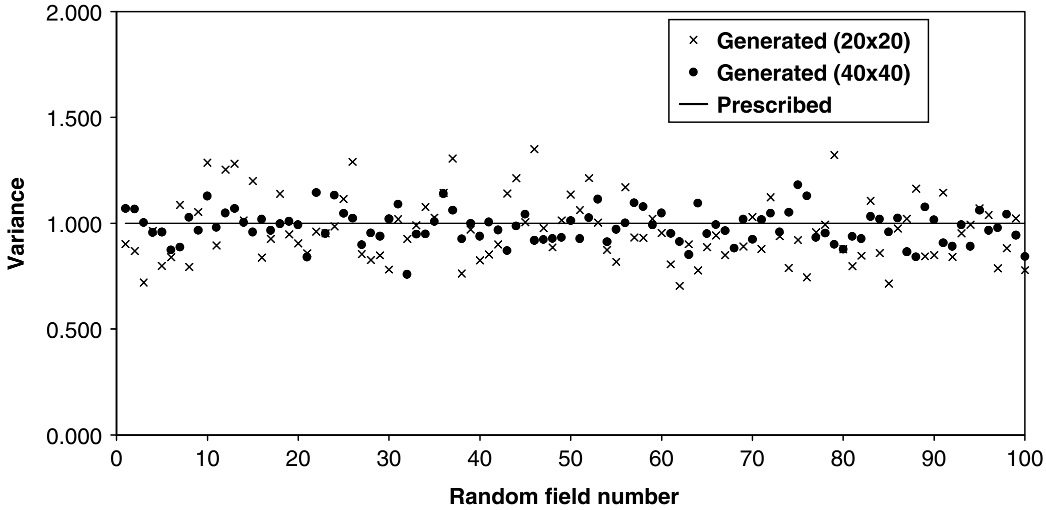

Figure 4. Comparison of the generated and prescribed variances

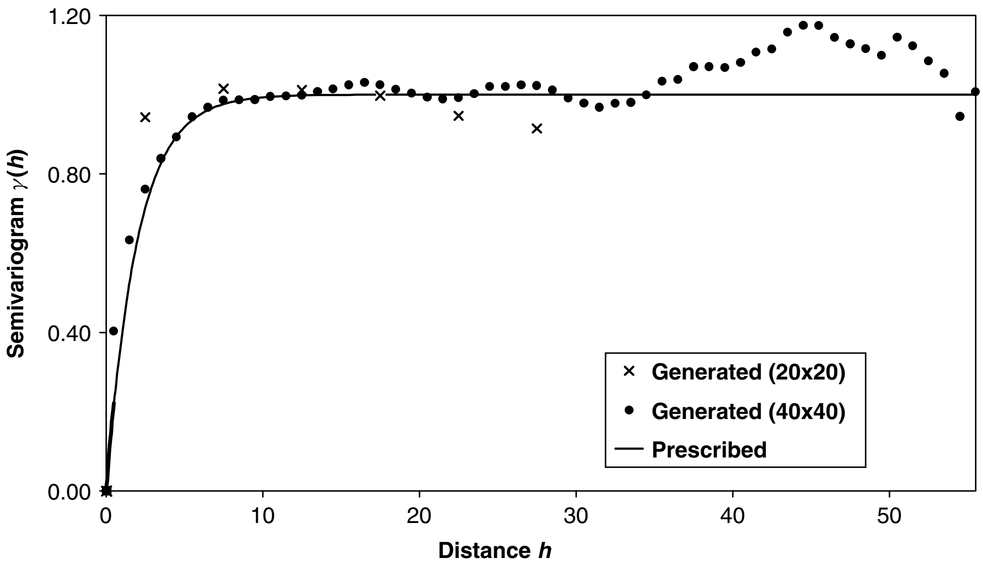

Figure 5. Comparison of the generated and prescribed semivariograms

value (i.e., 3·9 and 11·7), while holding the other parameters constant at the corresponding initial value. In the same way, the values of the sill in Cases 4 and 5 are varied around the initial value of 0·66, using multiples of 50 and 150% of the initial value (i.e., 0·33 and 0·99), while holding the other parameters constant at the corresponding initial value. Table III summarizes the parameter values used in simulations.

In each case, the numbers of measurements are varied from 2 to 50. For each number of measurements, the random field generator is used to generate 1000 sets of spatially correlated random fields and the relative standard errors of the estimated semivariograms are evaluated.

For Case 1 ($a = 7{\cdot}8$ and $w = 0{\cdot}66$), the simulated results show that the relative standard error of the estimated semivariogram decreases from 22·1% to 3·9% as the number of measurements increases from 2

Table III. The parameter values used in simulations

Parameter	Value				
	Case 1	Case 2	Case 3	Case 4	Case 5
Mean	2·883	2·883	2·883	2·883	2·883
Variance	0·66	0·66	0·66	0·33	0·99
Correlation length a	7·8	3·9	11·7	7·8	7·8
Sill w	0·66	0·66	0·66	0·33	0·99

to 25. It is found that increasing the number of measurements is a very effective method for reducing the uncertainty of the estimated semivariogram.

A comparison of the simulated results of Case 1 ($a = 7{\cdot}8$ and $w = 0{\cdot}66$) with those of Cases 2 ($a = 3{\cdot}9$ and $w = 0{\cdot}66$) and 3 ($a = 11{\cdot}7$ and $w = 0{\cdot}66$) is presented in Figure 6. It shows that both the standard deviation and the relative standard error of the estimated semivariogram decrease rapidly as the correlation length a increases. Figure 7 compares the simulated results of Case 1 ($a = 7{\cdot}8$ and $w = 0{\cdot}66$) with those of Cases 4 ($a = 7{\cdot}8$ and $w = 0{\cdot}33$) and 5 ($a = 7{\cdot}8$ and $w = 0{\cdot}99$). The standard deviation of the estimated semivariogram decreases with decreasing sill w, but the relative standard error of the estimated semivariogram is independent of sill w.

Jackknife analysis

Jackknife analysis is undertaken in order to estimate confidence limits of semivariograms from single actual data at each observation well in the Pingtung Plain. Maximum partitioning is used herein, because the computations are not exhaustive for the sample size of the Pingtung Plain and maximum partitioning always yields stable variance.

The jackknife estimate is used to approximate the semivariogram for log T based on the actual data. The $\pm 2\sigma_J(h)$ confidence limits of $J[\hat{\gamma}(h)]$ are also evaluated and shown in Figure 8. The upper and lower confidence limits have correlation lengths of 7·1 and 8·9, respectively.

Determination of the required number of measurements

On the basis of the jackknife analysis, the correlation length with a value of 7·1 and sill with a value of 0·66 are used in the following simulations to determine the required number of measurements for the existing groundwater monitoring network in the Pingtung Plain. The numbers of measurements are varied from 2 to 100. In a similar manner, the random field generator simulates 1000 sets of spatially correlated random fields for each number of measurements. Then, the relative standard errors of the estimated semivariograms are calculated.

In practice, we set various values of relative standard error in order to determine the required number of measurements. Figure 9 shows the required number of measurements for various relative standard errors. The required numbers of measurements for four different relative standard errors (10%, 7·5%, 5% and 2·5%) are calculated and summarized in Table IV. For the existing groundwater monitoring network in the Pingtung Plain, one needs 5, 7, 14 and 53 measurements at each observation well to achieve 10%, 7·5%, 5% and 2·5% relative standard errors, respectively, of the semivariogram of the transmissivity.

SUMMARY AND CONCLUSIONS

To estimate the transmissivity by kriging, determination of the semivariogram is an essential stage. The semivariogram has great influence on the estimated results, but it is usually determined based on limited

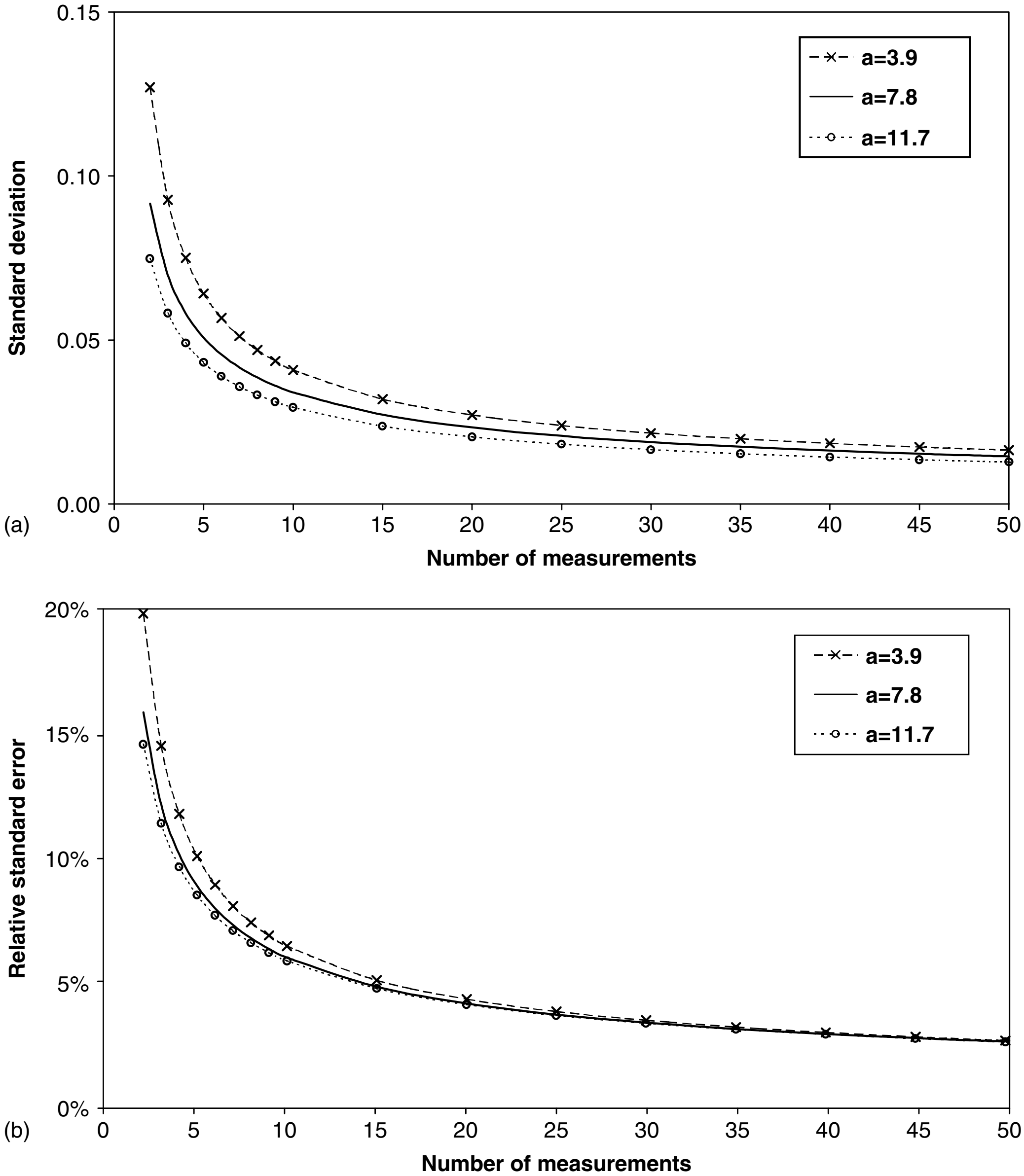

Figure 6. Number of measurements versus (a) standard deviation and (b) relative standard error of the semivariogram for three different correlation lengths

measured data at observation wells. Hence it is important to quantify the uncertainty of the estimated semivariogram.

In this paper, the relative standard error of the estimated semivariogram serves as a quantitative index of the uncertainty of the estimated semivariogram. A method, which is capable of quantifying the uncertainty of the estimated semivariogram using a random field generator, is presented. Actual application of the proposed method is performed. According to the transmissivity data of an existing groundwater monitoring network, five cases with different parameter values of the semivariogram model are simulated. Besides, the influence

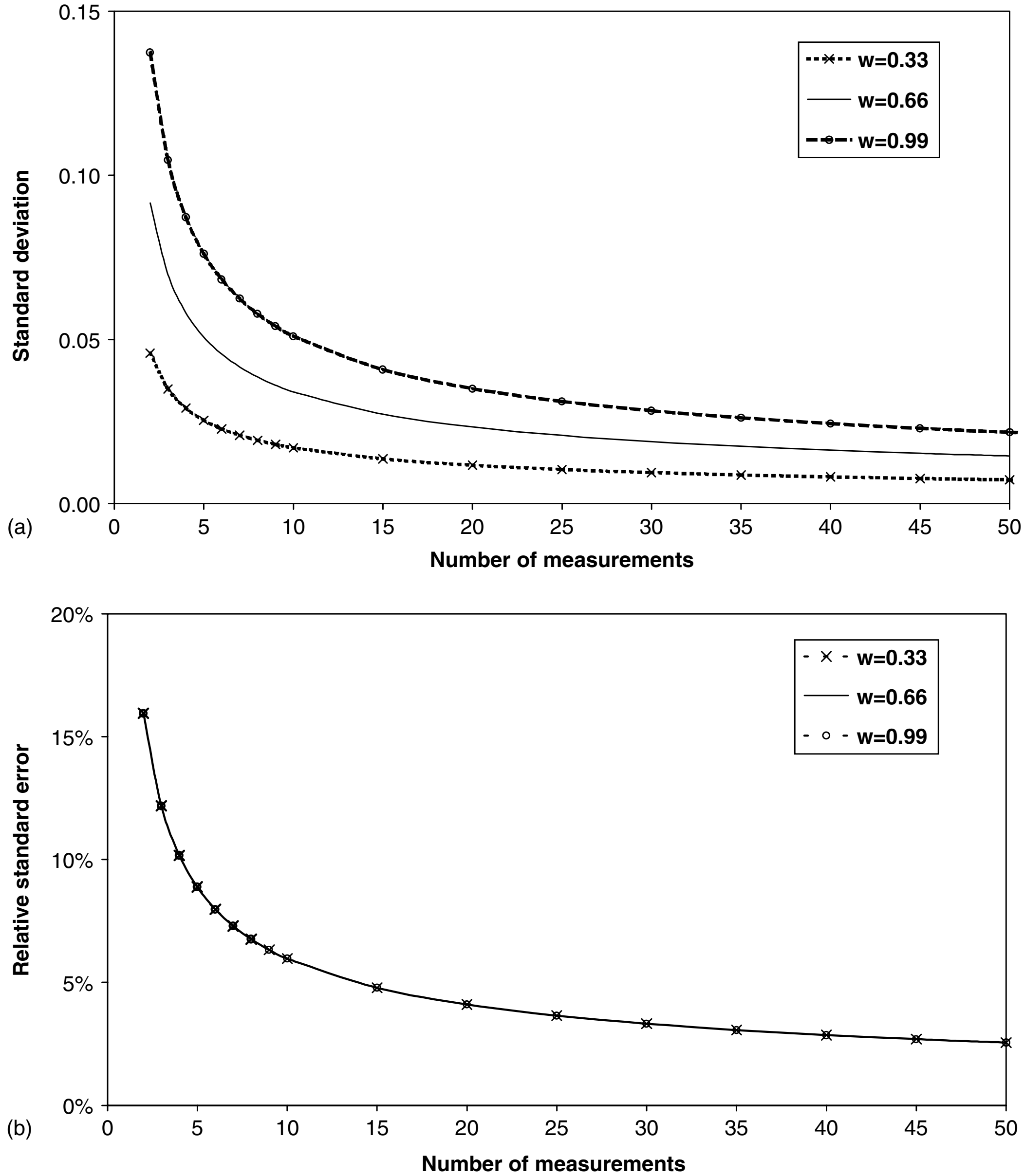

Figure 7. Number of measurements versus (a) standard deviation and (b) relative standard error of the semivariogram for three different sills

of parameter values on the uncertainty of the semivariogram is quantified and compared. Finally, the required numbers of measurements for different values of relative standard error are obtained. The simulated results show that increasing the number of measured data can effectively reduce the uncertainty of the estimated semivariogram. To obtain reliable semivariograms of the transmissivity for the groundwater monitoring network studied in this paper, one needs 5, 7, 14 and 53 measurements of transmissivity to achieve 10%, 7·5%, 5% and 2·5% relative standard errors, respectively.

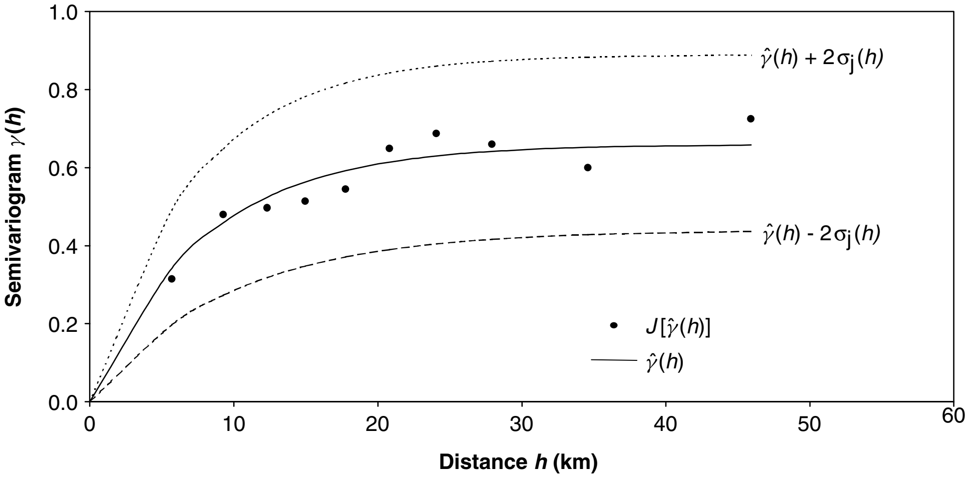

Figure 8. Jackknife estimate of the semivariogram with upper and lower confidence limits

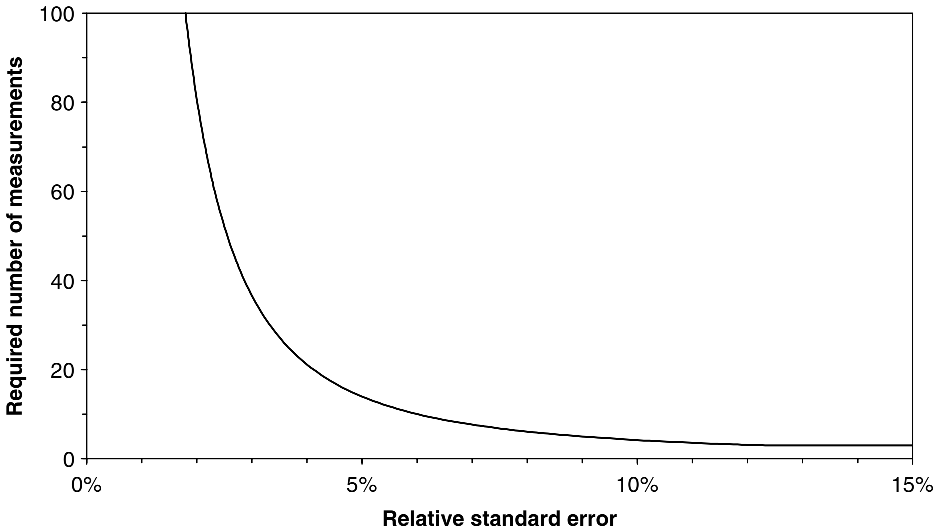

Figure 9. Required number of measurements versus relative standard error for the existing groundwater monitoring network in the Pingtung Plain

Table IV. Required number of measurements for four relative standard errors

Relative standard error (%)	Required number of measurements
10	5
7.5	7
5	14
2.5	53

REFERENCES

Bartlett MS. 1975. *The Statistical Analysis of Spatial Pattern*. Chapman and Hall: London; 90 pp.

Bastin G, Lorent B, Duque C, Gevers M. 1984. Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resources Research* **20**(4): 463–470.

Bellin A, Rubin Y. 1996. HYDRO_GEN: A spatially distributed random field generator for correlated properties. *Stochastic Hydrology and Hydraulics* **10**: 253–278.

Bras RL, Rodriquez-Iturbe I. 1985. *Random Functions and Hydrology*. Dover Publications: New York; 559 pp.

Delhomme JP. 1979. Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. *Water Resources Research* **15**(2): 269–280.

Gelhar LW. 1993. *Stochastic Subsurface Hydrology*. Prentice Hall: Englewood Cliffs, NJ; 390 pp.

Journel AG, Huijbregts CJ. 1978. *Mining Geostatistics*. Academic Press: New York; 600 pp.

King PR, Smith PJ. 1988. Generation of correlated properties in heterogeneous media. *Mathematical Geology* **20**(7): 863–877.

Mantoglou A, Wilson JL. 1982. The turning band method for simulation of random field using line generation by a spectral method. *Water Resources Research* **18**(5): 1379–1394.

Quenouille MH. 1956. Notes on bias in estimation. *Biometrika* **43**: 353–360.

Russo D, Jury WA. 1987. A theoretical study of the estimation of the correlation scale in spatial variable fields. 1. Stationary fields. *Water Resources Research* **23**(7): 1257–1268.

Shafer JM, Varljen MD. 1990. Approximation of confidence limits on sample semivariograms from single realizations of spatially correlated random fields. *Water Resources Research* **26**(8): 1787–1802.

Smith M, Schwartz FW. 1980. Mass transport 1, a stochastic analysis of macrodispersion. *Water Resources Research* **16**(2): 303–313.

Water Resources Bureau. 2000. *Groundwater quality of the Taiwan groundwater monitoring network*. Technical Report, Ministry of Economic Affairs, Taipei, Taiwan; 120 pp. (in Chinese).

Webster R, Oliver MA. 1992. Sample adequately to estimate variograms of soil properties. *Journal of Soil Science* **43**: 177–192.

