

지하수위 분포 모델링을 위한 UNIVERSAL KRIGING의 응용

1. K 階의 고유 확률함수

Application of universal kriging for modeling a groundwater level distribution

1. Intrinsic random function of order k

정상용 (Chung, Sang Yong)

부산수산대학교 응용지질학과

요약/Abstract

非正常 확률함수인 지하수위 자료의 추정을 위하여 k계 의 고유확률함수(IRF-k)를 이용하였다. IRF-k의 우수성을 판정하기 위해 지하수위 자료가 정상 확률함수라고 가정하여 ordinary kriging과 비교 검토하였다. 교차 타당성(cross validation)에 의하면 IRF-k가 ordinary kriging에 비해서 지하수 분포 추정에 더 적합한 모델이다. 또한 추정치의 통계적 오차에 의하여도 IRF-k가 ordinary kriging에 비해 우수하다. IRF-k와 ordinary kriging은 지하수위 등고선과 3차원 표면 그림에 있어서 차이를 보이는데, IRF-k에 의해 만들어진 도면들이 더 정밀하다.

Intrinsic random function of order k(IRF-k) was used to estimate groundwater levels of nonstationary random functions. The accuracy of IRF-k was compared to that of ordinary kriging assuming that the data of groundwater levels compose a stationary random function. Cross validation and statistical errors show that IRF-k is superior to ordinary kriging for the estimation of water levels. IRF-k and ordinary kriging made different contour and 3-D surface maps. The maps of IRF-k are more accurate than those of ordinary kriging.

서 언

유선망(flow net)은 지하수의 유동방향, 지하수 유동속도 및 지하수 유량의 결정에 중요한 역할을 한다. 따라서 유선망의 등압선(equipotential line)을 이루는 지하수위의 분포의 파악은 지하수 유동상태의 이해에 매우 중요하다. 지하수위 분포의 파악을 위한 등고선도(contour map)의 작성은 격자(grid)상의 표본자료를 필요로 하는데, 실제로 격자상의 자료 획득은 시간적, 경제적 문제로 어렵다. 그러므로 거의 대부분 확률 공간 분포 자료(random-spaced data)를 이용하는 것이 통례인데, 등고선도의 작성을 위해 격자상의 지점에서 표본자료를 이용한 자료의 推定이 필연적으로 요구된다.

지질학적 자료(예 : 광상의 품위, 지진의 강도, 지층의 두께, 지형의 고도 등)의 추정에는 Krige에 의해 개발되고 Matheron(1969, 1970)에 의해 이론적으로 확립된 Kriging이 좋은 방법이다. 특히 수문 및 수리지질학에 있어서 Kriging이 중요한 역할을 한다(Neuman, 1984). 현재 Kriging은 대수층 상수의 추정(Yeh et. al., 1983 : Aboufirassi와 Marino, 1984 : Kiitanidis와 Vomvoris, 1983)이나 지하수위 또는 수리수두경사(hydraulic head gradient)의 추정에 활발히 이용되고 있다(Aboufirassi와 Marino, 1983 : Philp과 Kitanidis, 1989 : Neuman과 Jacobson, 1984). 그러나 공간상에 분포하는 표본자료의 평균이 상수가 아닌 위치의 함수로 표현될 때 보통의 Kriging으로는 未知의 지점에서 정밀한 추정치를 얻

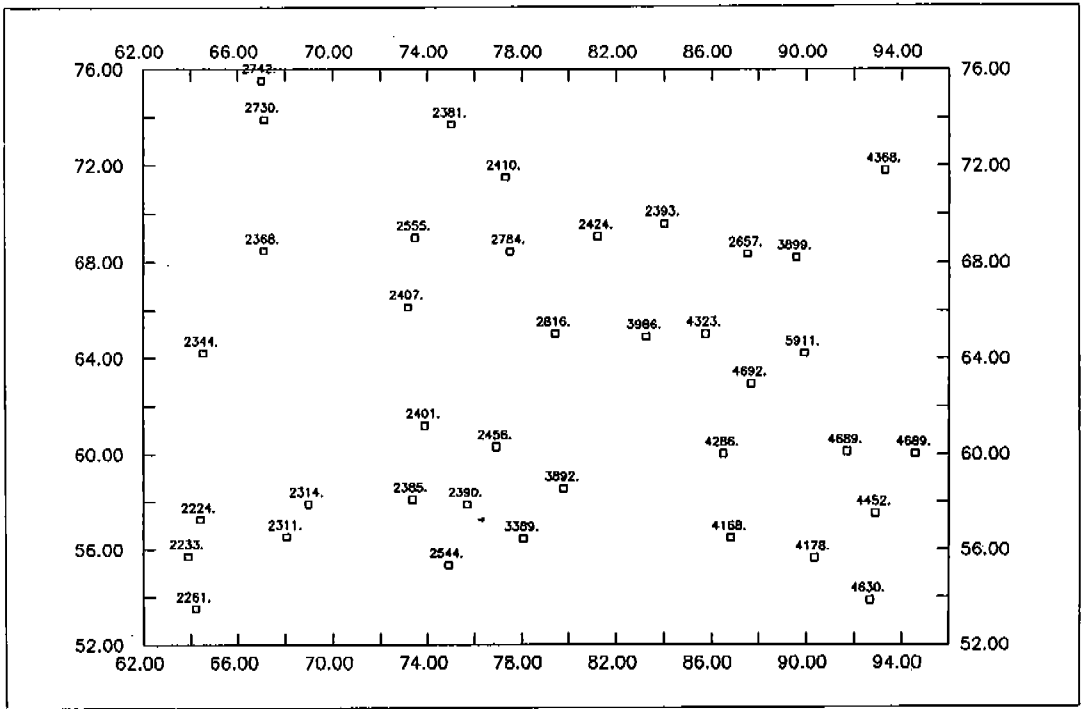


Fig. 1 Groundwater levels of study area in Nevada, USA. The unit of groundwater levels is feet.

기가 어렵다. 표본자료의 평균(이를 drift 또는 trend라고 한다)이 상수가 아닌 경우에는 비정상 확률함수(nonstationary random function)를 이루는데, k계의 고유 확률함수(intrinsic random function of order k: IRF-k)는 이런 자료의 정확한 추정을 위해 Matheron (1973)에 의해 개발되고 Delfiner(1976)에 의해 완성되었다.

이 연구의 목적은 일반적으로 비정상적인 현상을 보여주는 소수의 지하수위 자료를 가지고 지구통계학적 방법을 이용하여 필요한 지점에서 지하수위를 추정하여 실험 지역의 지하수 유동상태를 파악하는데 있다. 이용되는 지구통계학적 방법은 ordinary kriging과 IRF-k인데, 2가지 방법의 정밀성이 통계학적 측면과 등고선도(contour maps)들에 의해서 비교되어 진다.

그림 1은 미국 네바다주에서 얻어진 38개의 지하수위 자료로서, 그 평균은 3210 ft, 분산(variance)은 $1.03 \times 10^6 \text{ ft}^2$, 그리고 표준편차는 $1.02 \times 10^3 \text{ ft}$ 이다.

Variogram의 계산

공간상에 분포하는 표본자료의 변화의 특성은 variogram에 의해서 나타내어진다. variogram은 공간상에 분포하는 임의의 두 자료간의 차이의 분산(variance)이다.

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} \quad (1)$$

여기서 $\gamma(h)$ 는 semivariogram이고, $2\gamma(h)$ 는 variogram이다. $Z(x)$ 는 임의의 지점 x 에 있는 자료이며 $Z(x+h)$ 는 $Z(x)$ 에서 h 만큼 떨어진 지점의 자료이다.

표본자료에서 얻어진 semivariogram의 분

포는 그림 2에 나타나 있다. 실험자료는 비정상 현상을 보여준다. 정상(stationary) 자료라면 semivariogram이 그 자료의 분산(variance)근방에서 일정한 값을 유지해야만 한다. 그림 2에서 실험자료가 정상인 경우라고 가정하여서 球상의(spherical)모델이 실험자료의 variogram에 적용되었다. 이때 nugget은 $1.0 \times 10^5 \text{ ft}^2$, sill은 $1.03 \times 10^6 \text{ ft}^2$ 이며 semivariogram range는 12.4ft이다.

구상의 semivariogram은 다음의 식으로 결정된다.

$$\begin{cases} \gamma(h) = CO + C(1.5\frac{h}{r} - 0.5\frac{h^3}{r^3}) & (h < r) \\ \gamma(h) = \text{sill} & (h \geq r) \end{cases} \quad (2)$$

h 는 실험자료간의 거리이며, CO 는 nugget 이고 $C = \text{sill} - CO$, r 은 구상의 variogram이 안정될때 까지의 거리이다.

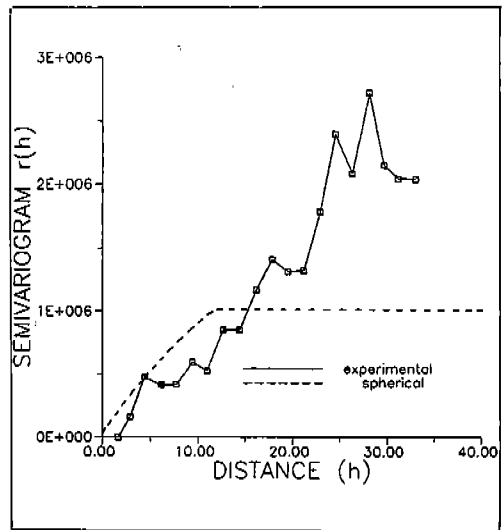


Fig. 2 Variogram of sample data. Solid and dashed lines represent experimental and theoretical variograms, respectively.

Ordinary Kriging

Kriging은 미지의 특성에 최선의 선형 불편 추정자(best linear unbiased estimator, BLUE)를 제공하는 통계적 방법이다. 그것의 수학적 표현은

$$Z_k^* = \sum_{i=1}^n \lambda_i Z_i \quad (3)$$

여기서 Z_k^* 는 추정치이고, Z_i 는 추정치 주변에 존재하는 참값들이며, λ_i 는 가중치이다. 이 가중치는 kriging이 불편(unbiased)이고, 추정분산(estimation variance)이 최소화 되도록 결정되어야 한다. 즉,

$$E\{Z_v - Z_k^*\} = 0 \quad (4)$$

$$\begin{aligned} E\{Z_v - Z_k^*\} &= \\ E\{Z_v\} - 2E\{Z_v Z_k^*\} + E\{Z_k^{*2}\} & \\ = \text{minimal} \end{aligned} \quad (5)$$

식(4)로부터 가중치의 총합계는 1.0이 된다. 식(5)로부터 추정분산은 자료들간의 공분산(covariance)을 이용해서 다음과 같이 표현된다.

$$\begin{aligned} E\{Z_v - Z_k^*\} &= \bar{C}(V, V) - 2\sum_i \lambda_i \bar{C}(V, v_i) \\ + \sum_i \sum_j \lambda_i \lambda_j \bar{C}(v_i, v_j) & \\ = \text{minimal} \end{aligned} \quad (6)$$

여기서 $\bar{C}(V, v_i)$ 는 공분산함수 $C(h)$ 의 평균값인데, $C(h)$ 는 영역 V 와 v_i 에서의 거리 h 만큼 떨어진 두 표본자료간의 공분산이다. 추정분산은 kriging 분산(σ_k^2)이라고도 하는데 kriging system을 이용하여 다음과 같이 표현될 수 있다(Journal and Huijbregts, 1978).

$$\begin{aligned} \sigma_k^2 &= E\{(Z_v - Z_k^*)^2\} \\ &= \bar{C}(V, V) + \mu - \sum_{i=1}^n \lambda_i \bar{C}(v_i, V) \end{aligned} \quad (7)$$

여기서 μ 는 Lagrangian 상수이다.

k階의 고유 확률 함수

n 개의 실험자료의 선형결합인 $\sum_{i=1}^n \lambda_i Z_i$ 는 다음의 조건에서 일반적 증가(generalized increment)가 된다.

$$\sum_{i=1}^n \lambda_i \chi_1^{p_1} \chi_2^{p_2} \cdots \chi_d^{p_d} = 0 \quad (8)$$

여기서 p_1, p_2, \dots, p_n 은 0이상의 정수이며, 그것들 전부의 합은 k 이하이다. d 는 공간의 차원이고 k 는 다항식의階(order)이다. 따라서 임의의 확률함수에 있어서 k 階의 증가의 평균이 0이고 약한 정상(weakly stationary)일 경우 그 확률함수는 k 階의 고유 확률함수(intrinsic random function of order k ; IRF- k)라고 말해진다.

확률함수 $Z(\chi)$ 가 비정상 상태일때, 그 함수는 평균값(drift)과 잔류값(residual)의 합으로 표현된다.

$$Z(\chi) = m(\chi) + Y(\chi) \quad (9)$$

이때 $E\{Y(\chi)\} = 0$ 이다. 그리고

$$m(\chi) = E\{Z(\chi)\} = \sum_{i=1}^k a_i f_i(\chi) \quad (10)$$

여기서 k 는 drift의 계수(order)로서 0, 1 또는 2가 될 수 있다. a_i 는 drift 모델의 미지의 상수이며, $f_i(\chi)$ 는 낮은 계수의 단항식이다. 1차적인 drift에 대해서

$$m(\chi) = a_1 + a_2\chi \quad (1-D) \quad (11)$$

$$m(u, v) = a_1 + a_2u + a_3v \quad (2-D) \quad (12)$$

2차적인 drift에 대해서

$$m(\chi) = a_1 + a_2\chi + a_3\chi^2 \quad (1-D) \quad (13)$$

$$m(u, v) = a_1 + a_2u + a_3v + a_4u^2 + a_5v^2 + a_6uv \quad (2-D) \quad (14)$$

IRF-k는 추정치의 계산을 위해 특정한 variogram 모델대신에 일반적 共分散(generalized covariance : GC)을 이용한다. 즉

$$K(h) = C\delta(h) + \sum_{j=0}^k \alpha_j |h|^{2j+1} \quad (15)$$

여기서

$$\delta(h) = \begin{cases} 0, & (|h| > 0) \\ 1, & (|h| = 0) \end{cases} \quad (16)$$

k의 값에 따른 일반적 공분산(GC)은 다음과 같다.

k = 0일 경우

$$GC = C\delta(h) + \alpha_0 |h| \quad (17)$$

k = 1,

$$GC = C\delta(h) + \alpha_0 |h| + \alpha_1 |h|^3 \quad (18)$$

k = 2,

$$GC = C\delta(h) + \alpha_0 |h| + \alpha_1 |h|^3 + \alpha_2 |h|^5 \quad (19)$$

상수의 한계는

$$\alpha_0 \leq 0, \alpha_2 \leq 0, \alpha_1 \geq -\sqrt{10}\sqrt{\alpha_0\alpha_2}$$

(for 3-D) 또는 $\alpha_1 \geq -\frac{10}{3}\sqrt{\alpha_0\alpha_2}$ (for 2-D)이다.

이 상수의 값들을 Delfiner(1976)는 가중최

소 자승법(weighted least squares method)에 의해서 결정하였다. Delfiner와 Delhomme(1973)은 IRF-k를 위한 컴퓨터 BLUEPACK-3D를 개발하였고, Kafritas와 Bras(1981)는 Delfiner의 알고리즘에 기초하여 반복회귀 접근(iterative regression approach)를 이용하여 컴퓨터 코드 AKRIP을 개발하였다.

실험자료의 IRF-k에 응용

미국 네바다주에서 얻어진 38개의 지하수위 자료를 가지고 자료가 없는 지점에서 추정치를 얻기 위해 IRF-k를 이용한다. 이 IRF-k의 실험자료에 대한 적합성을 판단하기 위해 ordinary kriging에 의한 추정치와 비교한다. 이 때에는 실험자료를 정상 확률함수로 가정하여 球狀의 variogram 모델을 이용한다.

비정상 확률함수를 이용한 未知의 지점에서 자료의 추정에 IRF-k와 ordinary kriging를 비교하기 위해 교차 타당성(cross validation)을 이용한 통계학적인 방법과 격자상의 추정치로부터 작성된 지하수위 등고선도와 추정 분산등고선도등의 그림에 의한 방법을 이용한다.

일반적 공분산(GC)의 결정

실험자료로부터 결정된 drift의 계수는 1 이고 일반적 공분산의 계수는 $C=15995$, $\alpha_0 = -33782$, $\alpha_1 = \alpha_2 = 0$ 이다. 따라서 이용될 GC는 다음과 같다.

$$GC = 15995 - 33782 |h|$$

교차타당성(cross validation)

실험자료로부터 선정된 variogram 모델이나

공분산모델의 적합성을 판정하기 위해서 이용되는 것이 교차타당성이다. 그 절차는 다음과 같다(Davis, 1987).

첫째, 1개의 자료를 전체의 실험자료에서 뽑아낸다. 그 지점에서 나머지 자료들과 선정된 variogram이나 공분산모델을 가지고 Kriging을 이용하여 값을 추정해낸다.

둘째, 추정치의 오차를 알고있는 참값(true value)으로 부터 계산한다. Kriging 분산을 이용하여 감소오차(reduced error)를 계산한다. 즉

$$RE = \frac{Z(\chi) - Z^*(\chi)}{\sqrt{\sigma_k^2}} \quad (20)$$

선정된 모델이 가장 적합하다면 감소오차의 평균(Mean Reduced Errors, MRE)이 0에 가까워야 하고 감소오차의 표준편차(Standard Deviation of Reduced Errors, SDRE)가 1.0에 가까워야 한다(Davis, 1987; Solow, 1990). 즉

$$MRE = \frac{1}{N} \sum_{i=1}^n \left[\frac{Z(\chi) - Z^*(\chi)}{\sqrt{\sigma_k^2}} \right] \approx 0 \quad (21)$$

$$VRE = \frac{1}{N-1} \sum_{i=1}^n (RE - MRE)^2 \quad (22)$$

$$SDRE = \sqrt{VRE} \approx 1.0 \quad (23)$$

몇가지의 다른 통계값들도 선정된 모델의 적합성을 판정하기 위해 이용된다.

1. 평균 오차는 0에 가까워야 한다.

$$ME = \frac{1}{N-1} \sum_{i=1}^n [Z(x) - Z^*(\chi)] \quad (24)$$

2. 평균 제곱오차는 최소가 되어야 한다.

$$MSE = \frac{1}{N} \sum_{i=1}^n [Z(x) - Z^*(\chi)]^2 \quad (25)$$

3. 평균 제곱오차의 제곱근 역시 최소가 되어야 한다.

$$SRMSE = \sqrt{MSE} \quad (26)$$

4. 평균 추정분산(mean estimation variance) 또는 kriging 분산이 최소가 되어야 한다.

$$MKV = \frac{1}{N} \sum_{i=1}^{N-1} (\sigma_k^2)_i \quad (27)$$

5. 오차의 분산은 평균 Kriging 분산과 같아야 한다(Samper, 1986).

$$VE = \frac{1}{N-1} \sum_{i=1}^N (Error - ME)^2 = MKV \quad (28)$$

표1은 ordinary kriging과 IRF-k에서 교차타당성을 위해 얻어낸 통계적 값들을 보여준다. MRE는 IRF-1이 OK보다 0에 가까운 값을 갖고, SDRE는 IRF-1이 OK보다 1.0에 훨씬 가까운 값을 갖는다. ME와 MKV에 있어서 IRF-1이 OK보다 적은 값을 갖고, MSE, SRMSE 및 VE에서는 OK가 IRF-1보다 적은 값을 갖는다.

등고선도

그림 3과 4는 각각 ordinary kriging과 IRF-1으로 부터 만들어진 등고선도이다. 전반적인 등압력선들의 분포상태는 비슷하지만 2개의 그림을 겹쳐보면 세부적으로 많은 차이를 보인다. 그림 5와 6은 각 등고선도에 이용된 격자상의 각 추정치들이 갖는 추정 분산치의 분포를 나타내준다. ordinary kriging이 IRF-1

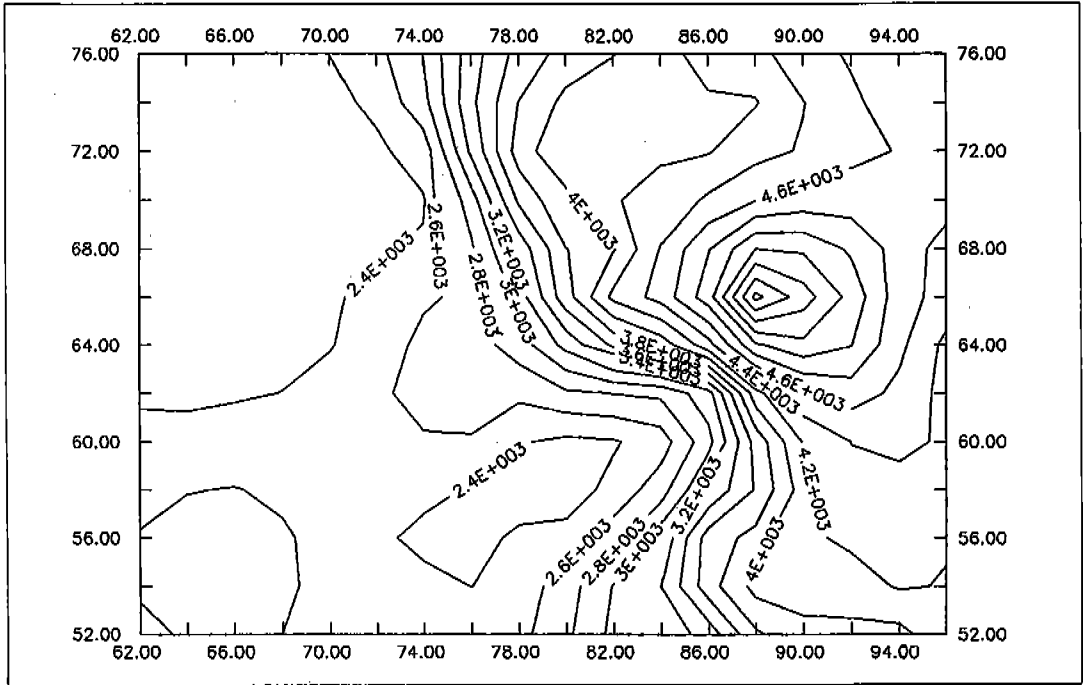


Fig. 3 Contour map of groundwater levels produced by ordinary kriging.

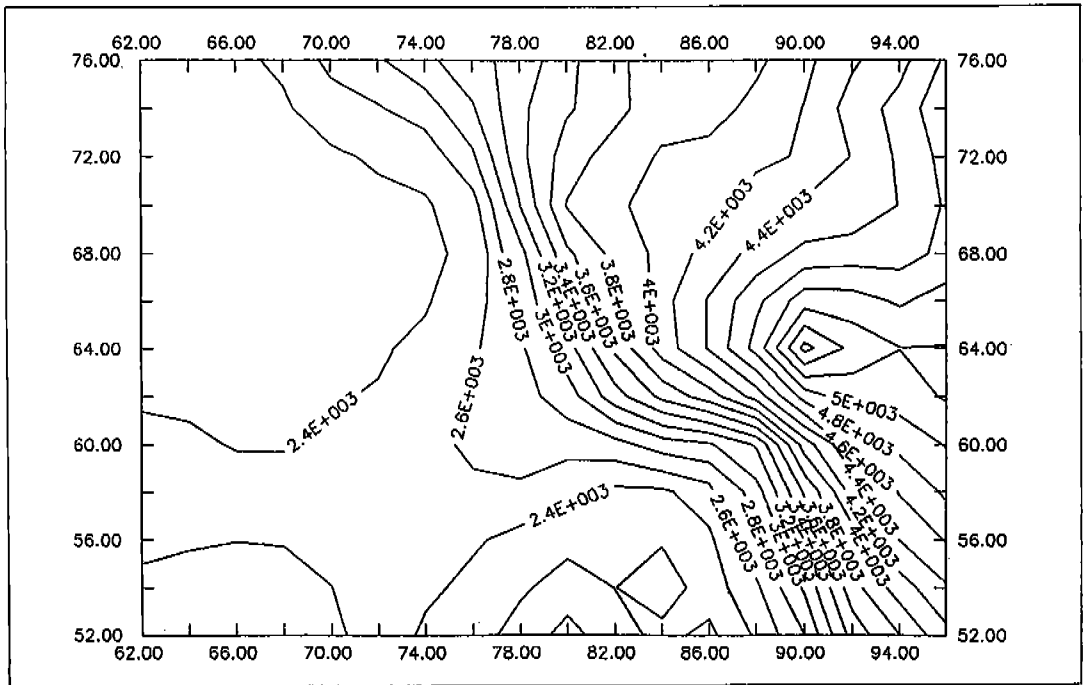


Fig. 4 Contour map of groundwater levels produced by IRF-k

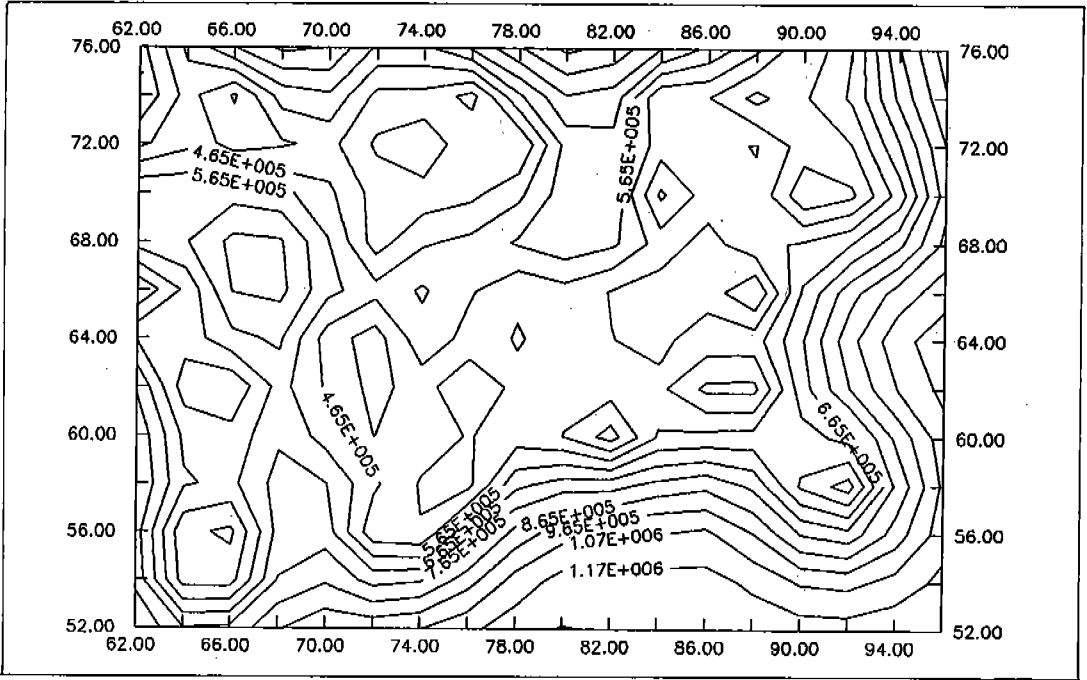


Fig. 5 Contour map of estimation variance produced by ordinary kriging.

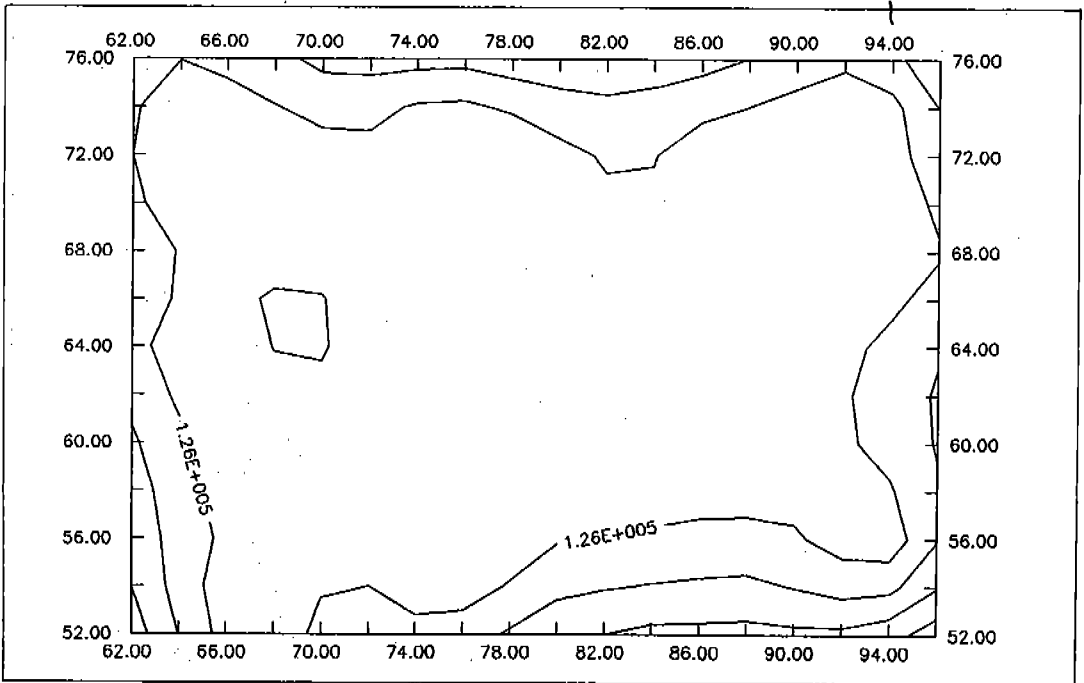


Fig. 6 Contour map of estimation variance produced by IRF-k

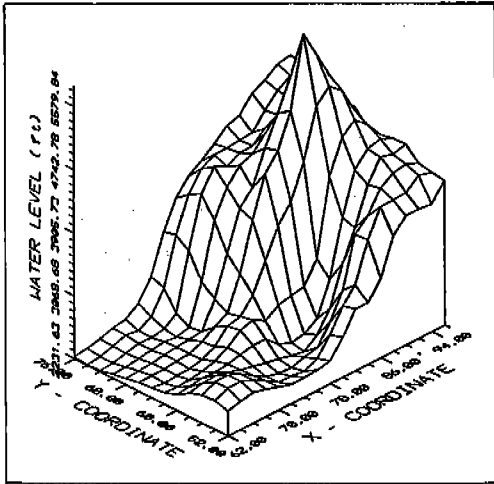


Fig. 7 Surface map of groundwater levels produced by ordinary kriging.

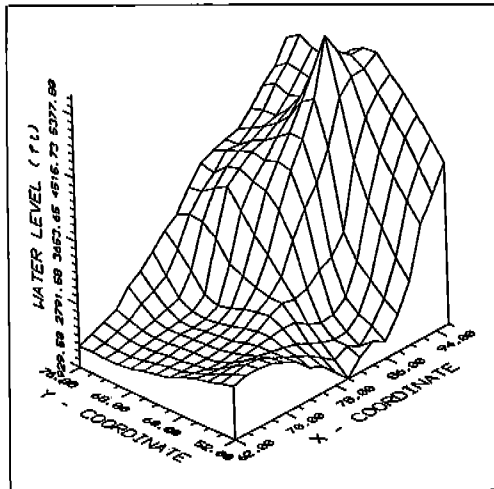


Fig. 8 Surface map of groundwater levels produced by IRF-k

에 비해서 큰 추정 분산치를 갖는다. 그림 7과 8은 ordinary kriging과 IRF-1에 의해서 만들어진 추정치를 3차원의 표면에서 비교해보기 위한 그림이다. 전체적으로 비슷해 보이지만 2개의 그림을 겹쳐보면 세부적으로 많은 차이가 난다.

토의 및 결론

교차타당성에 의하면 오차의 분산(VE)과 평균 자승오차의 제곱근(SRMSE)에 있어서만 ordinary kriging이 IRF-1보다 적은 값을 갖는다. 특히 교차타당성에 있어서 평균 감소오차(MRE)와 감소오차의 표준편차(SDRE)가 모델의 적합성 판정에 중요한 역할을 하는데, IRF-1이 ordinary kriging에 비해 좋은 결과를 보여준다. 격자상의 추정치에서 구해진 추정 분산값(estimation variance)을 보면, ordinary kriging이 IRF-1보다 훨씬 큰 분포를 갖는다. 좋은 모델은 최소의 추정분산치를 갖는다.

지하수위 등고선도와 3차원의 표면도를 보면 ordinary kriging과 IRF-1이 세부적으로 많은 차이를 갖는다. 지하수위 유동방향, 지하수유량, 지하수 유동속도 등이 유선망으로부터 구해질 수 있다. 이 유선망은 등압선을 이루는 지하수위에서 결정되므로, 정밀한 지하수위분포의 추정이 정밀한 지하수 상태의 파악을 가능하게 한다.

본 연구의 결론을 요약하면 다음과 같다.

- 1) 미국 네바다주에서 얻은 지하수위 자료는 1개의 비정상 확률함수이다.
- 2) 비정상 확률변수에 대한 자료추정은 IRF-k가 ordinary kriging보다 적은 통계적 오차를 갖는 우수한 방법이다.
- 3) 정밀한 지하수위 등고선 분포도는 정확한 지하수 상태의 파악에 필수적인 요소이다. IRF-k에 의해서 만들어진 지하수위 등고선이 ordinary kriging에 의한 것보다 더 정밀하다.

사 사

본 연구를 위해 지하수위 자료를 제공해 준 미국 네바다주립 연구소 Desert Research Institute에 감사한다.

참고문헌

- Aboufirassi, M., and M.A. Marino, 1983, Kriging of Water Levels in the Souss Aquifer, Morocco, *Mathematical Geology*, 15(4), pp. 537-551.
- Aboufirassi, M., and M.A. Marino, 1984, Cokriging of Aquifer Transmissivities from Field Measurements of Transmissivity and Specific Capacity, *Mathematical Geology*, 16(1), pp. 19-35.
- Davis, B.M., 1987, Uses and Abuses of Cross-Validation in Geostatistics, *Mathematical Geology*, 19(3), pp. 241-248.
- Delfiner, P. and Delhomme, J.P., 1973, Presentation du Programme BLUEPACK, Technical Report, Ecole des Mines de Paris, Fontainebleau.
- 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, pp. 49-68, D. Reidel Publishing Company, Dordrech-Holland.
- Journel, A.G., and C.H.J. Huijbregts, 1978, *Mining Geostatistics*, 600p., Academic Press.
- Kafritas and Bras, 1981, The Practice of Kriging, Ralph M. Parsons Laboratory, Dept of Civil Engineering, MIT Technical Report No. 263.
- Kitanidis, P.A., and E.G. Vomvoris, 1983, A Geostatistical Approach to the Inverse Problems in Groundwater Modeling (Steady State) and One Dimensional Simulations, *Water Research Research*, 19(3), pp. 677-690.
- Philip, R.D., and P.K. Kitanidis, 1989, Geostatistical Estimation of Hydraulic Head Gradients, *Groundwater*, 27(6), pp. 855-865.
- Matheron, G., 1969, *Le Krigeage Univesal*, Cah. Centre Morphol, Math. 1
- Matheron, G., 1970, The Theory of Regionalized Variable and Its Applications, Cah. Centre Morphol. Math. 5
- Matheron, G., 1973, The Intrinsic Random Functions and Their Applications, *Advanced Applied Probability*, Vol. 5 pp. 439-468
- Neuman, S.P., 1984, Role of Geostatistics in Subsurface Hydrology, *Geostatistics for Natural Resources Characterization*, Parts 2, edited by G. Verly et al., pp. 287-816.
- Neuman and Jacobson, 1984, Analysis of Nonintrinsic Spatial Variability by Residual Kriging with Application to Regional Groundwater Levels, *Mathematical Geology*, 16(1), pp. 19-35.
- Solow, A.R., 1990, Geostatistical Cross-Validation: A Cautionary Note, *Mathematical Geology*, 22(6), pp. 637-639.
- Samper, F.J., 1986, *Statistical Methods of Analyzing Hydrological, Hydrochemical, and Isotopic Data from Aquifers*, Ph.D. Dissertation, Dept. of Hydrology and Water Resources, University of Arizona, Tucson.

Yeh, W.W-G., Y.S., Yoon, and K.S.Lee, 1983,
Aquifer Parameter Identification with Kri-
ging and Optimum Parameterization, Wa-
ter Resourses Research, 19(1), pp. 677--
690.

정상용 :

부산수산대학교

해양과학대학, 응용지질학과

부산시 남구 대연3동 599-1

608-023

TEL : (051) 620-6235

FAX : (051) 627-1096