

## 토양층 내 Cu, Pb 농도의 평균 및 중앙값을 이용한 지시크리깅의 확률도 비교 분석

유훈선<sup>1</sup> · 정상용<sup>2</sup> · 강동환<sup>2</sup> · 김병우<sup>3</sup> · 김철<sup>3</sup>

<sup>1</sup>동의과학대학 동의분석센터

<sup>2</sup>부경대학교 환경지질학과

<sup>3</sup>동의과학대학 환경보건과

yuhunsun@naver.com

### 요 약 문

본 연구에서는 공단지역 내 토양층(지표면하 1 m 이내)에 존재하는 Cu, Pb 성분의 평균값보다 높은 지역과 중앙값보다 높은 지역을 확률적으로 추정하고자 하였다. 먼저, 연구지역의 Cu, Pb 성분의 농도 분포를 파악하기 위해 일반통계분석을 하였고, 공단지역 내 Cu, Pb 성분의 평균값과 중앙값을 제한값으로 하여 지시크리깅을 수행하여 평균값보다 높은 확률도와 중앙값보다 높은 확률도를 작성하였다. 평균값과 중앙값에 의해 작성된 확률도를 비교하여, 지시크리깅의 제한값으로서의 적합성을 분석하였다.

**key words** : 지시크리깅, 평균값, 중앙값, 확률도

### 1. 서론

토양층 내 중금속 성분의 농도값들로 구성된 모집단이 정규분포(normal distribution)를 이루는 경우에는 모수적(parametric) 통계기법인 정규크리깅(ordinary kriging)에 의해 좋은 추정치를 구할 수 있다. 그러나, 모집단(population)이 정규분포를 이루지 못하거나 분포함수를 정의할 수 없는 경우에는 현실적인 대안적인 추정기법을 적용하여야 한다. 이러한 경우의 대안적인 통계기법인 모집단의 분포에 대한 가정 없이 이론을 전개해 나가는 비모수적(nonparametric) 통계기법인 지시크리깅(indicator kriging)에 의한 확률도의 작성이 효과적이다. 지시크리깅을 적용한 사례로는 국내에서는 지하수의 수질오염분포에 적용하였으며(강동환 외, 2003), 또한 토양층 내 중금속 성분의 분포 및 오염가능성(유훈선 외, 2005)에 대한 연구가 수행되었다. 국외에서는 지시크리깅의 이론 및 수학적 연구(Carr and Mao, 1993; Goovaerts, 1994; Posa and Marcotte, 1992)가 수행되었으며, 현장적용사례로는 토양지질로깅 자료를 이용하여 토양의 투수성 분포도를 작성한 연구(Johnson et al, 1989; Sminchak et al, 1996) 등이 있다.

본 연구에서는 확률분포함수(probability distribution function)를 정의할 수 없는 토양층 내 Cu, Pb 성분값들에 대한 확률도(probability map)를 작성하여 평균값보다 높은 지역과 중앙값보다 높은 지역의 범위를 추정하였다. 먼저, Cu, Pb 성분의 값들을 평균값과 중앙값보다 초과되면 "0", 그 외의 경우에는 "1"의 값으로 지시변환(indicated transformation) 하였다. 지시변환된 자료들의 공간적인 상관성과 방향성을 파악하기 위하여 베리오그램 분석을 실시하여, 최적의 베리오그램 모델을 선정하였다. 선정된 모델에 의한 지시크리깅을 수행하여 확률도를 작성하였다. 본 연구를 통해, 평균과 중앙값을 제한값으로 적용한 경우 확률도의 적합성을 비교 분석하였다.

### 2. 본론

#### 2.1. 지시크리깅 이론

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

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

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 변환한다(Johnson et al, 1989). 제한값(cutoff value)이  $z_c$ 인 2진 지시함수(binary indicator function)는 식 (2)와 같이 정의된다.

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

확률함수(random function)  $I(z; x_c)$ 의 평균(mean)과 분산(variance)은 각각 식 (3), (4)와 같이 표현된다.

$$E[I(x; z_c)] = 1P[z(x) \leq z_c] + 0P[z(x) > z_c] = P[z(x) \leq z_c] = m(z_c) \quad (3)$$

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

지시확률함수의 실험적인 베리오그램은 식 (5)와 같이 정의된다.

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

베리오그램에 이용되는 자료값들은  $[0, 1]$  사이의 값으로 변환되어지며, 이를 이용하여 베리오그램 모델이 선정되어진다. 2진 지시크리깅은 선정된 베리오그램 모델을 이용하여 제한값보다 크거나 동일한 값들에 대한 확률의 추정으로 계산된다(Johnson et al, 1989).

## 2.2. 시료채취 및 일반통계분석

연구지역은 울산시에 위치하고 있는 울산석유화학공단, 여천공단, 용연공단이며, 울산석유화학공단에서 27개 지점, 여천공단 25개 지점, 용연공단 26개 지점을 격자방식으로 선정하였다. 지표면하 1 m 까지 시료를 채취하여 토양층 내 성분의 정확한 농도값을 구하기 위해 0.0~0.2 m,

0.2~0.5 m, 0.5~1.0 m로 시료를 나누어 분석한 후, Cu, Pb 성분값의 최소값, 최대값, 평균값, 중앙값, 왜도 및 첨도를 정리하였다(Table 1). 일반통계분석 결과, 최소값은 두 성분 모두 0으로 나타났다으며, 최대값은 Cu 19.718 mg/kg, Pb 43.617 mg/kg, 평균값은 Cu 3.900 mg/kg, Pb 8.966 mg/kg, 중앙값은 Cu 2.687 mg/kg, Pb 6.617 mg/kg으로 나타났다. 연구지역 내에는 Cu 성분에 비해 Pb 성분의 농도가 높게 분포하고 있으며, 두 성분 모두 농도의 평균보다 낮은 값의 개수가 많이 존재하고 있다.

Table 1. General statistic of Cu, Pb components in soil of study area

Statistic \ Component	Cu	Pb
Minimum (mg/kg)	0.000	0.000
Maximum (mg/kg)	19.718	43.617
Mean (mg/kg)	3.900	8.966
Median (mg/kg)	2.687	6.617
Skewness	2.389	2.077
Kurtosis	7.230	6.463

### 2.3. 성분별 classed post map

Cu 성분의 평균값인 3.900 mg/kg 보다 높은 지점은 '▲'으로 표기하였으며, 중앙값인 2.687 mg/kg 보다 크고 평균값인 3.900 mg/kg보다 작은 지점은 '⊕'으로 표기하였으며, 중앙값인 2.687 mg/kg 보다 작은 지점은 '+'으로 표기하였다(Fig. 1(left)). Pb 성분도 평균값인 8.966 mg/kg 보다 높은 지점은 '▲', 중앙값인 6.617 mg/kg 보다 크고 평균값인 8.966 mg/kg 보다 작은 지점은 '⊕', 중앙값인 2.687 mg/kg 보다 작은 지점은 '+'으로 표기하였다(Fig. 1(right)). 작성된 classed post map의 분포를 비교하여 보면, Cu 성분은 평균값보다 높은 값들이 상대적으로 많으며, Pb 성분은 평균과 중앙값 사이의 값들이 상대적으로 많이 존재하고 있다.

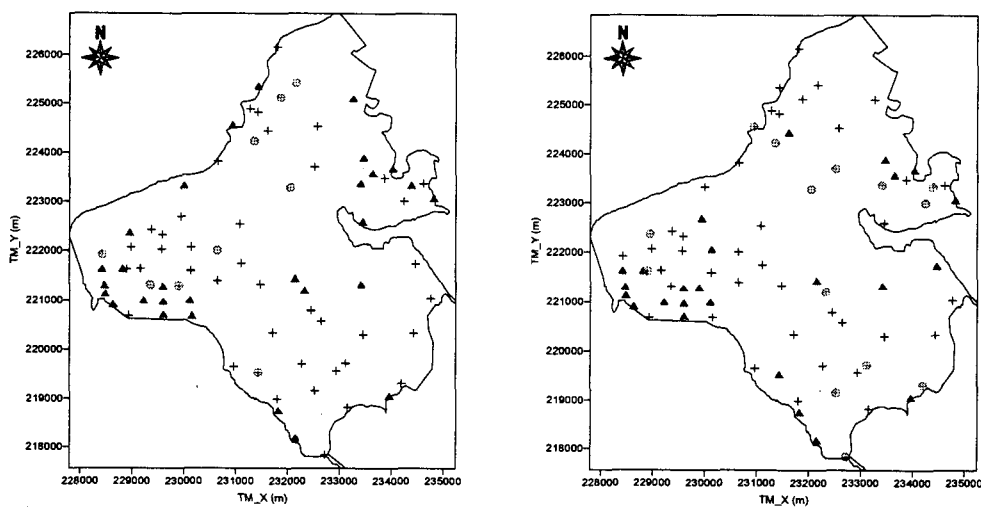


Fig. 1. Classed post maps of Cu (left) and Pb (right) components.

(▲ : value > mean, ⊕ : median < value ≤ mean, + : value ≤ median)

#### 2.4. 베리오그램 작성

지시변환된 자료의 베리오그램을 작성하기 위하여 Cu, Pb 성분의 지시변환자료에 대한 베리오그램 분석을 수행하였다. 평균에 의해 지시변환된 Cu 성분의 베리오그램은 지수형모델(exponential model)이 가장 적합하게 나타났으며, 이 외의 베리오그램은 구상형모델(spherical model)이 가장 적합한 것으로 나타났다.

Table 2. Variogram parameters of Cu, Pb components

Cut-off values	Heavy metal	Model	Effective range (m)	Nugget	Sill	R <sup>2</sup>	RSS
Mean	Cu	Exponential	960	0.033	0.240	0.702	$7.287 \times 10^{-04}$
	Pb	Spherical	590	0.011	0.220	0.035	$2.655 \times 10^{-03}$
Median	Cu	Spherical	950	0.012	0.257	0.808	$1.167 \times 10^{-03}$
	Pb	Spherical	600	0.012	0.253	0.061	$2.069 \times 10^{-03}$

#### 2.5. 평균과 중앙값에 의한 확률도의 비교분석

Cu 성분의 평균과 중앙값을 제한값으로 적용하여 작성한 확률도는 Fig. 2에 도시하였으며, 제한값으로 중앙값을 적용하였을 때의 분포도가 적합하게 나타났다. 평균을 제한값으로 적용한 경우의 확률도에서는 이상대(원으로 표기)가 나타났다. Pb 성분의 평균과 중앙값을 제한값으로 적용하여 작성한 확률도는 Fig. 3에 도시하였으며, 제한값으로 중앙값을 적용한 확률도가 더욱 적합하게 나타났다. 자료가 부족한 지점들에서의 이상대(원으로 표기)는 평균값을 적용한 경우의 확률도에서 많이 발생하였다.

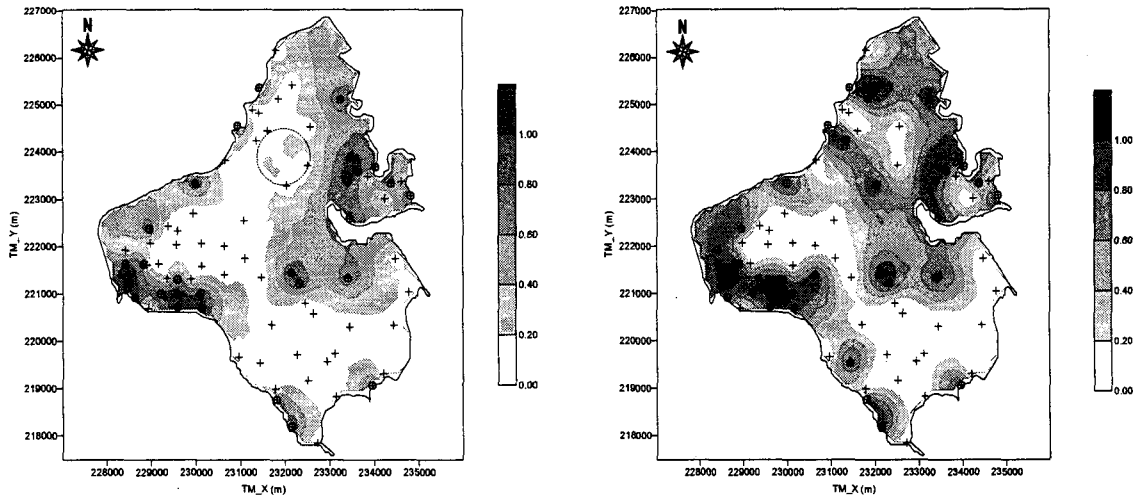


Fig. 2. Probability maps of mean (left) and median (right) using Cu component.

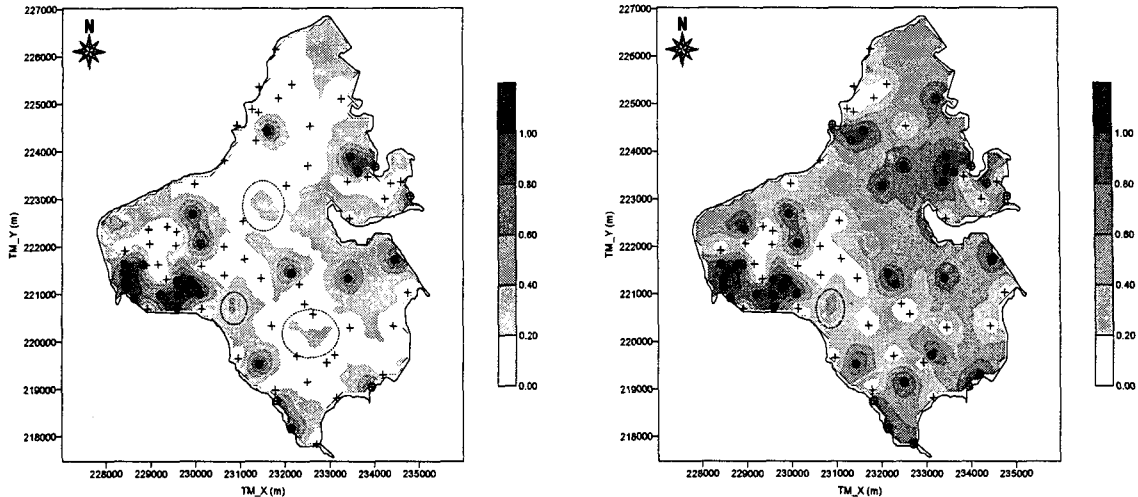


Fig. 3. Probability maps of mean (left) and median (right) using Pb component.

### 3. 결론

Cu, Pb 성분의 일반통계분석 결과, 최소값은 두 성분 모두 0으로 나타났으며, 최대값은 Cu 19.718 mg/kg, Pb 43.617 mg/kg, 평균값은 Cu 3.900 mg/kg, Pb 8.966 mg/kg, 중앙값은 Cu 2.687 mg/kg, Pb 6.617 mg/kg으로 나타났다. 연구지역 내에서는 Cu 성분에 비해 Pb 성분의 농도가 높게 나타났으며, 두 성분 모두 농도의 평균보다 낮은 값의 개수가 많았다.

지시변환된 Cu, Pb 성분자료의 베리오그램 분석결과, 평균에 의해 지시변환된 Cu 성분의 베리오그램은 지수형모델이 가장 적합하게 나타났으며, 이 외의 베리오그램은 구상형모델이 가장 적합한 것으로 나타났다.

Cu 및 Pb 성분의 평균과 중앙값을 제한값으로 적용하여 작성한 확률도를 분석한 결과, 제한값으로 중앙값을 적용하였을 때의 분포도가 더욱 적합한 것으로 판단되었다. 평균을 제한값으로 적용한 경우의 확률도에서는 현실적이지 못한 이상대가 많이 발생하였다.

### 4. 참고문헌

- 10) 강동환, 2003. Indicator 크리깅을 이용한 부산 지하수 수질의 오염도 연구, 한국지하수토양환경학회 2003년도 총회 및 추계학술발표회, p.249-253.
- 11) 유훈선, 2005. 크리깅을 이용한 울산미포산업단지의 토양층 내 중금속성분의 농도분포 및 오염가능성 연구, 2005년 대한지질학회 제 60차 정기총회 및 추계학술발표회, p.176.
- 12) 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.
- 13) Goovaerts, P., 1994, Comparative Performance of Indicator Algorithms for Modeling Conditional Probability Distribution Functions, *Mathematical Geology*, 26(3), p.389-411.
- 14) Johnson, N. M., and S. J. Dreiss, 1989, Hydrostratigraphic Interpretation Using Indicator Geostatistics, *Water Resources Research*, 25(12), p.2501-2510.
- 15) Sminchak, J. R., David F. Dominic, and Robert W. Ritz, Jr., 1996, Indicator Geostatistical Analysis of Sand Interconnections Within a Till, *Groundwater*, 34(6), p.1125-1131.
- 16) Posa, D., and D. Marcotte, 1992, Robustness of Kriging Weights to Non-Bias Conditions, *Mathematical Geology*, 24(7), p.759-773.