

지하수위 분포 모델링을 위한 UNIVERSAL KRIGING의 응용
2. 제한적 최대 偶度法
Application of universal kriging for modeling a groundwater
level distribution
2. Restricted maximum likelihood method

정상용(Chung, Sang Yong)

부산수산대학교 응용지질학과

요약/Abstract

非定常 현상을 나타내는 지하수위 자료의 추정을 위해서 제한적 최대 偶度值(RML)을 이용하는 universal kriging을 사용하였다. 제한적 최대우도법의 우수성을 판정하기 위해 가중 최소자승법(weighted least squares method)을 이용하는 IRF-k와 비교 검토하였다. 교차 타당성에 있어서 RML과 k계 고유 확률함수(IRF-k)는 거의 비슷한 결과를 갖고, 참값 대 추정치의 離散分布圖와 지하수위 등고선도에 있어서도 2가지 방법이 거의 같은 결과를 만들었다. RML을 이용한 universal kriging과 가중 최소자승법을 이용한 IRF-k가 비정상 함수자료의 추정에 있어서 거의 비슷한 결과를 만든 원인은 이용된 실험자료의 非正規分布(non-Gaussian distribution)와 적은 자료수에 기인하는 것으로 사료된다.

Restricted maximum likelihood(RML) method was used to determine the parameters of generalized covariance, and universal kriging with RML was applied to estimate a groundwater level distribution of nonstationary random function. Universal kriging with RML was compared to IRF-k with weighted least squares method for the comparison of their accuracies. Cross validation shows that two methods have nearly the same ability for the estimation of groundwater levels. Scattergram of estimates versus true values and contour maps of groundwater levels have nearly the same results. The reason why two methods produced the same results is thought to be the non-Gaussian distribution and the small number of sample data.

서 언

수리지질학에 있어서 水理水頭, 오염농도, 그리고 대수층 상수(투수량계수, 수리전도도, 저류계수 등)의 공간적 변화에 대한 연구가 지하수 또는 오염물질의 유동상태 및 분포 현황, 지하수 산출 그리고 오염물질의 처리 등에 아주 중요하다. 수리수두나 오염농도 등은 수치해석이나 지구통계학적 방법에 의해 결정되어질 수 있으며, 대수층 상수는 역산 문제(Inverse problem : Kitanidis와 Vomvoris, 1983 ; Koeksma와 Kitanidis, 1984, 1985a, b ; Weber, 1991)나 지구통계학적 방법(Chung, 1992)에 의해 推定되어질 수 있다.

Matheron(1969, 1970)에 의해서 개발된 kriging이 수문학적 자료는 물론 지질학적 자료에 대한 공간적 해석이나 추정을 위해 널리 이용되고 있다. Drift를 갖는 비정상 확률함수(nonstationary random function)의 자료는 ordinary kriging에 의한 추정보다는 universal kriging을 이용하는 것이 좋은 추정치를 얻는다(Jornel과 Huijbregts, 1978). 정(1993)은 k계 고유 확률함수(Intrinsic random function of order k, IRF-k)가 지하수위분포의 모델링을 위해 ordinary kriging보다 좋은 결과를 준다는 것을 보여 주었다. 그러나, IRF-k는 일반적 공분산(generalized covariame, GC)의 상수 결정에 가중최소자승법(weighted least squares method)을 이용하는데, 이는 약간의 결함을 갖는다. Puente와 Bras(1986)는 GC 결정을 위한 Delfiner(1976)의 방식은 큰 표본 자료에는 좋지만, 작은 표본자료에는 최소평균자승오차 및 등급화방식(the least mean square error and ranking rocedures)이 더 좋은 결과를 준다고 했다. Starks and Fang(1982)도

가중최소자승법이 종속변수의 지나친 편중성과 독립변수의 복합공선형(multicollinearity) 때문에 비정상 자료의 추정에 불안정하다고 했다.

일반적 공분산(GC)의 결정에 더 우수한 방법인 최대우도법(maximum likelihood method, ML)이 Kitanidis(1983)에 의해서 GC의 결정을 위해 적용되었고, Kitanidis와 Vomvoris(1983), Hoeksma와 Kitanidis(1984, 1985a) 그리고 Kitanidis와 Lane(1985)에 의해서 수리지질학에 이용되었다. 이 연구의 목적은 일반적 공분산의 결정을 위하여 최대우도법의 일종인 제한적 최대우도법(restricted maximum likelihood method, RML)을 미국 네바다주에서 획득한 지하수위 표본자료에 이용하여, 가중 최소자승법과 비교하고 연구 지역의 정확한 지하수위 분포를 파악하는 데 있다. 교차타당성(cross validation)과 추정치 및 추정분산(estimation variance)의 등고선도(contour map)들이 2가지 방법의 비교를 위해서 이용된다.

최대 偶度法

이 방법은 표본자료의 Gaussian분포를 가정한다. N개의 자료에 대한 Gaussian 공동확률밀도함수(Gaussian joint probability density function)는 다음과 같다.

$$P(Z/\theta) = (2\pi)^{-\frac{N}{2}} |Q|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (Z-\mu)^T Q^{-1} (Z-\mu) \right]$$

여기서 $\mu = E[Z/\theta]$

$$Q = E[(Z-\mu)((Z-\mu)^T/\theta)]$$

- |Q| = Q의 행렬값
- Z = N개 자료의 벡터
- θ = 일반적 공분산(GC) 상수의 벡터
- T = 전치행렬(transposed matrix)을 의미

GC의 상수벡터는 최대우도(ML) 추정이나 음대수 우도 함수(negative log-likelihood function)에 의해서 결정된다.

$$L(Z|\theta) = -\ln p(Z|\theta) = \frac{N}{2} \ln(2\pi)$$

$$+ \frac{1}{2} \ln |Q| + \frac{1}{2} (Z-\mu)^T Q^{-1} (Z-\mu)$$

음대수 우도 함수의 GC스칼라 상수 θ에 대한 미분형은 다음과 같다.

$$\frac{\partial L}{\partial \theta_i} = \frac{1}{2} \text{Tr}(Q^{-1} \frac{\partial Q}{\partial \theta_i}) - \frac{1}{2} (Z-\mu)^T Q^{-1}$$

$$\frac{\partial Q}{\partial \theta_i} Q^{-1} (Z-\mu) - (Z-\mu)^T Q^{-1} \frac{\partial \mu}{\partial \theta_i}$$

여기서

T = 전치행렬을 의미

Tr = 행렬의 trace를 의미

$$\frac{\partial}{\partial \theta_i} \ln |Q| = \text{Tr}[Q^{-1} \frac{\partial Q}{\partial \theta_i}]$$

$$\frac{\partial}{\partial \theta_i} Q^{-1} = -Q^{-1} \frac{\partial Q}{\partial \theta_i} Q^{-1}$$

음대수 우도함수의 미분값은 최소값 θ*에 대해서 0이 되어야 한다.

$$g_i = \left. \frac{\partial L}{\partial \theta_i} \right|_{\theta=\theta^*} = 0$$

GC의 상수결정을 위해서 음대수 우도 함수의 최소화를 위한 gradient($\frac{\partial L}{\partial \theta}$)에 근거한 반복법이 이용된다. 기본적인 반복은 아래와 같다.

$$\theta_{i+1} = \theta_i + \rho_i R_i g_i$$

여기서

θ_i = i번째 반복에 있어서 GC상수의 벡터

ρ_i = 스칼라인 간격 크기 (보통 1.0)

g_i = i번째 반복에 있어서 gradient vector

그러나 최대 우도법은 표본자료가 클 경우 drift와 GC상수를 동시에 결정할 경우 좋지 않은 결과를 만들어낸다. Kitanidis와 Lane (1985)은 최대우도법의 단점을 보완하기 위해 제한적 최대 우도법(Restricted maximum likelihood method; RML)을 GC의 결정에 적용하였다. RML방법은 표본자료의 평균을 0으로 변환시킨 후 GC의 상수를 결정한다. 또한 이 방법은 음대수 우도 함수의 2계미분(Hessian) 대신에 Fisher정보행렬(Fisher information matrix, M)을 이용한다.

즉

$$M_i = E \left\{ \left(\frac{\partial^2 L}{\partial \theta^2} \right) \middle| \theta = \theta_i \right\}$$

$$= E \left\{ \left(\frac{\partial L}{\partial \theta} \right) \left(\frac{\partial L}{\partial \theta} \right)^T \middle| \theta = \theta_i \right\}$$

Fisher정보행렬의 (j, k)요소는 다음과 같다.

$$M(j, k) = E \left(\frac{\partial L}{\partial \theta_j} \frac{\partial L}{\partial \theta_k} \right)$$

실험자료의 RML방법에의 응용

미국 네바다주에서 얻은 38개의 지하수위 자료(그림1)를 이용하여 semivariogram(그림 2)을 구하고, RML방법을 적용하였다. 이 방법의 우수성을 판정하기 위해 k계 고유확률함수(IRF-k)를 이용하여 추정치를 구하였다(정, 1993). 교차타당성(cross validation), 이산분포도(scattergram), 등고선도(contour map)등이 두 방법을 비교하기 위해 이용되었다.

실험자료에서 결정된 drift는 1이고, RML 방법에 의해 결정된 GC의 상수는 $C=2610$, $\alpha_0=-54700$, $\alpha_1=0.127$, $\alpha_2=0$ 이고, IRF-1에 의해 얻어진 GC의 상수는 $C=15995$, $\alpha_0=-33782$, $\alpha_1=\alpha_2=0$ 이다.

교차 타당성

IRF-1과 RML을 이용한 universal kriging으로부터 추정치를 구하여 각각 참값에서 오차를 얻어 교차타당성에 이용하였다. 선정된 GC의 모델이 적합하다면 감소오차의 평균(Mean Reduced Error, MRE)이 0에 가깝고, 감소오차의 표준편차(Standard Deviation of Reduced Error, SDRE)가 1.0에 가까워야 한다(Divis, 1987; Solow, 1990).

또한 평균오차(ME), 평균자승오차(MSE), 평균자승오차의 제곱근(SRMSE), 평균 Kriging 분산(MKV), 오차의 분산(VE) 등의 통계값이 교차확인을 위해 이용되었다. IRF-1과 RML의 오차를 비교하기 위해서 구해진 통계값들이 표1에 있다.

IRF-1의 MRE가 RML의 그것보다 약간 더

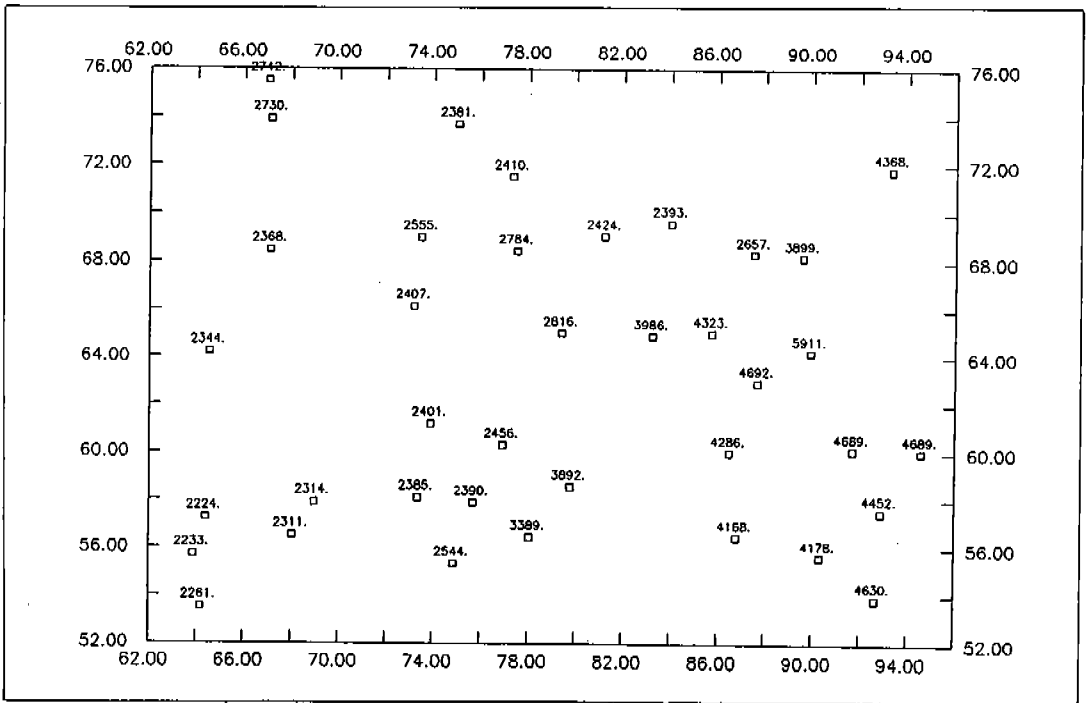


Fig. 1 Groundwater levels of study area in Nevada, U.S.A. The unit of groundwater levels is feet.

Table 1 Statistical errors resulting from IRF-k and RML method.

	ME	MSE	SRMSE	MKV	VE	MRE	SDRE
IRF-1	18.8	1.44×10^5	379	1.24×10^5	1.47×10^5	0.0141	1.14
RML	20.3	1.36×10^5	369	1.90×10^5	1.39×10^5	0.0156	0.895

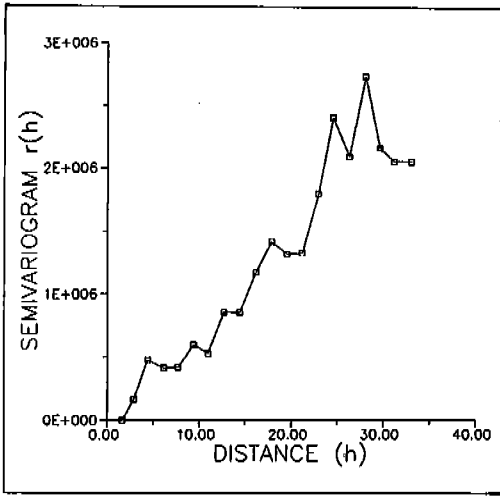


Fig. 2 Variogram of sample data.

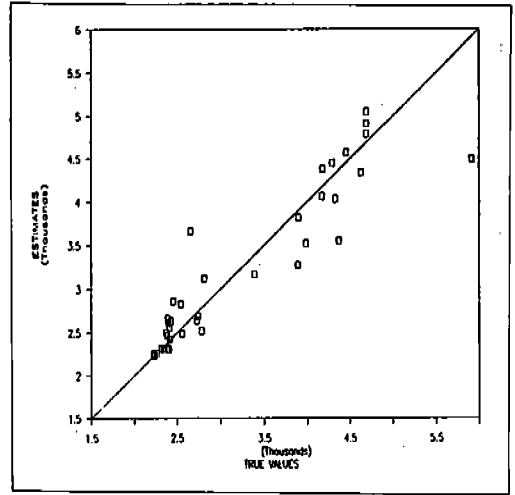


Fig. 3 Scattergram of estimates produced by IRF-k versus true values.

0에 가깝고, RML의 SDRE가 IRF-1의 그것보다 약간 더 1에 가까우나 그 차이들은 아주 적다. ME와 MKV에 있어서는 IRF-1이 RML보다 적은 값을 가지나, MSE, SRMSE, 및 VE에 있어서는 RML이 IRF-1보다 적은 값을 갖는다.

이산 분포도

참값에 대한 추정치의 분포를 나타내는 異散 분포도에서 추정치들이 1대 1비율의 대각선상에 위치하면 그것들은 참값과 같다는 것을 뜻한다. 추정치가 대각선 위에 놓이면

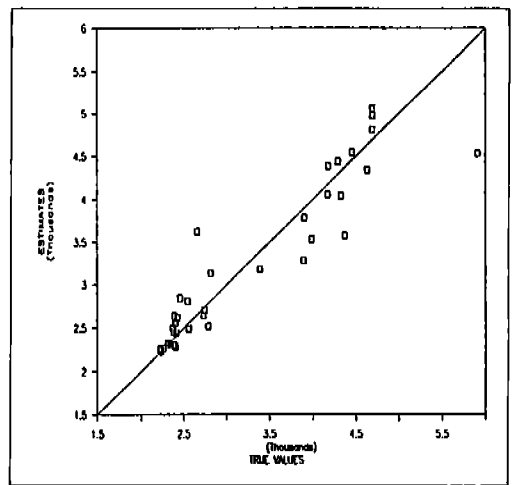


Fig. 4 Scattergram of estimates produced by RML method versus true values.

참값보다 높게 추정되었고, 대각선 밑에 놓이면 참값보다 낮게 추정되었다는 것을 의미한다. 따라서 이산분포도를 이용하면 선정된 모델의 우수성을 판정하는데 도움을 얻을 수 있다. 그림 3과 4는 각각 IRF-1과 RML에서 얻은 참값에 대한 추정치의 이산분포도들이다. 두 그림을 겹쳐보면 추정치들이 거의 같은 지점에 위치한다.

등고선도

그림 5는 IRF-1에서 얻은 지하수위 등고선도이고, 그림 6은 RML방법을 이용한 universal kriging으로 부터 얻은 지하수위 등고선도이다. 두 그림을 겹쳐보면 거의 일치할 하고 있다. 그림 7은 IRF-1에서 얻은 추정분산(esti-

mation variance)치의 등고선도이고 그림 8은 RML방법을 이용한 universal kriging으로 부터 얻은 추정분산치의 등고선도이다. IRF-1의 추정분산치가 RML의 추정분산치보다 적다는 것을 알 수 있다. 그림 9는 IRF-1에서 얻은 3차원 표면도이고, 그림 10은 RML방법을 이용한 universal kriging으로 부터 얻은 3차원 표면도이다. 두 그림을 겹쳐보면 거의 완전히 일치한다.

토의 및 결론

IRF-1과 RML을 이용한 universal kriging의 교차 타당성을 비교해 볼때, MRE값들이 거의 비슷하게 0에 가깝고, SDRE값이 거의 비슷

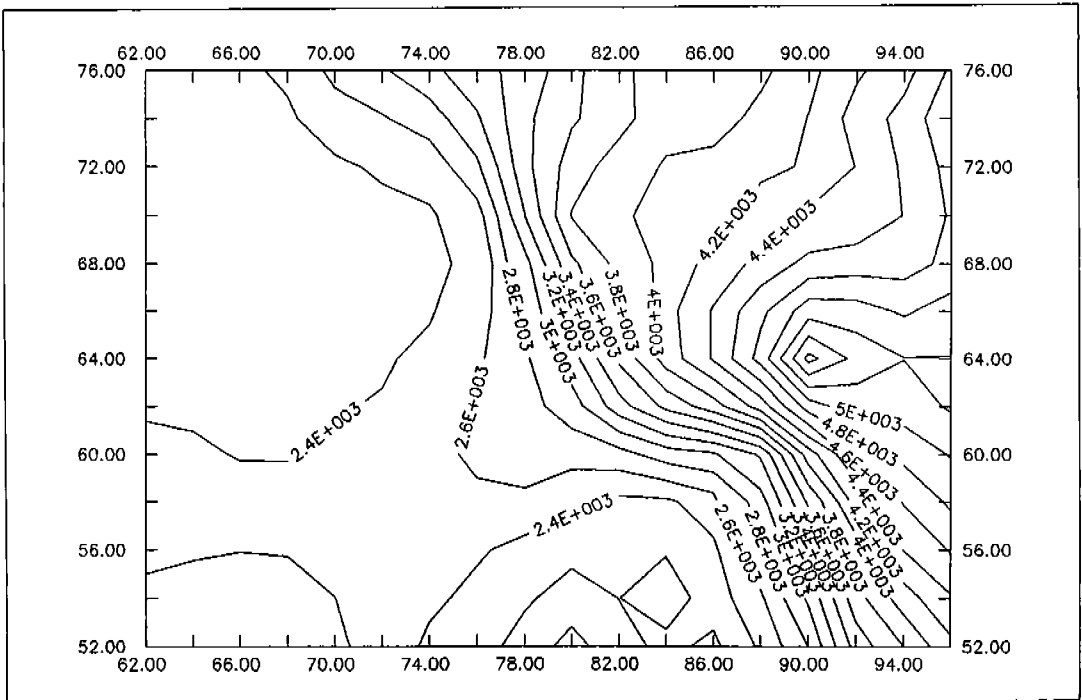


Fig. 5 Contour map of groundwater levels produced by IRF-k.

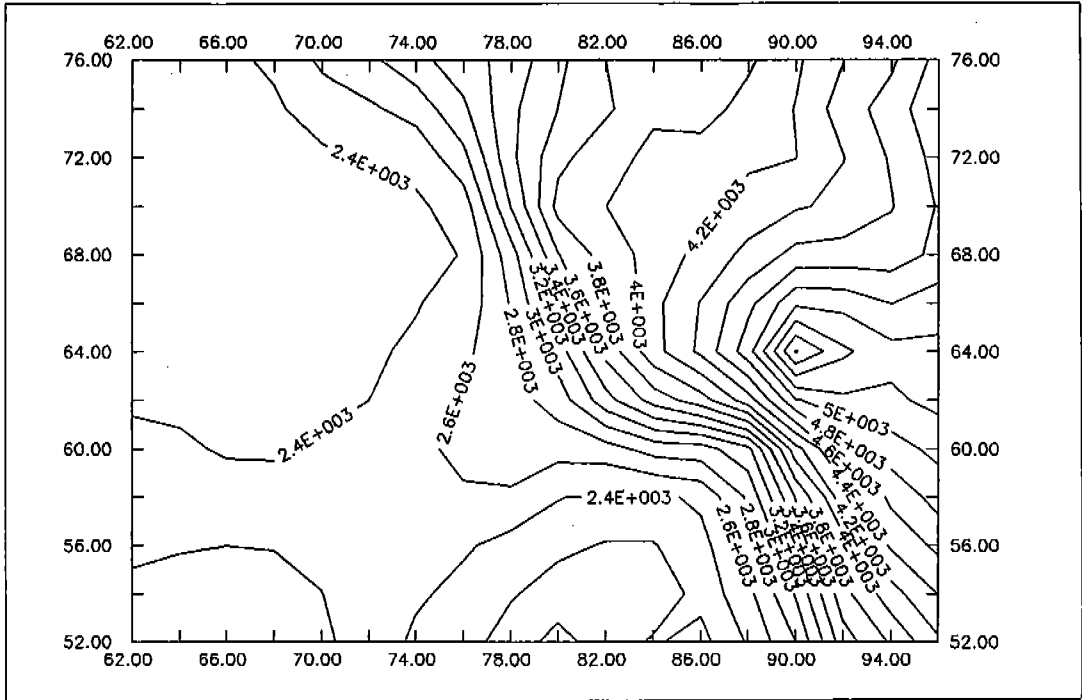


Fig.6 Contour map of groundwater levels produced by RML method.

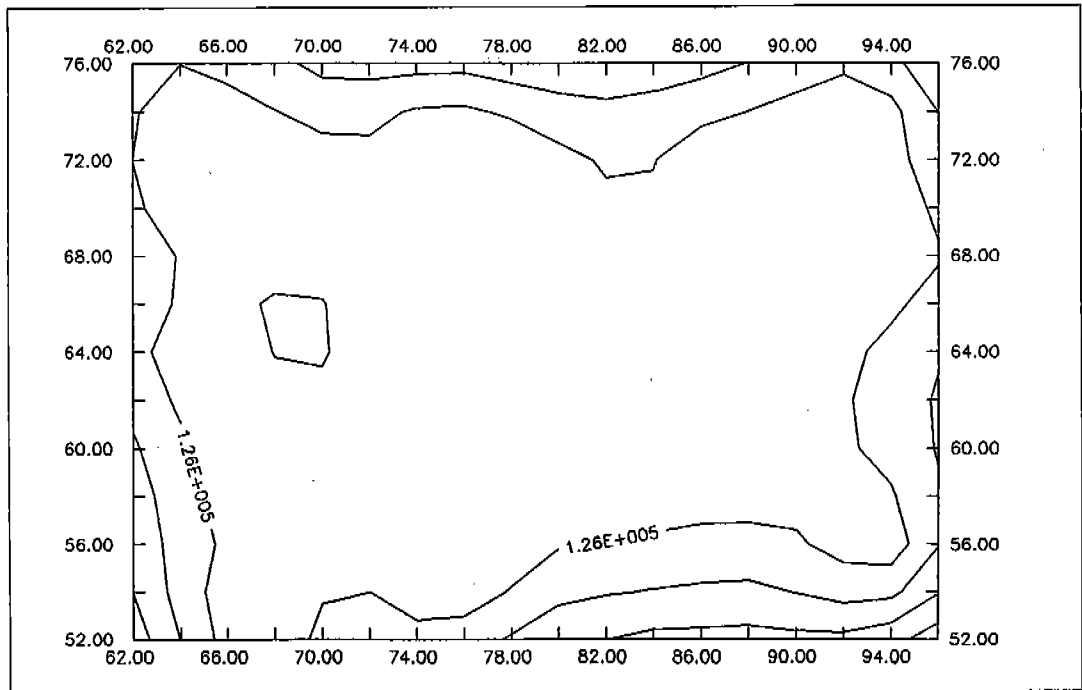


Fig.7 Contour map of estimation variance produced by IRF-k.

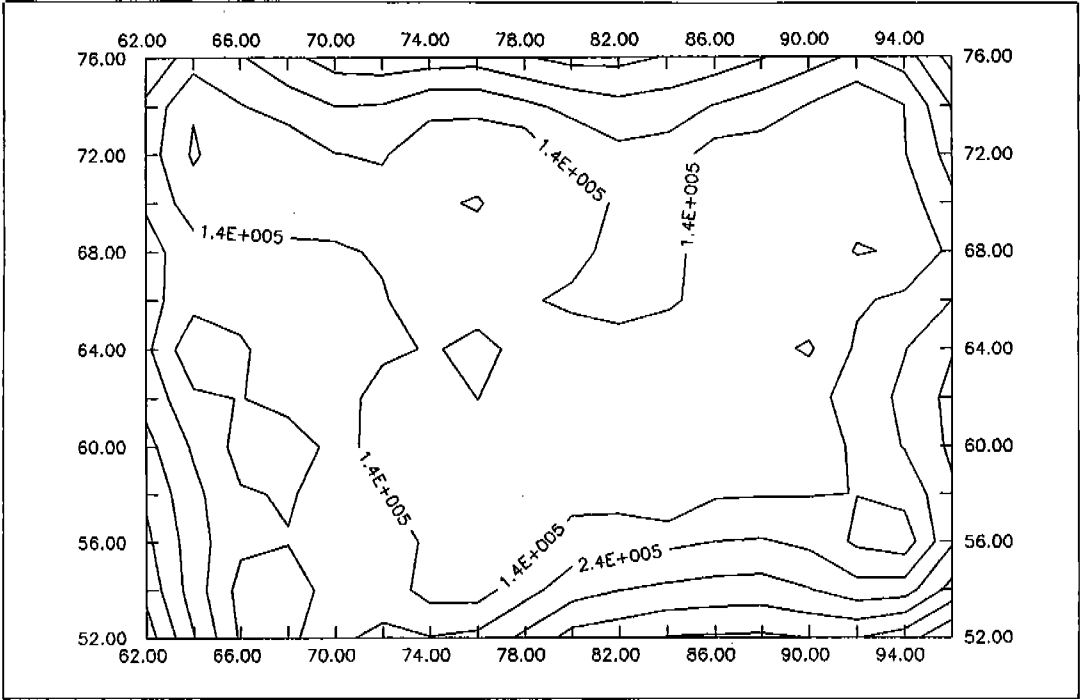


Fig. 8 Contour map of estimation variance produced by RML method.

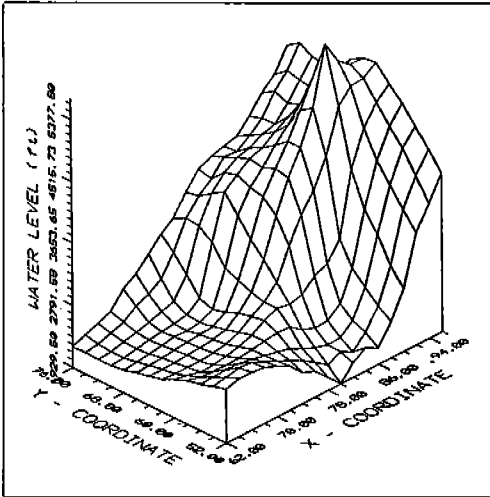


Fig. 9 Surface map of groundwater levels produced by IRF-k.

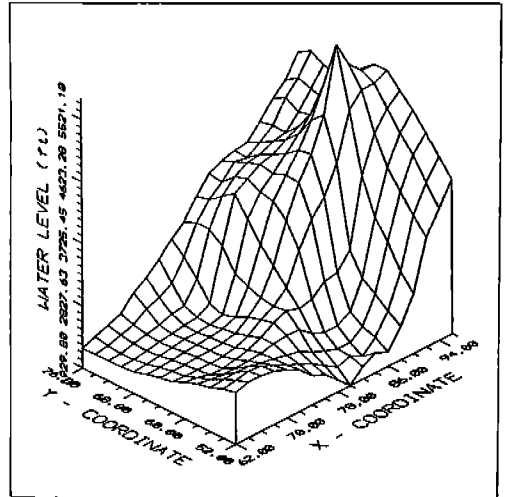


Fig. 10 Surface map of groundwater levels produced by RML method.

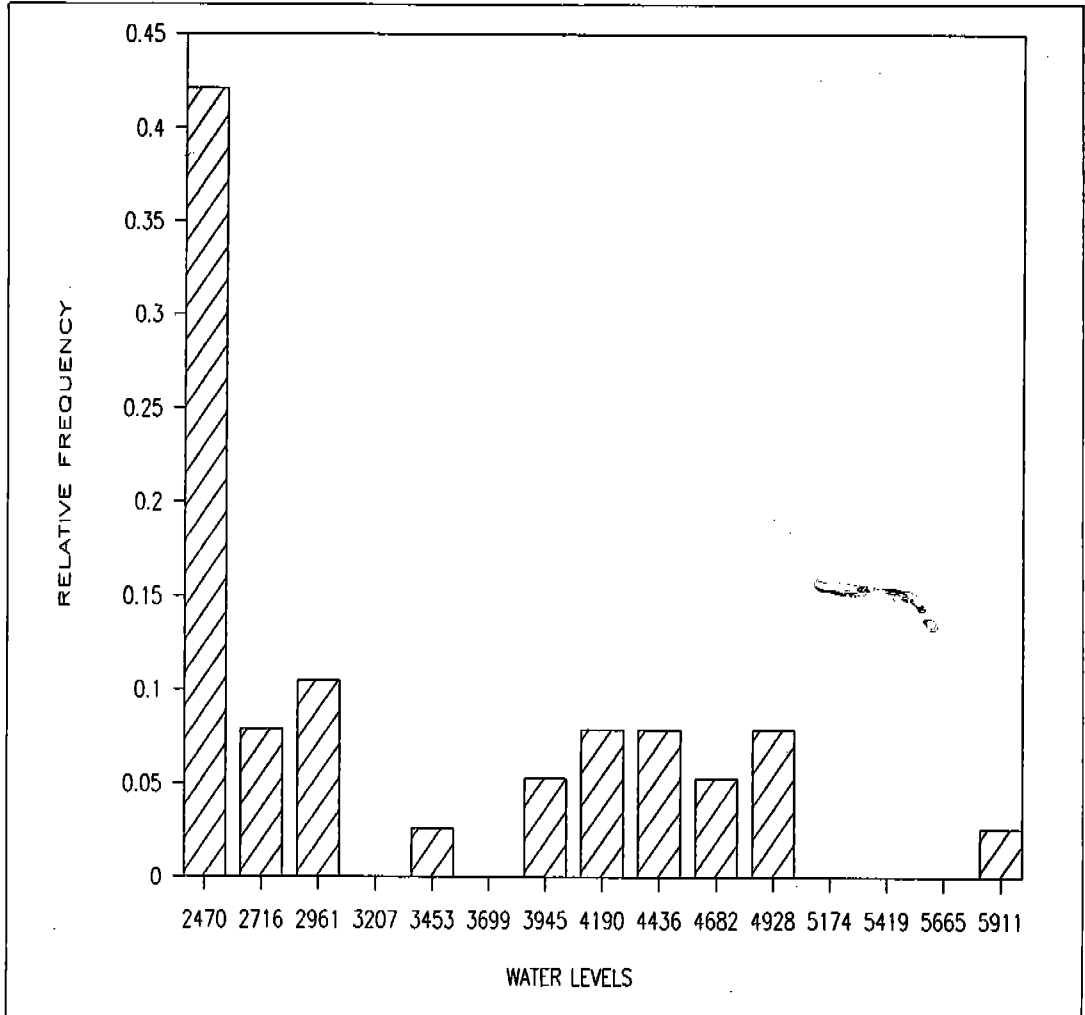


Fig. 11 Histogram of groundwater levels.

하게 1에 가깝다. 또한 ME, MSE, SRMSE 및 VE값도 RML들과 IRF-1이 큰 차이가 없다. 참값에 대한 추정치의 이산분포도에서도 IRF-1과 RML이 거의 비슷하다. 지하수위 등고선도와 3차원표면도를 비교해 보면 역시 IRF-1이 RML과 거의 같은 분포를 만들어냈다. 추정분산치에 있어서 IRF-1이 RML보다 적은 값을 만들어냈으나 전체적으로 볼때 두가지 방법은 지하수위 분포의 추정에서 거의 같은

능력을 보여 주었다.

RML방법은 표본자료의 정규분포(normal distribution)를 요구한다. 그러나 이용된 실험자료는 정규분포를 보이지 않는다(그림 11). 또한 RML방법은 실험자료의 수가 많을 경우에 좋은 결과를 준다. 그러나 이용된 실험자료의 수는 38개에 불과했다. 따라서 RML방법이 지하수위 추정에 있어서 IRF-k보다 크게 우수하지 않은 이유는 실험자료의

비정규분포와 자료의 수가 적은데 원인이 있는 것으로 사료된다.

사 사

본 연구를 위해 지하수위 자료를 제공해준 미국 네바다 주립 연구소 Desert Research Institute에 감사한다.

참고문헌

- 정상용, 1993, 지하수위분포 추정을 위한 Universal Kriging, 1, k계의 고유확률함수, 지질공학회지, 금번호.
- Chung, S.Y., 1992, Development of Inverse Fractional Interpolation Functions for Modeling Hydraulic Conductivity Distributions, Ph. D. Dissertation, University of Nevada-Reno, 207p.
- Davis, B.M., 1987, Uses and Abuses of Cross-Validation in Geostatistics, Math. Geol., 19(3), 241-248.
- Delfiner, P., 1976, Linear Estimation of Non Stationary Spatial Phenomena, Advanced Geostatistics in the Mining Industry, edited by M. Guarascio, M. David, and C. Huijbregts, 49-68pp, D. Reidel Publishing Company, Dordrecht-Holland.
- Hoeksema, R.F., and P.K. Kitanidis, 1984, An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling, Water Resour. Res., 20(7), 1003-1020.
- Hoeksema, R.F., and P.K. Kitanidis, 1985a, Analysis of the Spatial Structure of Properties of Selected Aquifers, Water Resour. Res., 21(4), 563-572.
- Hoeksema, R.F., and P.K. Kitanidis, 1985b, Comparison of Conditional Mean and Kriging Estimation in the Geostatistical Solution of the Inverse Problem, Water Resour. Res., 21(6), 825-836.
- Journel, A.G., and Ch. J. Huijbregts, 1978, Mining Geostatistics, 600p, Academic Press.
- Kitanidis, P.A., 1983, Statistical Estimation of Generalized Covariance Functions and Hydrologic Applications, Water Resour. Res., 19(4), 909-921.
- Kitanidis, P.K., and E.G. Vomvoris, 1983, A Geostatistical Approach to the Inverse Problem in Groundwater Modeling (Steady state) and One-Dimensional Simulations, Water Resour. Res., 19(3), 677-690.
- Kitanidis P.A., and R.W. Lane, 1985, Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes by the Gauss-Newton Method, Jour. of Hydrology, 79(1985), 53-71.
- Matheron, G., 1969, Le Krigeage Universel, Cah. Centre Morphol., Math., 1.
- Matheron, G., 1970, The Theory of Regionalized Variable and Its Applications, Cah. Centre Morphol., Math., 5.
- Puente, C.E., and R.L. Bras, 1986, Disjunctive Kriging, Universal Kriging, or No Kriging: Small Sample Results with Simulated Fields, Math. Geol., 18(3), 287-305.
- Solow, A.R., 1990, Geostatistical Cross-Validation: A Cautionary Note, Math. Geol., 22(6), 637-639.

- Starks, T.H. and J.H.Fang, 1982, On the Estimation of the Generalized Covariance Functions, *Math. Geol.*, 14(1), 57-64.
- Weber, D., D.Easley, and England, 1991, Probability of Plume Interception Using Conditional simulation of Hydraulic Head and Inverse Modeling. *Math. Geol.*, 23(2), 219-239.

정상용 :

부산수산대학교
해양과학대학 응용지질학과
부산시 남구 대연3동 599-1
608-023
TEL : (051) 620-6235
FAX : (051) 627-1096