

Indicator 크리깅을 이용한 부산지하수 수질의 오염도 연구

김동환, 정상용, 김병우, 심병원*, 성익환*, 조병욱*

부산광역시 남구 대연3동 599-1 부경대학교 환경·해양대학 환경지질과학과

*한국지질자원연구원 지하수지열연구부 지하수자원연구실 (ak1k1k1@mail1.pknu.ac.kr)

<요약문>

강서구를 제외한 부산 전지역에서 1998년도에 조사된 지하수 수질 중 6개 성분(pH, TS, KMnO₄, Cl, SO₄, NO₃-N)에 대한 일반통계분석 결과 pH 성분을 제외하고는 5개 성분의 중앙값이 평균보다 적은 값을 보이는 양성왜도를 보임으로써, 수질오염정도를 분석하기 위해 지시크리깅이라는 비모수적인 지구통계분석기법을 적용하였다. 6개 수질성분에 대해 음용수 기준치를 적용하여 음용가능은 "1"의 값이, 음용불가능은 "0"의 값이 주어졌다. 이렇게 변환된 자료를 이용하여 각 성분별로 실험적인 베리오그램 분석을 실시한 결과 pH, TS, SO₄ 성분은 선형모델이 선정되었으며, KMnO₄, Cl, NO₃-N 성분은 구상형모델이 선정되었다. 본 연구에서는 지시크리깅을 이용하여 6개 성분의 분포도를 작성하고 부산지역의 오염정도를 분석하였다. 지시크리깅기법은 연구지역 전체의 정량적인 분포를 나타내지는 못하지만, 오염의 유·무와 오염의 크기를 정확하게 파악할 수 있으며 또한, 이상치(outlier)가 크게 영향을 미칠수 있는 통계학적인 오류를 보완할 수 있다.

주요어 : 비모수통계분석, 지시크리깅, 지하수 수질오염.

1. 서론

우리나라 최고의 항구도시인 부산에서는 지하수에 대한 의존도가 상당히 높은 편이다. 그러나, 부산의 지하수는 많은 인구에 의해 발생된 생활하수와 각종 쓰레기, 공장에서 배출된 산업폐수와 자동차 배기 가스 등으로 인한 오염이 심각하며, 또한 바다를 접하고 있어 해수에 의한 오염의 가능성이 높다. 부산의 수질현황을 파악하기 위해 부산시 보건환경 연구원에서 1998년도에 조사된 수질자료 중에서 pH, TS, KMnO₄ 소비량(이하 KMnO₄), Cl, SO₄, NO₃-N 성분의 수질자료를 이용하였으며, 강서구는 자료가 부족하여 제외되었다.

모수적 통계기법인 정규크리깅(ordinary kriging)은 모집단의 정규분포 가정이 만족되는 경우에는 좋은 추정치를 구할수 있지만(심병원 외, 2000; 정상용 외, 1999), 모집단의 분포가 정규분포와 다른 왜곡된 분포를 갖는 경우에는 오류를 범할 확률이 높다. 즉, 특정지점의 이상치(outlier)에 의한 영향성이 매우 크게 나타날 수 있다는 것이다. 이러한 단점을 줄이기 위해 모집단의 분포에 대한 아무런 가정 없이 이론을 전개해 나가는 비모수적 통계기법인 지시크리깅(indicator kriging)을 이용하여 수질자료의 추정치를 구하였다. 지시크리깅을 적용한 사례로는 현장에서 작성된 토양지질로깅 자료를 이용하여 토양의 투수성분포도를 작성한 연구들이 있다(Johnson et al, 1989; Sminchak et al, 1996). 그리고, 지시크리깅

의 비모수적 통계분석방법으로서의 수학적 연구는 1990년대 초·중반에 활발하게 연구되었다(Carr and Mao, 1993; Goovaerts, 1994; Posa and Marcotte, 1992). 본 연구에서는 지시크리깅 분석을 위해 먼저 일반통계분석을 실시하여 수질자료들이 정규분포합수를 이루는지를 판단한 후, 수질자료값들을 "0"과 "1"의 값으로 변환하였다. 또한, 자료들의 분포특성과 공간적 연속성을 파악하기 위하여 베리오그램 분석을 실시하였다. 본 연구에서 변환된 수질성분 자료들이 정상성(stationarity)을 가지는 것으로 판명되어, 지시크리깅을 적용하여 수질성분별 분포도를 작성하여 오염정도를 분석하였다.

2. 본 론

2.1. 지시크리깅 이론

지구통계학에서의 기본적인 전제조건은 변수들의 공간적인 특성(spatial structure)이 베리오그램(variogram)에 의해 정의되어 진다는 것이다(Journel and Huijbregts, 1978). 2차 정상성(second-order stationarity)으로 가정하면, 베리오그램은 자료의 위치에는 무관하며 자료들간의 거리(separation distance)와 방향(direction)의 함수가 된다. 실험적인 베리오그램($\gamma^*(h)$)은 확률변수 $Z(x)$ 의 이산된 조사값들인 $z(x_i)$ 로부터 식 (1)과 같이 계산되어 진다.

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i+h) - z(x_i)]^2 \quad (1)$$

여기서 $N(h)$ 는 동일거리 h 만큼 떨어진 자료쌍들의 개수이다.

지시크리깅(indicator kriging)은 특정한 한계치(threshold value)보다 적거나 혹은 동일한 변수값들의 확률을 추정하기 위해 자료를 변환(transform)한다(Johnson et al, 1989). 제한값(cutoff value)이 z_c 인 2진지시함수(binary indicator function)는 식 (2)와 같이 정의되어 진다.

$$\begin{aligned} I(x; z_c) &= 1 && \text{if } Z(x) \leq z_c \\ I(x; z_c) &= 0 && \text{if } Z(x) > z_c \end{aligned} \quad (2)$$

확률함수(random function) $I(x; z_c)$ 의 평균(mean)과 분산(variance)은 식 (3)과 (4)와 같이 표현되어 질 수 있다.

$$\begin{aligned} E [I(x; z_c)] &= 1 P [Z(x) \leq z_c] + 0 P [Z(x) > z_c] \\ &= P [Z(x) \leq z_c] = m(z_c) \end{aligned} \quad (3)$$

$$\text{Var} [I(x; z_c)] = m(z_c)[1 - m(z_c)] \quad (4)$$

지시함수의 실험적인 베리오그램은 식 (5)와 같이 정의되어 진다.

$$\gamma_I^*(h; z_c) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [I(x_i+h; z_c) - I(x_i; z_c)]^2 \quad (5)$$

베리오그램에 이용되는 자료값들은 [0,1]사이의 값으로 변환되어 지며, 이를 이용하여 베리오그램 모델이 선정되어 진다. 2진지시크리깅은 선정된 베리오그램 모델을 이용하여 제한값보다 적거나 동일한 값들에 대한 확률의 추정으로 계산되어 진다(Johnson et al, 1989).

2.2. 일반통계와 베리오그램 분석

각 수질성분별 자료들에 대한 일반통계분석 결과가 Table 1에 정리되어 있다. pH 성분을 제외하고는 모든 성분의 중앙값(median)이 평균(mean)보다 적은 값을 보이는 양성왜도를 보이고 있다. 즉, pH 성분은 정규분포함수로의 가정이 적합하나 나머지 성분들은 정규분포함수로의 가정이 이론상 부적합한 것으로 판단된다.

Table 1. Basic statistics of data used for varigram and indicator kriging

Component Statistics	pH	TS	KMnO4	Cl	SO ₄	NO ₃ -N
Number of data	821	910	906	820	822	912
Range	4.80	5577.00	300.20	527.00	1136.00	48.00
Minimum	4.80	12.00	0.00	1.00	0.00	0.00
Maximum	9.60	5589.00	300.20	528.00	1136.00	48.00
Mean	7.23	236.83	1.80	37.71	32.14	4.29
Median	7.20	204.00	0.60	29.00	23.00	3.40
Standard deviation	0.52	232.96	13.14	41.93	54.38	4.03
Variance	0.27	54272.10	172.72	1758.31	2957.56	16.23
Skewness	0.01	14.36	18.21	5.61	13.20	2.94
Kurtosis	0.62	310.82	366.27	46.87	237.74	20.07

각 수질성분별 제한값은 음용수 기준치를 적용하여 TS 성분은 500mg/ℓ, KMnO₄ 성분은 10mg/ℓ, Cl 성분은 150mg/ℓ, SO₄ 성분은 200mg/ℓ, NO₃-N 성분은 10mg/ℓ의 값이 주어졌다. 제한값보다 적거나 동일한 값을 가진 지점에서는 “1”의 값이 주어지고, 제한값보다 높은 값을 가진 지점에서는 “0”의 값이 주어진다. 그리고, pH 성분은 5.8~8.5 사이의 값을 가진 지점에서는 “1”의 값이, 나머지는 “0”의 값이 주어진다. 이러한 방식으로 변환된 자료를 이용하여 각 성분별로 실험적인 베리오그램 분석을 실시하여 선정된 모델과 매개변수들이 Table 2에 정리되어 있다. pH, TS, SO₄ 성분은 선형모델(linear model)이 선정되었으며, KMnO₄, Cl, NO₃-N 성분은 구상형모델(spherical model)이 선정되었다.

Table 2. Varigram model and parameters of chemical components for indicator kriging

Component	Model	Nugget (C ₀)	Sill (C ₀ +C)	Effective range	R ²	RSS
pH	linear	0.00	0.00	32.54	0.87	2.58×10 ⁻³
TS	linear	0.02	0.02	32.51	0.04	1.47×10 ⁻⁴
KMnO ₄	spherical	0.00	0.01	2.02	0.00	6.75×10 ⁻³
Cl	spherical	0.01	0.04	69.37	0.75	1.07×10 ⁻⁴
SO ₄	linear	0.00	0.00	32.52	0.71	4.29×10 ⁻³
NO ₃ -N	spherical	0.02	0.07	42.11	0.94	1.21×10 ⁻⁴

2.3. 수질성분별 오염분포도

Figure 1에서 6에는 선정된 베리오그램 모델을 적용하여 추정한 각 성분별 추정치를 이용하여 작성된 오염분포도가 도시되어 있다. pH 성분은 국부적으로 3개 지점에서 오염이 발생한 것으로 나타났다. TS 성분과 Cl 성분은 해안가를 따라 오염정도가 심하게 나타났으며, TS 성분은 내륙에서도 오염분포가 나타났다. KMnO₄ 성분은 3개 지점 정도에서 오염이 심한 것으로 나타났으며, 기장군을 제외한 지역에서

산발적으로 약간의 오염이 발생한 것으로 나타났다. SO₄ 성분은 5개 지점 정도에서 오염원이 발생되어 있는 것으로 나타났다. 그리고, NO₃-N 성분은 기장군과 해안가 주변 지역에서 광범위하게 오염되어 있는 것으로 나타났다.

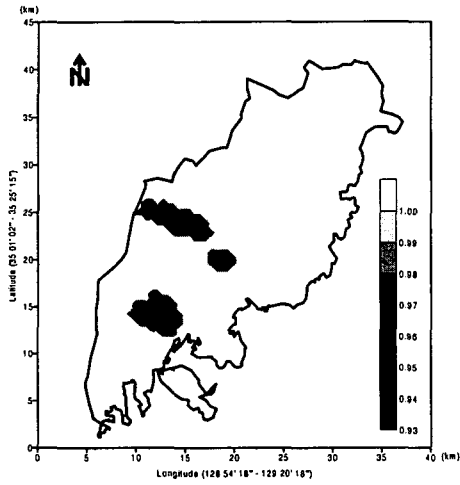


Figure 1. Contamination map of pH.

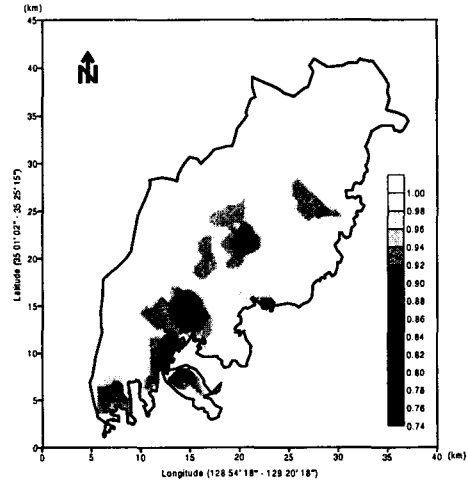


Figure 2. Contamination map of TS.

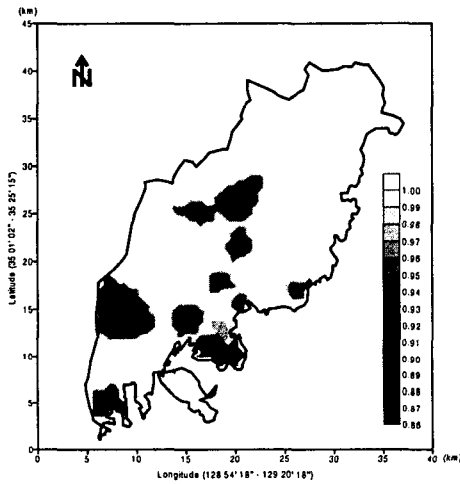


Figure 3. Contamination map of KMnO₄.

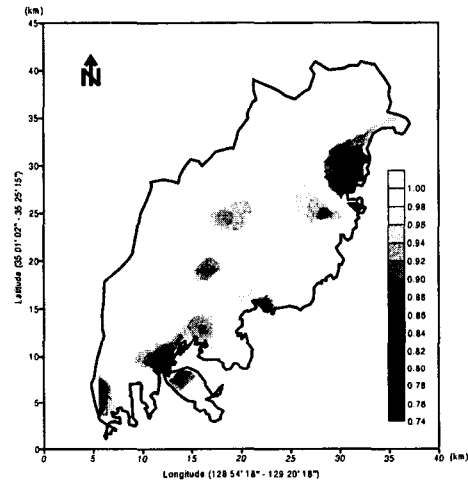


Figure 4. Contamination map of Cl.

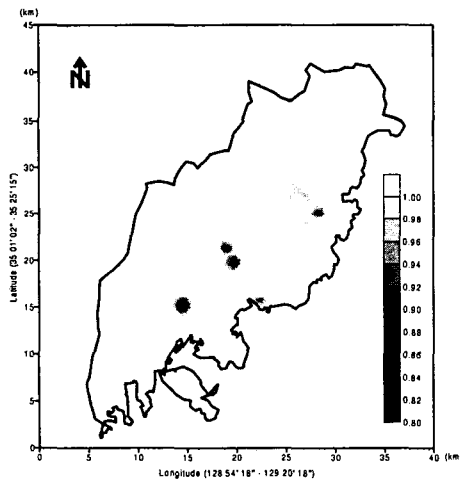


Figure 5. Contamination map of SO₄.

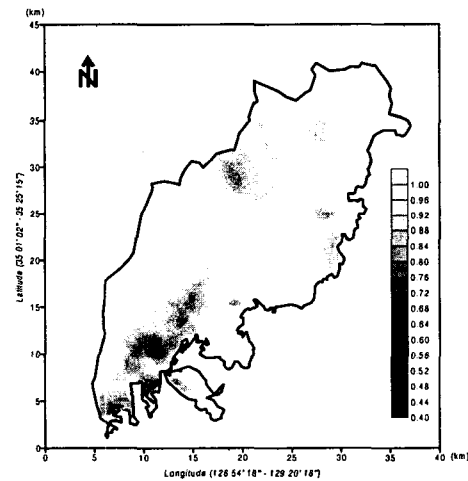


Figure 6. Contamination map of NO₃-N.

3. 결론

각 수질성분별 자료들에 대한 일반통계분석 결과 pH 성분을 제외하고는 모든 성분의 중앙값이 평균보다 적은 값을 보이는 양성왜도를 보임으로써, 정규분포함수로의 가정은 이론상 부적합한 것으로 판단되었다. 각 수질성분별 제한값을 음용수 기준치인 TS 500mg/l, KMnO₄ 10mg/l, Cl 150mg/l, SO₄ 200mg/l, NO₃-N 10mg/l 을 적용하여 제한값보다 적거나 동일한 값을 가진 지점에서는 “1”의 값이 주어졌으며, 제한값보다 높은 값을 가진 지점에서는 “0”의 값이 주어졌다. 그리고, pH 성분은 5.8~8.5 사이의 값을 가진 지점에서는 “1”의 값이, 나머지는 “0”의 값이 주어졌다. 변환된 자료를 이용하여 각 성분별로 실험적인 베리오그램 분석을 실시한 결과 pH, TS, SO₄ 성분은 선형모델이 선정되었으며, KMnO₄, Cl, NO₃-N 성분은 구상형모델이 선정되었다.

선정된 베리오그램 모델을 적용하여 추정된 각 성분별 추정치를 이용하여 오염분포도를 작성하여 분석한 결과, pH 성분은 국부적으로 3개 지점에서 오염이 발생한 것으로 나타났다. TS 성분과 Cl 성분은 해안가를 따라 오염정도가 심하게 나타났으며, TS 성분은 내륙에서도 오염분포가 나타났다. KMnO₄ 성분은 3개 지점에서 오염이 심한 것으로 나타났으며, 기장군을 제외한 지역에서 산발적으로 약간의 오염이 발생한 것으로 나타났다. SO₄ 성분은 5개 지점 정도에서 오염원이 발생되어 있는 것으로 나타났다. 그리고, NO₃-N 성분은 기장군과 해안가 주변 지역에서 광범위하게 오염되어 있는 것으로 나타났다.

지시크리깅을 이용한 수질오염분석은 연구지역 전체의 정량적인 분포를 나타내지는 못하지만, 오염의 유·무와 오염의 크기를 정확히 파악할 수 있으며 또한, 이상치가 크게 영향을 미칠수 있는 통계학적인 오류를 보완한다.

4. 사 사

본 연구는 1998~1999년도 한국지질자원연구원의 위탁연구비와 1998년도 한국학술진흥재단의 학술연구비의 지원을 받았습니다.

5. 참고문헌

- 1) 심병완, 정상용, 강동환, 김규범, 박희영, 2000, 영산강·섬진강 유역의 지하수 데이터베이스 자료에 대한 지구통계학적 분석, 지질공학회지, 10(2), p.131-142.
- 2) 정상용, 유인걸, 윤명재, 권해우, 허선희, 1999, 불균질·이방성 대수층의 지하수 유동분석에 지구통계기법의 응용, 지질공학회지, 9(2), p.147-159.
- 3) Goovaerts, P., 1994, Comparative Performance of Indicator Algorithms for Modeling Conditional Probability Distribution Functions, Mathematical Geology, 26(3), p.389-411.
- 4) Johnson, N. M., and S. J. Dreiss, 1989, Hydrostratigraphic Interpretation Using Indicator Geostatistics, Water Resources Research, 25(12), p.2501-2510.
- 5) Journel, A. J., and C. Huijbregts, 1978, Mining Geostatistics, Academic, San Diego, Calif., 600p.
- 6) Sminchak, J. R., David F. Dominic, and Robert W. Ritzi, Jr., 1996, Indicator Geostatistical Analysis of Sand Interconnections Within a Till, Groundwater, 34(6), p.1125-1131.
- 7) Posa, D., and D. Marcotte, 1992, Robustness of Kriging Weights to Non-Bias Conditions, Mathematical Geology, 24(7), p.759-773.
- 8) Carr, J. R., and Nai-hsien Mao, 1993, A General Form of Probability Kriging for Estimation of the Indicator and Uniform Transforms, Mathematical Geology, 25(4), p.425-438.