Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods

Chapter · June 2019

DOI: 10.1016/B978-0-12-815413-7.00004-3

5 authors, including:

Sang Yong Chung
Pukyong National University

Venkatramanan Senapathi
Alagappa University

Hussam Eldin Elzain

Selvam .S
V.O.Chidambaram College, Tuticorin

Chapter 4

Supplement of Missing Data in Groundwater-Level Variations of Peak Type Using Geostatistical Methods

Sang Yong Chung*, S. Venkatramanan*,†,‡,§, Hussam Eldin Elzain*, S. Selvam¶ and M.V. Prasanna‖

**Department of Earth and Environmental Sciences, Pukyong National University, Busan, South Korea* †*Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam* ‡*Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam* §*Department of Geology, Alagappa University, Karaikudi, India* ¶*Department of Geology, V.O. Chidambaram College, Thoothukudi, India* ‖*Department of Applied Geology, Faculty of Engineering and Science, Curtin University Malaysia, Miri, Malaysia*

Chapter Outline

4.1 INTRODUCTION

Many groundwater monitoring wells are installed for the effective conservation and management of groundwater. Groundwater level and quality are measured from automatic recorders in the monitoring wells. However, groundwater-level data are often missed because of power outage or digital-sensor problems. The missing data need to be interpolated using proper statistical methods, because they degrade the continuity of monitored data.

The use of geostatistics is necessary for the reproduction of missing data, because groundwater-levels show various changes and irregularities. Kriging can interpolate the missing intervals with the minimum errors in the case of small fluctuation in water-level data. Chung et al. (2001) reproduced missing data in a sinuous-type long-term groundwater-level with few errors using ordinary kriging.

Examples of kriging applications include the estimation of aquifer parameters (Loaiciga et al., 1996), the analysis of groundwater flow (Jensen et al., 1996), and the estimation of groundwater level and hydraulic gradient (Philip and Kitanidis, 1989). Recently kriging has been used for the optimization of groundwater-level observation networks (Theodossiou and Latinopoulos, 2006), the risk assessment of nitrate contamination (Hu et al., 2005), and the evaluation of arsenic-contamination potential (Liu et al., 2004).

In the case of irregularly fluctuating water-level data, kriging has limitations in reproducing the variability of water-level data. Conditional simulation can produce the irregular fluctuations with relatively few statistical errors (Journel and Huijbregts, 1978; Chiles and Delfiner, 1999). Chung and Wheatcraft (1993) showed that conditional simulation was far superior to kriging in the estimation of two-dimensional hydraulic conductivity distributions using hydraulic conductivity data of the Borden site in Canada. Conditional simulation is widely used for groundwater-flow modeling and contaminant-transport modeling through the simulation of hydraulic conductivity distribution (Gómez-Hernández et al., 1997; Capilla et al., 1997; Hendricks Franssen et al., 2003).

GIS and Geostatistical Techniques for Groundwater Science. https://doi.org/10.1016/B978-0-12-815413-7.00004-3

4.2 GROUNDWATER LEVEL DATA WITH A PEAK TYPE VARIATION

In this study, the groundwater-level data of alluvium and bedrock in the Daegu Bisan National Groundwater Monitoring Well, Korea were used to compare kriging with conditional simulation for the interpolation of missing data. Fig. 4.1 shows the variations of two groundwater-level data sources from the Daegu Bisan National Groundwater Monitoring Well. They have very irregular fluctuations and there are several missing intervals of groundwater level. Table 4.1 shows the general statistical values for the groundwater levels of alluvium and bedrock.

FIG. 4.1 Missing interval in groundwater levels of a groundwater monitoring well.

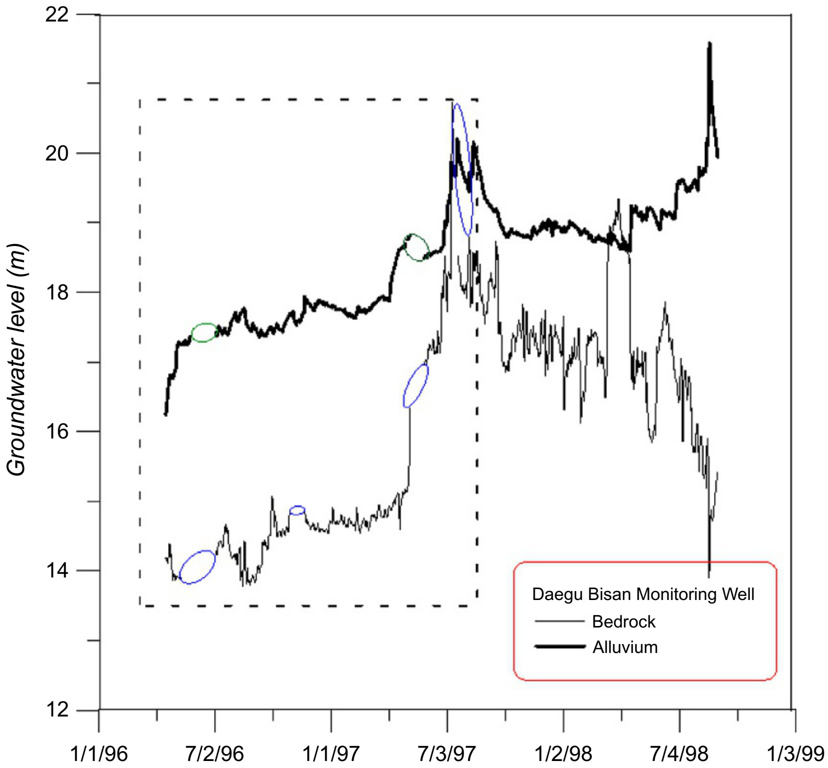

TABLE 4.1 General Statistics for Groundwater Levels of a Groundwater Monitoring Well

Statistics	Statistical Values	
	Alluvium	Bedrock
No. of data	383	409
Mean	17.77	15.43
Median	17.72	14.72
Standard deviation	0.48	1.55
Variance	0.23	2.40
Skewness	0.51	1.16
Kurtosis	1.02	−0.09
Minimum value	16.24	13.77
Maximum value	19.31	20.74

4.3 GEOSTATISTICAL METHODS

4.3.1 Kriging

Kriging is a local estimation technique of the best linear unbiased estimator (BLUE) for the unknown values of spatial and temporal variables. Kriging is expresses as:

$$Z_K^* = \sum_{i=1}^{n} \lambda_i Z_i \tag{4.1}$$

where Z_K^* is an estimate by kriging, λ_i is a weight for Z_i, and Z_i is a variable. The weight is determined to ensure that the estimator is unbiased and that the estimation variance is minimal (Journel and Huijbregts, 1978).

The unbiased condition of kriging is:

$$E\{Z_V - Z_K^*\} = 0 \tag{4.2}$$

where Z_V is an actual value and Z_K^* is an estimated value.

The sum of weights is:

$$\sum_{i=1}^{n} \lambda_i = 1.0 \tag{4.3}$$

The estimation variance of kriging variance is:

$$\sigma_K^2 = E\left\{[Z_V - Z_K^*]^2\right\} = \overline{C}(V,V) + \mu - \sum_{i=1}^{n} \lambda_i \overline{C}(v_i, V) \tag{4.4}$$

where $\overline{C}(V,V)$ represents the covariances between sample variables, μ is Langrange parameter, and $\overline{C}(v_i, V)$ represents the covariances between the sample variable and the estimates.

Various kinds of kriging have been developed to be suitable for the characteristics of used data, i.e., ordinary kriging for stationary data, universal kriging for nonstationary data, cokriging for a group of correlated data, etc. In this study, ordinary kriging was used to produce the graphs of groundwater-level data.

4.3.2 Variogram

The spatial dependence between sample data is necessary for the determination of kriging weights. The measure of the spatial dependence is the semivariogram expressed as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{4.5}$$

where $Z(x_i)$ and $Z(x_i+h)$ are observed variables at sampling point x_i and x_i+h, and $N(h)$ is the number of pairs of samples separated by the lag h. An experimental semivariogram needs to be fitted to a theoretical semivariogram model for the kriging interpolation.

The covariances between sample data are obtained from the following relation:

$$C(h) = sill - \gamma(h) \tag{4.6}$$

where $C(h)$ is covariance, and $\gamma(h)$ is semivariogram.

4.3.3 Conditional Simulation

The principles of conditional simulation are expressed as:

$$Z^*_{SC}(x) = Z^*_{OK}(x) + \left[Z_S(x) - Z^*_{SK}(x)\right] \tag{4.7}$$

where $Z^*_{SC}(x)$ is a conditional simulation and $Z^*_{OK}(x)$ is a kriged value at a point x.

$Z_S(x)$ is a nonconditional realization at a point x and $Z^*_{SK}(x)$ is a kriged value of a nonconditional realization $Z_S(x)$.

Turning band method (TBM; Journel, 1974) was used for a nonconditional realization $z_s(x)$ at a point x. TBM turns multidimensional simulations into several independent one-dimensional simulations for reasonable computer costs. $z_s(x)$ is a realization of three-dimensional random function $Z_S(x) = Z_s(u, v, w)$ which is a second-order stationary, and has a zero expectation and a covariance of $C(h) = E\{Z_s(x)Z_s(x+h)\}$.

The equation of TBM is expressed as:

$$z_s(x) = \frac{1}{\sqrt{(N)}} \sum_{i=1}^{N} z_i(x) \tag{4.8}$$

where $z_s(x)$ is multidimentional nonconditional realization at a point x, $z_i(x)$ is one-dimensional nonconditional realization, and N is the number of turning band lines ($N = 15$ in three-dimensional realization). TBM was developed by Matheron (1973) and practically applied by Journel (1974).

4.4 COMPARISON OF INTERPOLATION CAPABILITY

4.4.1 Statistical Validation Test

Some statistical errors were used for the accuracy validation between kriging and conditional simulation.

Mean error (ME):

$$ME = \frac{1}{N}\sum_{i=1}^{n} [Z(x) - Z^*]_i \tag{4.9}$$

Standard deviation of error (SDE):

$$VE = \frac{1}{N-1}\sum_{i=1}^{N} (Error - ME)^2 \tag{4.10}$$

$$SDE = \sqrt{VE} \tag{4.11}$$

Square root-mean-squared errors (SRMSE):

$$MSE = \frac{1}{N}\sum_{i-1}^{m} [Z(x) - Z^*(x)]_i^2 \tag{4.12}$$

$$SRMSE = \sqrt{MSE} \tag{4.13}$$

4.4.2 Interpolation of Artificial Missing Data

Alluvial groundwater-level data collected by the Daegu Bisan National Groundwater Monitoring Well were used for a comparison of the interpolation capability of kriging with that of conditional simulation. Experimental data were sampled from November 1, 1997 to April 15, 1998. Four kinds of missing intervals were arbitrary chosen from these data. Ten missing segments of data were identified from January 1, 1998 to January 10, 1998; 20 from January 1, 1998 to January 20; 40 from January 1, 1998 to February 19; and 60 from January 1, 1998 to March 10, respectively.

Fig. 4.2 shows the results of interpolation for the missing data (*black* color) using kriging and conditional simulation. The distributions of data interpolated by kriging (*red* color) are almost linear in shape, but those interpolated by conditional simulation (*blue* color) show similar fluctuations as the original distributions.

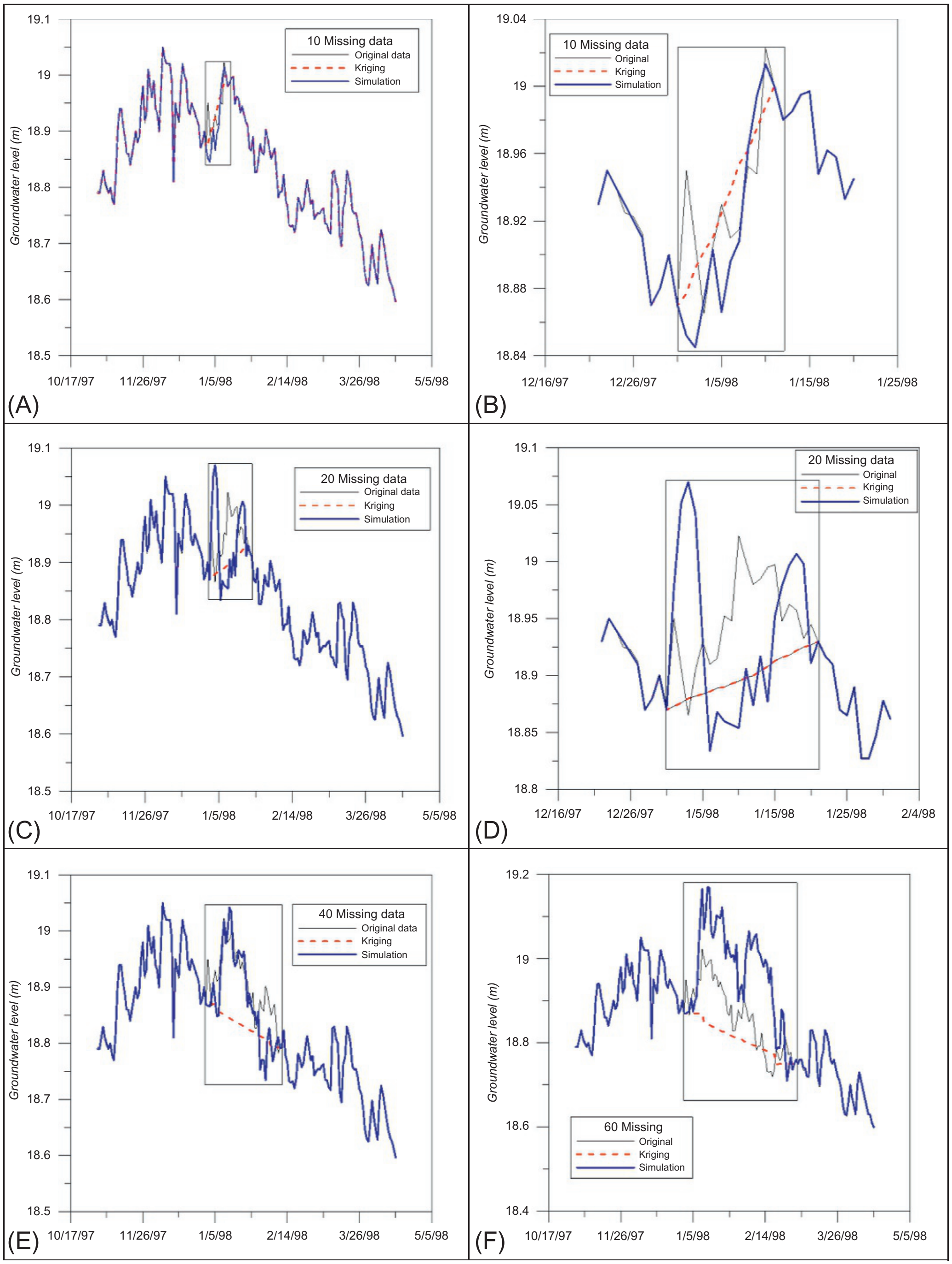

FIG. 4.2 Interpolations of (A) 10 missing data segments, (B) magnified graph of 10 missing data segments, (C) 20 missing data segments, (D) magnified graph of 20 missing data segments, (E) 40 missing data segments, and (F) 60 missing data segments.

Table 4.2 shows the statistical errors resulting from the interpolations of four kinds of missing data. The SRMSEs of kriging are smaller than conditional simulation for the time periods with 10, 20, and 60 missing data segments, but the SRMSE of conditional simulation is smaller than kriging for the time period with 40 missing data segments. Table 4.3 is the number of smaller deviations from original data. Kriging has more numbers than conditional simulation for the time period with 20, and 60 missing data segments, but conditional simulation has more numbers than kriging for the time period with 40 missing data segments. Thus conditional simulation is far superior to kriging for the interpolation of the 40 missing data segments.

TABLE 4.2 Statistics of Errors for Four Kinds of Missing Data

Statistics		10 Missing Data Segments	20 Missing Data Segments	40 Missing Data Segments	60 Missing Data Segments
Mean error (ME)	Kriging	0.002	−0.051	−0.0774	−0.052
	Conditional simulation	−0.019	−0.012	−0.0258	0.108
Standard deviation of error (SDE)	Kriging	0.0354	0.0355	0.0448	0.0538
	Conditional simulation	0.0429	0.0981	0.0503	0.0935
Square root of mean square error (SRMSE)	Kriging	0.0336	0.0620	0.0892	0.0748
	Conditional simulation	0.0450	0.0964	0.0560	0.1424

TABLE 4.3 Number of Smaller Deviations From Original Data

	Kriging		Conditional Simulation	
Data Segments	No. of Data Segments	Ratio (%)	No. of Data Segments	Ratio (%)
10 Missing	5	50	5	50
20 Missing	13	65	7	35
40 Missing	17	42.5	23	57.5
60 Missing	40	66.7	20	33.3

4.5 INTERPOLATION OF ACTUAL MISSING DATA

Kriging and conditional simulation were used for the interpolation of groundwater-level data actually missing from that collected by the Daegu Bisan National Groundwater Monitoring Well with peak-type variations.

4.5.1 Application of Kriging

Fig. 4.3 shows the results of interpolation by kriging of missing data for the groundwater-levels in alluvium and bedrock. Kriging reproduced the missing data with minimum errors. However, the distributions of reproduced data in alluvium and bedrock were nearly linear and didn't show the fluctuations of groundwater-levels.

4.5.2 Application of Conditional Simulation

Fig. 4.4 shows the results of interpolation by conditional simulation of missing data for groundwater-levels in alluvium and bedrock. Conditional simulation reproduced the missing data of alluvium and bedrock with reasonable fluctuations.

4.5.3 Cross Validation Test

Cross validation test was developed by Davis (1987) to examine the suitability of a specified vaiogram or covariance model for the given data.

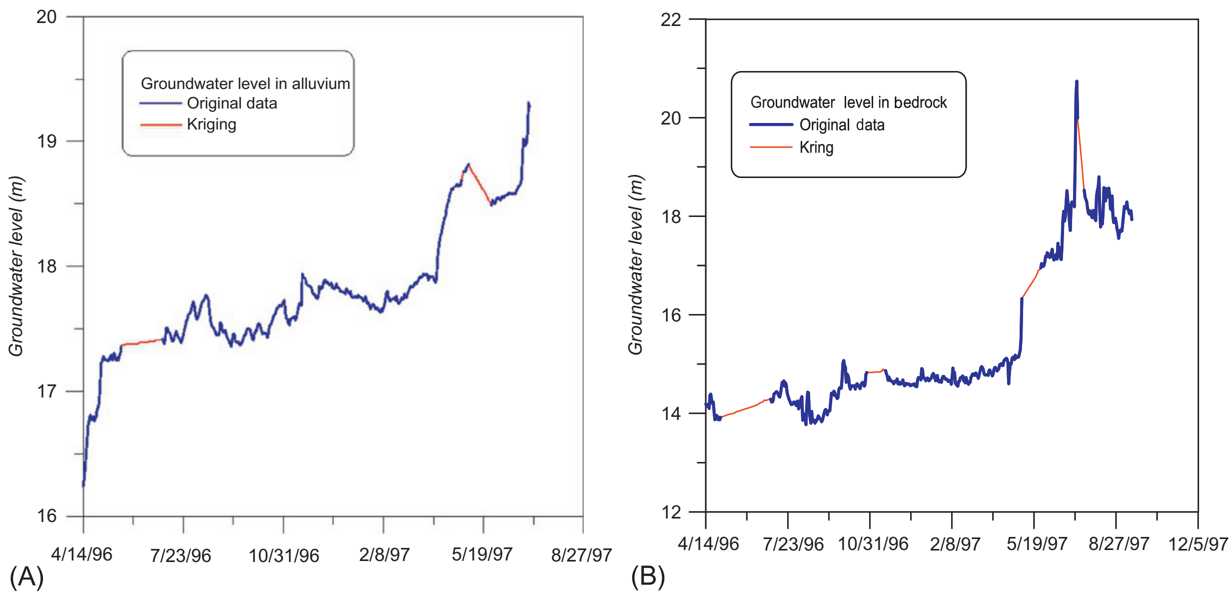

FIG. 4.3 Interpolation of missing data for the groundwater levels of (A) alluvium and (B) bedrock using kriging.

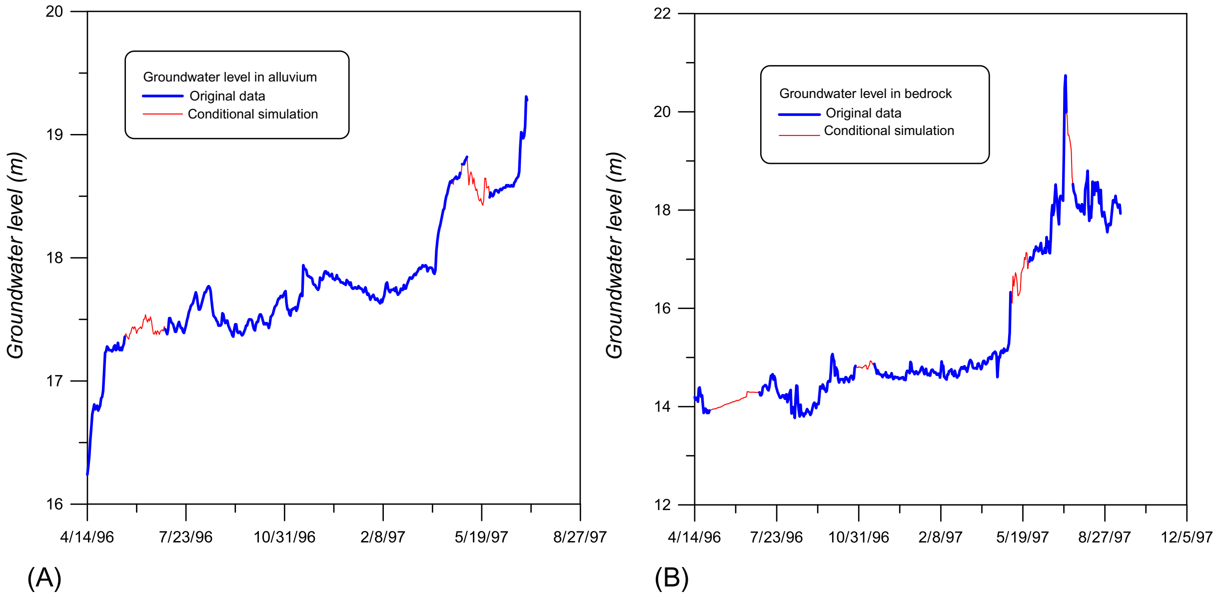

FIG. 4.4 Interpolation of missing data for groundwater levels of (A) alluvium and (B) bedrock using conditional simulation.

The test is using a reduced error (RE) which is defined by the error ($Z(x) - Z^*(x)$) divided by the square root of kriging variance $\left(\sqrt{\sigma_K^2}\right)$ of sample data:

$$RE = \frac{Z(x) - Z^*(x)}{\sqrt{\sigma_K^2}} \tag{4.14}$$

Kriging variance is calculated from kriging, but cannot be calculated from conditional simulation. Thus, mean error (ME), standard deviation of error (SDE) and Square Root of Mean Square Error (SRMSE) were used for the cross validation test between kriging and conditional simulation.

TABLE 4.4 Statistics of Errors for the Cross-Validation Test of Original Data

Statistics		Alluvium	Bedrock
Mean error (ME)	Kriging	0.00003	−0.00032
	Conditional Simulation	0.00028	−0.00064
Standard deviation of error (SDE)	Kriging	0.02199	0.10755
	Conditional Simulation	0.05171	0.23271
Square root of mean square error (SRMSE)	Kriging	0.02198	0.10742
	Conditional Simulation	0.05164	0.23245

Table 4.4 shows the statistical errors of cross-validation test. Statistical errors of kriging were smaller than conditional simulation, but the fluctuations of groundwater-levels couldn't be generated by kriging.

4.6 CONCLUSIONS

Alluvial groundwater-level data from the Daegu Bisan National Groundwater Monitoring Well were used for the comparison of the interpolation capabilities of kriging with those of conditional simulation. Experimental data were sampled from November 1, 1997 to April 15, 1998. Four kinds of missing intervals, i.e., 10, 20, 40, and 60 missing data segments, were arbitrarily chosen. The distributions of the missing data interpolated by kriging was almost linear in shape, although the statistical errors of kriging were smaller than those of conditional simulation. However, the data interpolated by conditional simulation showed the similar fluctuations to the original distributions, even though conditional simulation made larger statistical errors than kriging.

Groundwater-level data reproduced by kriging at the actual intervals of a groundwater monitoring well were completely consistent with the original data, but the missing data formed a nearly linear shape because kriging produced estimates with minimum errors. On the other hand, conditional simulation produced very similar fluctuations to the original distributions, although its estimates contained larger statistical errors than the kriging estimates. Therefore conditional simulation is widely used for the reproduction of irregular fluctuations such as groundwater level, hydraulic conductivity, and tidal fluctuation.

REFERENCES

Capilla, J.E., Gómez-Hernández, J.J., Sahuquillo, A., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 2. Demonstration on a synthetic aquifer. J. Hydrol. 203, 175–188.

Chiles, J.-P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons, Inc. 695 p.

Chung, S.Y., Wheatcraft, S.W., 1993. Application of fractal interpolation functions for modeling hydraulic conductivity distributions. J. Geol. Soc. Korea 29, 168–175.

Chung, S.Y., Shim, B.W., Kang, D.H., Kim, B.W., Park, H.Y., Won, J.H., Kim, G.B., 2001. Interpolation of missing groundwater level data using kriging at a national groundwater monitoring well. J. Geol. Soc. Korea 37, 421–430.

Davis, B.M., 1987. Uses and abuses of cross-validation in geostatistics. Math. Geol. 19, 241–248.

Gómez-Hernández, J.J., Sahuquillo, A., Capilla, J.E., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, part 1, theory. J. Hydrol. 204, 162–174.

Hendricks Franssen, H.J., Gómez-Hernández, J.J., Sahuquillo, A., 2003. Coupled inverse modeling of groundwater flow and mass transport and the worth of concentration data. J. Hydrol. 281, 281–295.

Hu, K., Huang, Y., Li, H., Li, B., Chen, D., White, R.E., 2005. Spatial variability of shallow groundwater level, electrical conductivity and nitrate concentration, and risk assessment of nitrate contamination in North China Plain. Environ. Int. 31, 896–903.

Jensen, J.L., Corbett, P.W.M., Pickup, G.E., Ringrose, P.S., 1996. Permeability semivariograms, geological structure, and flow performance. Math. Geol. 28 (4), 419–436.

Journel, A.G., 1974. Geostatistics for conditional simulation of ore bodies. Econ. Geol. 69, 673–687.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press. 600 p.

Liu, C.W., Jang, C.S., Liao, C.M., 2004. Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). Sci. Total Environ. 321, 173–188.

Loaiciga, H.A., Leipnik, R.B., Hudak, P.F., Marino, M.A., 1996. 1-, 2-, and 3-Dimensional effective conductivity of aquifers. Math. Geol. 28 (5), 563–584.

Matheron, G., 1973. The intrinsic random functions and their applications. Adv. Appl. Probab. 5.

Philip, R.D., Kitanidis, P.A., 1989. Geostatistical estimation of hydraulic gradients. Ground Water 27 (6), 855–865.

Theodossiou, N., Latinopoulos, P., 2006. Evaluation and optimization of groundwater observation networks using the kriging methodology. Environ. Model. Softw. 21, 991–1000.