

현장 조사 자료의 공간 보간을 위한 다변량 크리깅을 이용한 범주형 자료의 통합

Integration of Categorical Data using Multivariate Kriging for Spatial Interpolation of Ground Survey Data

박 노 옥*

No Wook Park

요약 이 논문에서는 공간적으로 소수의 지점에서 획득된 현장 조사 자료의 공간 보간 과정에 범주형 자료를 결합하는 다변량 크리깅 기법을 제안하고자 한다. 범주형 자료를 결합하는 과정에서 기존 범주형 자료의 속성별로 대푯값을 할당하는 단일 지역 평균 기반의 단순 크리깅 방식 대신에, 영역-점 변환 크리깅을 이용하여 원하는 해상도로 상세화시킨 추정값을 가변적 지역 평균으로 이용하였다. 지화학 원소 구리의 공간 보간에 지질도를 이용하는 사례연구를 통해 제안 기법을 제시하였다. 교차 검증 결과, 제안 기법이 단변량 정규 크리깅과 기존 단일 지역 평균 기반의 단순 크리깅 기법에 비해 각각 15%와 25%의 예측 능력의 향상을 나타내었다. 따라서 범주형 자료를 부가 자료로 이용하는 공간 보간에 이 논문에서 제안한 기법이 효율적으로 적용될 수 있을 것으로 기대된다.

키워드 : 크리깅, 상세화, 공간 보간, 범주형 자료, 지질도

Abstract This paper presents a multivariate kriging algorithm that integrates categorical data as secondary data for spatial interpolation of sparsely sampled ground survey data. Instead of using constant mean values in each attribute of categorical data, disaggregated local mean values at target grid points are first estimated by area-to-point kriging and then are used as local mean values in simple kriging with local means. This algorithm is illustrated through a case study of spatial interpolation of a geochemical copper element with geological map data. Cross validation results indicates that the presented algorithm leads to significant respective improvement of 15% and 25% in prediction capability, compared with univariate ordinary kriging and conventional simple kriging with constant mean values. It is expected that the multivariate kriging algorithm applied in this study would be effectively applied for spatial interpolation with categorical data.

Keywords : Kriging, Disaggregation, Spatial Interpolation, Categorical Data, Geological Map

1. 서론

공간적으로 소수의 지점에서 획득된 현장 조사 자료는 보통 관심 영역의 모든 지점에서 값을 예측하는 공간 보간을 통해 일종의 주제도의 형태로 분석에 사용된다. 현장의 많은 지점에서 자료를 획득하게 되면 공간 보간에 수반되는 오차를 줄일 수 있지만, 시간, 비용 등의 측면에서 항상 많은 현장 조사

자료 획득이 가능한 것은 아니다. 공간 보간을 거친 주제도는 응용 목적에 따라 이를 입력 자료로 사용하는 이후 분석에 직접적으로 사용되기 때문에 자료가 획득되지 않은 지점에서의 값 예측에 수반되는 불확실성 혹은 오차는 이후 분석에 영향을 미치게 된다. 따라서 이러한 오차 전파 문제를 완화시키기 위해서는 최대한 신뢰성 높은 공간 보간 결과를 얻어야 한다.

*이 연구는 한국지질자원연구원 기관목적사업 "GIS 기반 국토지질정보시스템 실용화 기술 개발"의 위탁과제 연구비 지원으로 수행되었음.

*인하대학교 지리정보공학과 조교수 nwpark@inha.ac.kr(교신저자)

이러한 공간 보간과 관련하여 많은 기법들이 제안되었으며, 상용/비상용 GIS 소프트웨어에 그 방법론들이 포함되어 공간 보간에 많이 활용되고 있다. 대표적인 방법론으로는 거리역비례 가중치법(inverse distance weighting), 지역 경향면 모델(local trend surface model), 불규칙 삼각망(triangulated irregular network), 크리깅(kriging) 등 다양하다[11, 15]. 이 중 크리깅은 공간 보간 과정에 샘플 자료와 보간 위치의 공간적 자기상관성 정보를 직접적으로 가중치 결정에 사용하는 고급 보간 기법에 해당된다[3].

보간에 의한 값 추정에 수반되는 오차가 거의 없다고 가정하거나 오차를 무시하는 역거리 가중치 법과 같은 기존 결정론적 기법과 달리 크리깅은 이론적으로 추정하고자 하는 값 자체를 확률변수로 간주하여 모델링을 수행한다[1, 2]. 따라서 공간 보간과 더불어 샘플링되지 않은 지점에서의 불확실성 표현이 가능하며, 추계론적 시뮬레이션과 결합함으로써 공간 불확실성 추정이 가능하다[3].

전술한 결정론적 방법론과의 차별성과 더불어, 크리깅은 다양한 형태의 부가 자료를 통합할 수 있는 방법론을 제공한다. 만약 현장 조사 자료와 연관성이 높으면서 공간적으로 더 조밀하게 샘플링된 부가 자료의 이용이 가능하다면, 이 자료를 공간 보간 과정에 포함시킴으로써 샘플링되지 않은 지점의 불확실성을 낮출 가능성이 높아진다. 국가 GIS 사업 등으로 지형도, 지질도, 토양도, 임상도 등과 같은 수치 주제도의 이용이 가능해졌고, 다양한 해상도의 원격 탐사 자료의 이용이 가능하기 때문에 이러한 자료 통합은 기존 보간 방법론과 차별화된 크리깅의 특징으로 간주할 수 있다. 그동안 크리깅을 이용한 자료 통합에는 고도, 원격탐사 자료, 지구물리 탐사 자료와 같은 연속형 자료가 주로 이용되어 왔다[4, 9, 10, 12, 13, 14]. 이러한 기존 연구들에서는 연속형 자료를 통합함으로써 단변량 크리깅 기법에 비해 예측 능력이 향상되고, 크리깅의 일반적인 특징인 평활화 효과가 감소되었다고 보고하였다. 반면 범주형 주제도를 부가 자료로 사용하는 경우는 상대적으로 많이 연구되지 못하였는데, 범주형 주제도의 각 속성별로 포함되는 샘플 자료의 평균값을 대푯값으로 간주하여 단일 지역 평균 기반의 단순 크리깅(simple kriging with local means: SKLM)을 적용한 사례연구가 있다[5]. 즉 폴리곤으로 표현되는 범주형 자료의 개별 속성에 대푯값을 부여한 후에 단순 크리깅

적용 과정에 이용하였다. 그러나 폴리곤별로 하나의 대푯값으로 표현하기 때문에 실제로는 폴리곤 내부적으로 나타날 수 있는 속성값의 변화 양상을 표현하지 못하는 한계가 있다. 이러한 한계는 시군구와 같은 보다 큰 행정단위에서 단일 대푯값으로 인구수를 표현하더라도, 실제 집계구나 읍면동 단위별로는 인구 분포가 변하는 것과 유사하다. 또한 각 속성별 폴리곤에 일정한 대푯값을 부여함으로써 실제 통합된 결과에서 폴리곤 경계에서 불연속적인 속성값의 분포 양상이 나타날 수 있는 한계가 있다.

이 연구에서는 현장 조사 자료의 공간 보간을 위해 범주형 자료를 효과적으로 통합할 수 있는 다변량 크리깅 기법을 제안하고자 한다. 앞서 살펴본 것처럼, 기존 범주형 자료의 각 속성별로 단일 대푯값을 부여하는 방식을 보완하기 위해 범주형 주제도의 상세화(disaggregation)와 결합된 다변량 크리깅 기법을 제안하고자 한다. 우선 범주형 주제도의 각 속성별로 현장 조사 자료의 평균값을 계산한 후에, 이 값들에 대해 영역-점 변환 크리깅(area-to-point kriging, [7])을 적용하여 속성의 폴리곤 내부에서 값이 변하는 지역 평균을 유추하였다. 그리고 상세화시킨 범주형 주제도를 이용하는 가변적 지역 평균 기반의 단순 크리깅을 적용하여 최종적인 공간 분포도를 작성하였다. 태백산 광화대 일부 지역에서 획득한 지화학 자료와 지질도를 대상으로 사례 연구를 수행하여 제안 기법의 적용 과정을 예시하였다.

2. 분석 방법론

2.1 크리깅

크리깅은 일반화된 최소자승 회귀분석 방법론에 기반을 둔 공간 보간기법으로 알려져 있는데, 이론적으로 미지의 관측지점에서의 속성값은 그 위치에서 정의된 공간적으로 상관된 확률 변수의 조합으로 모델링된다는 확률 함수의 개념에 기반을 두고 있다[3]. 이 때, 공간적 상관성은 2차 정상성(stationary)의 전제하에 일반적으로 베리오그램을 통해 크리깅 시스템에 반영된다.

n 개의 샘플 자료 $\{z(\mathbf{u}_a), a=1, \dots, n\}$ 가 주어졌을 때, 샘플링 되지 않은 미지의 위치 \mathbf{u} 에서의 크리깅 예측치 $\hat{z}^*(\mathbf{u})$ 는 식 (1)과 같이 주변 관측값의 선형 가중합으로 표현된다.

$$z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) [z(\mathbf{u}_{\alpha}) - m(\mathbf{u}_{\alpha})] \quad (1)$$

여기서 $\lambda_{\alpha}(\mathbf{u})$ 는 관측값에 부여되는 가중치를 나타내며, $m(\mathbf{u})$ 와 $m(\mathbf{u}_{\alpha})$ 는 확률 변수 $Z(\mathbf{u})$ 와 $Z(\mathbf{u}_{\alpha})$ 의 평균 혹은 기댓값을 각각 나타낸다.

식 (1)에서 기댓값을 어떻게 정의하느냐에 따라 다양한 크리깅 기법으로 치환된다. 기댓값이 알려져 있으며 일정하다고 가정하면 단순 크리깅(simple kriging)으로 적용되며, 기댓값이 알려져 있지 않지만 주변 탐색창에서는 일정하다고 가정하면 정규 크리깅(ordinary kriging)으로 적용된다. 또한 기댓값이 공간적으로 변화한다고 가정하면 경향 크리깅(kriging with a trend model)으로 대체된다[3].

현장 조사 자료이외에 부가 자료가 존재할 경우, 부가 자료는 식 (1)의 평균 혹은 기댓값의 계산에 이용되거나 직접적으로 크리깅 시스템에 반영된다. 이러한 다변량 크리깅 기법으로는 지역 평균 기반의 단순 크리깅, 외부 추세 크리깅(kriging with an external drift), 공동 크리깅(cokriging) 등이 있다[2, 3]. 공동 크리깅은 부가 자료가 크리깅 시스템에 직접적으로 사용되는데, 원 자료의 베리오그램, 부가자료의 베리오그램 및 원 자료와 부가자료 사이의 교차 베리오그램을 계산하고 모델링해야 하는 번거로움이 존재한다. 외부 추세 크리깅은 부가 자료가 원 자료와 선형 관계가 존재할 때 적용가능하며, 부가 자료를 이용하여 개별 크리깅 대상 그리드에서 선형 회귀분석의 계수를 재추정한다. 공동 크리깅과 외부 추세 크리깅은 부가 자료로 범주형 자료에는 적용이 불가능하고 연속형 변수에만 적용이 가능하며, 변수의 개수가 늘어나면 이에 수반되는 모델링 과정이 복잡해진다[3, 4]. 반면 지역 평균 기반 단순 크리깅은 부가 자료를 현장 조사 자료에 대해 캘리브레이션한 값을 기댓값으로 사용하기 때문에 서로 다른 특성을 가진 다양한 자료의 이용이 가능한 장점이 있다. 또한 기존 연구 [4]에서는 고도자료를 이용하는 강우 자료의 공간 보간에 전술한 다변량 크리깅을 적용하여 성능 비교를 하였는데, 3가지 방식의 예측 능력이 거의 유사한 것으로 보고하였다. 그러나 실제 적용 과정의 간편성과 실용성 측면에서는 지역 평균 기반의 단순 크리깅이 가장 우수하다고 보고하였다. 이러한 배경에서 이 연구에서는 다변량 크리깅 기법으로 지역 평균 기반의 단순 크리깅을 적용하였으며, 크리깅 식은 (2)와 같다.

$$z_{sklm}^*(\mathbf{u}) = m^*(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) [z(\mathbf{u}_{\alpha}) - m^*(\mathbf{u}_{\alpha})] \quad (2)$$

우선 관측 자료와 그 지점에서의 기댓값의 차이인 잔차($z(\mathbf{u}_{\alpha}) - m^*(\mathbf{u}_{\alpha})$)를 계산하고, 이 잔차를 크리깅한 결과에 연구지역의 모든 위치에서 알고 있는 기댓값 $m^*(\mathbf{u})$ 을 더함으로써 추정치를 얻게 된다. 전술한 것처럼 지역 평균 기반의 단순 크리깅은 기댓값 $m^*(\mathbf{u})$ 을 계산하는 과정에서 연속형 변수와 범주형 변수의 사용이 모두 가능하다. 만약 연속형 변수를 사용할 경우, 현장 조사 자료와 연속형 변수의 선형/비선형 회귀분석을 적용한다. 반면 범주형 변수를 사용할 경우에는 범주형 속성에 포함되는 현장 조사 자료의 평균값을 기댓값으로 보통 사용한다[4].

2.2 가변적 지역 평균 기반의 단순 크리깅

이 연구에서는 식 (2)에서 $m^*(\mathbf{u})$ 을 추정하는 단계에서 범주형 속성별로 일정한 대푯값을 부여하지 않고, 속성별 폴리곤 내에서 속성값의 상세화를 위해 영역-점 변환 크리깅을 적용하였다.

Kyriakidis [7]에 의해 제안된 영역-점 변환 크리깅은 저해상도 자료로부터 상대적으로 고해상도의 그리드에서 속성값을 예측하는 크리깅을 의미한다. 즉 영역-점 변환 크리깅은 다양한 형태의 폴리곤 단위에서 대푯값을 가지고 있는 상태에서 폴리곤 단위를 이산화시킨 지점에서의 상세화된 값을 예측한다.

연구지역에 N 개의 폴리곤으로 구성된 범주형 주제도의 속성값 $\{z(v_i), \alpha=1, \dots, N\}$ 이 있을 경우, 폴리곤을 이산화시키는 그리드 \mathbf{u} 에서의 정규 크리깅 기반 영역-점 변환 크리깅 추정치 $z_p(\mathbf{u})$ 는 K 개의 주변 자료로부터 식 (3)을 이용하여 구하게 된다[7].

$$z_p(\mathbf{u}) = \sum_{i=1}^K \lambda_i(\mathbf{u}) z(v_i) \quad (3)$$

크리깅 가중치와 라그랑지안 상수 $\mu(\mathbf{u})$ 는 식 (4)의 크리깅 방정식을 이용하여 계산된다.

$$\sum_{j=1}^K \lambda_j(\mathbf{u}) C(v_i, v_j) + \mu(\mathbf{u}) = C(v_i, \mathbf{u}) \quad i=1, \dots, K$$

$$\sum_{j=1}^K \lambda_j(\mathbf{u}) = 1 \quad (4)$$

여기서 $C(v_i, v_j)$ 와 $C(v_i, \mathbf{u})$ 는 폴리곤 자료간의 공분산과 폴리곤 자료와 이산화된 그리드 사이의 공분산을 각각 나타낸다.

식 (3)과 (4)를 정리하면, 폴리곤 기반의 자료로부터 상세 그리드에서의 속성값은 기본적으로 이웃한

폴리곤 자료에 가중치를 부여하는 선형 결합 방식으로 결정하게 된다. 이를 위한 가중치를 결정하는 단계에서, 베리오그램으로부터 폴리곤 자료 사이의 공간적 자기상관성과 폴리곤 자료와 상세 그리드 지점 사이의 자기 상관성을 반영하게 된다.

영역-점 변환 크리깅은 일종의 제한조건으로 이산화된 그리드에서 얻어진 영역-점 변환 크리깅 추정치가 속하는 폴리곤 단위에서 평균값을 계산하면, 즉 업스케일링 혹은 합산(agggregation)하였을 때에 원래 폴리곤 자료의 대푯값을 재생산해야 한다. 예를 들러, 특정 폴리곤의 대푯값이 100이라면, 이 폴리곤 내부에 포함되는 영역-점 변환 크리깅 추정치들의 평균값을 계산하면 100이 나와야 한다는 것이다. 만약 영역-점 변환 크리깅에 사용된 K 개의 폴리곤 자료가 P_{β} 개의 이산화 포인트 \mathbf{u} 의 추정에 동일하게 사용된다면 이러한 재생산 조건을 만족하게 된다.

일정한 대푯값으로 표현되는 폴리곤을 상세화시켰을 때, 폴리곤 내부의 상세화시킨 값들의 평균값이 원래 폴리곤의 대푯값과 같아지는 조합은 무한히 많다. 영역-점 변환 크리깅은 이러한 조합을 찾는 과정에 폴리곤, 이산화 포인트 사이의 공간적 자기상관성을 고려하면서 추정 오차의 분산을 최소화시키는 결과를 얻게 된다.

이러한 제한조건을 만족하는 영역-점 변환 크리깅 추정치 $z_p(\mathbf{u})$ 를 얻게 되면, 이 값을 식 (2)의 $m^*(\mathbf{u})$ 로 대체하여 사용하면 최종 추정값을 얻게 된다.

영역-점 변환 크리깅을 가변적 지역 평균으로 사용하는 제안 방법론의 실제 적용 절차는 그림 1에 제시하였으며, 세부 과정은 다음과 같다.

- (1) 범주형 주제도의 각 속성 폴리곤별에 포함되는 현장 조사 자료의 평균값을 대푯값으로 할당
- (2) (1)의 자료를 대상으로 현장 조사 자료의 베리오그램 모델을 이용하여 영역-점 변환 크리깅을 적용하여 공간 보간의 대상 그리드별 상세화된 속성값을 유추하고 지역 평균값으로 간주
- (3) 현장 조사 자료에서의 속성값과 (2) 단계를 통해 얻어진 현장 조사 자료 위치에서의 속성값의 차이인 잔차를 계산
- (4) 잔차에 대한 베리오그램 모델을 이용하는 단순 크리깅을 적용하여 공간 보간 대상 그리드별 잔차값을 유추

- (5) (2)의 유추된 지역 평균값과 (4)에서 유추된 잔차값을 합하여 공간 보간 대상 그리드에서의 최종 추정값 계산

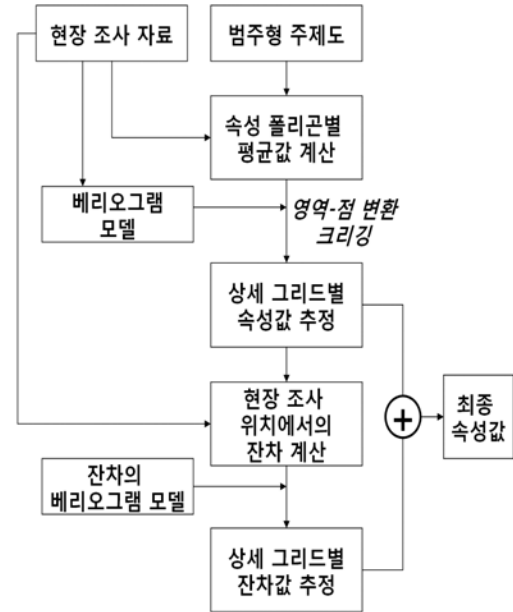


그림 1. 가변적 지역 평균 기반의 단순 크리깅의 실제 적용 절차

3. 사례연구

3.1 연구지역 및 사용 자료

이 연구에서는 제안한 다변량 크리깅 기법의 적용 가능성을 예시하기 위해 강원도 태백산 광화대 호명도폭([16])내의 177개 지점에서 획득된 하상 퇴적물의 구리(Cu) 성분 자료에 대한 사례연구를 수행하였다. 범주형 주제도로는 수치 지질도 강릉 도폭을 이용하였다.

그림 2는 지질 경계와 샘플 자료의 위치를 나타내고 있다. 대부분의 고 함량치는 대상 지역의 북서부와 중앙부에 나타나고 있으며, 도폭의 하단과 남동쪽에서는 상대적으로 소량이 나타나고 있다. 연구 지역에서 고 함량치가 나타는 지역은 소량의 화성암과 Cambro-Orodovician 변성 퇴적암으로 구성되어 있다고 보고되었는데[16], 이 지역은 지질도에서는 하부대석회암층군, 상부대석회암층군 및 양덕층군에 해당된다.

샘플 자료의 기본 통계값을 살펴보면(그림 3), 평

균은 130.9 ppm이지만 표준편차가 101.64 ppm으로 높게 나타나면서 양으로 치우친 분포를 나타내고 있다. 이러한 극단적인 양의 왜도 값을 고려하여 원 자료를 크리깅에 사용하지 않고, 극단값의 효과를 줄이기 위해 상용로그를 취한 값을 크리깅 분석에 사용하였고, 이후 역변환을 통해 원 자료 스케일로 변환하였다.

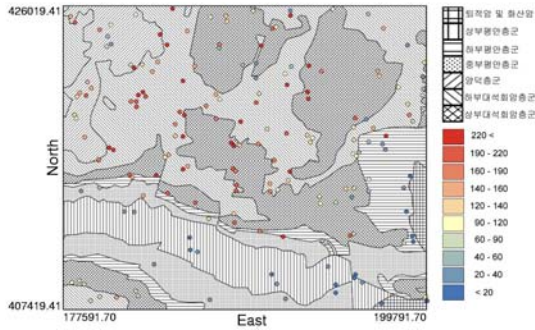


그림 2. 샘플 자료의 위치도와 지질 경계

로 북서쪽은 값이 높고 우측과 하부는 상대적으로 값이 낮음을 확인할 수 있다. 따라서 그림 3의 각 지질 속성별 평균값을 대푯값으로 사용할 경우, 이러한 세부적인 변화 양상을 설명할 수 없게 된다. 이 연구에서는 그림 4의 값을 영역-점 변환 크리깅에 이용하였으며, 비교를 목적으로 기존 지역 평균 기반의 단순 크리깅의 평균값으로 사용하였다.

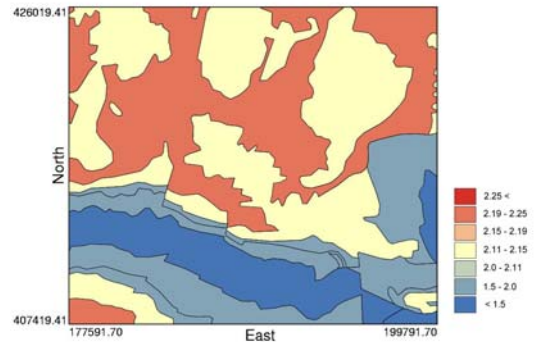


그림 4. 지질별 평균 로그 변환된 구리 분포도

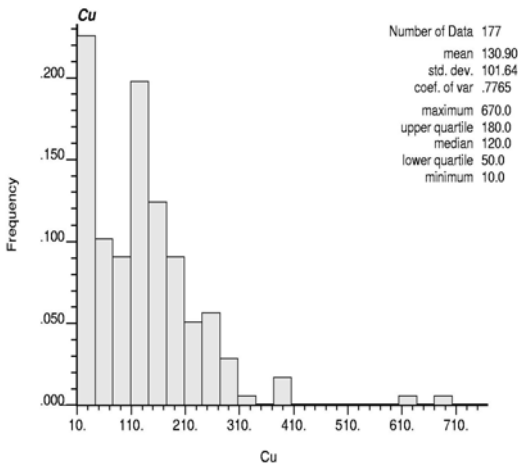


그림 3. 샘플 자료의 히스토그램과 요약통계치

3.2 분석 결과

우선 영역-점 변환 크리깅을 적용하기 위해, 각 지질별로 포함되는 샘플 자료들의 평균값을 계산하였는데, 그림 4는 각 지질별 로그 변환된 구리의 평균값의 공간 분포를 나타내고 있다.

그림 4를 살펴보면, 하부대석회암층에서 가장 높은 값을 나타내고 있으나, 이 속성에 포함되는 그림 2의 샘플 자료의 분포를 살펴보면 이 속성 내부적으로

그림 4의 자료에 대해 영역-점 변환 크리깅을 적용하기 위해 우선 로그 변환된 구리 자료의 실험적 베리오그램을 계산하고 모델링을 수행하였다(그림 5). 실험적 베리오그램은 대상 지역의 전체 크기를 고려하여 [6]에서 제시한 경험적 방법에 따라 대상 지역의 최대 범위의 1/2에 해당하는 12km까지 계산하였다. 실험적 베리오그램을 계산한 후에는 자료의 잡음 효과를 나타내는 너겟 효과, 공간적 자기상관성이 최대 유지되는 거리인 상관거리, 상관거리에서의 베리오그램 값인 문턱값, 이론적 베리오그램 함수 등과 같은 베리오그램 모델 파라미터를 결정해야 한다. 이 연구에서는 우선 가중 최소 자승법[8]에 따라 자동으로 초기 모델을 설정한 후에, 시각적인 판단으로 최종 베리오그램 모델 파라미터들을 선정하였다. 모델링 결과, 약 1.6km의 근거리 및 약 12km 정도의 원거리에서 상관거리가 나타났는데 이러한 양상은 2개의 구상 모델을 이용하여 모델링하였다. 이후 그림 5의 베리오그램 모델을 영역-점 변환 크리깅 변환에 필요한 점 기반 베리오그램 모델로 사용하였다. 영역-점 변환 크리깅에 필요한 폴리곤 사이의 공분산과 폴리곤과 그리드 점 사이의 공분산의 계산에는 폴리곤을 이산화시킨 두 지점 사이의 점 기반 공분산의 평균값을 이용하였다.

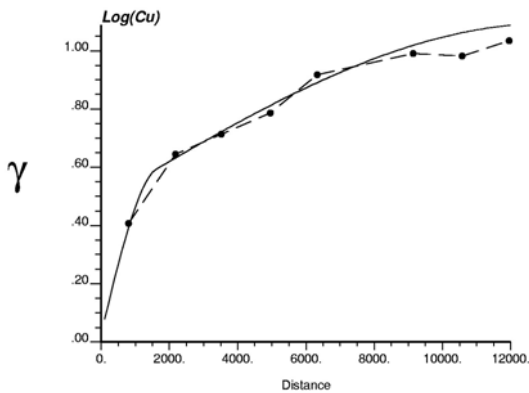


그림 5. 로그 변환된 구리 자료의 실험적 베리오그램과 모델링 결과

그림 6은 영역-점 변환 크리깅 적용 결과를 나타내고 있다. 영역-점 변환 크리깅을 통해 각 지질 속성을 나타내는 폴리곤 내부적으로 로그 변환된 구리 값이 변하는 양상을 얻을 수 있다. 연구지역의 중앙에 위치하는 상부대석회암층군은 그림 4의 일정한 평균값의 분포와 다르게 주변을 둘러싸고 있는 상대적으로 평균값이 가장 높았던 하부대석회암층군의 영향으로 값이 높게 나타나는 지역이 내부적으로 나타나고 있다. 또한 연구지역의 동쪽 하단에 위치하는 상부대석회암층군도 역시 폴리곤 내부에서 상대적으로 값이 높은 지역이 나타나고 있다. 즉 영역-점 변환 크리깅을 적용함으로써, 동일한 지질 속성 내부적으로 로그 변환된 구리값의 분포가 변하는 양상을 얻게 된다.

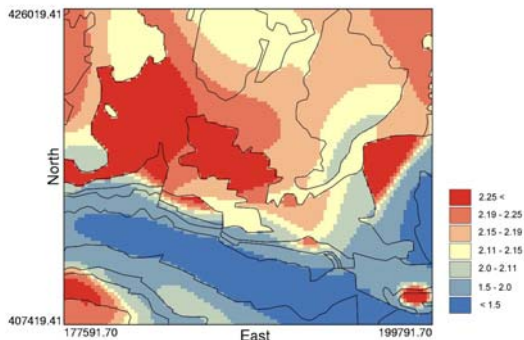


그림 6. 영역-점 변환 크리깅 적용 결과

영역-점 변환 크리깅의 결과가 재생산 조건을 만족하는지 살펴보기 위해, 그림 6의 200미터 해상도의 포인트 기반 로그 변환된 구리값들을 원 지질 속

성 폴리곤별로 평균값을 구한 결과와 그림 4의 원 지질 속성 폴리곤별 로그 변환된 구리값 사이의 산점도를 작성하였다(그림 7). 작성 결과, 폴리곤 내부의 국소적으로 속성값이 변하지만 원 폴리곤 자료의 스케일에서의 속성의 대푯값을 그대로 재생산하는 제한조건을 잘 만족하고 있음을 확인할 수 있다.

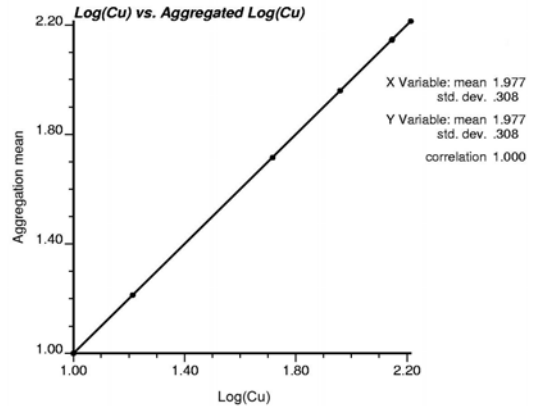


그림 7. 원 폴리곤 자료의 속성값과 영역-점 변환 크리깅 결과의 폴리곤별 업스케일링 결과의 산점도

지역 평균 기반의 단순 크리깅의 실제 적용에는 원 자료의 베리오그램을 사용하는 것이 아니라, 관측 지점에서의 값과 평균값의 차이인 잔차의 베리오그램을 사용해야 한다. 그래서 177개 샘플 위치에서의 그림 4의 평균값과 그림 6의 평균값을 이용하여 잔차를 각각 계산한 후에 베리오그램 모델링을 수행하였다. 그림 5와 마찬가지로, 자동 모델링 과정과 시각적 판독에 의한 수정 과정을 함께 베리오그램 모델링에 사용하였다. 모델링 결과는 그림 8에 제시하였는데, 2가지 잔차 모두에서 근거리와 원거리의 공간 상관성이 존재함을 확인할 수 있으며, 이러한 공간 상관성은 잔차의 크리깅 수행시 사용된다. 영역-점 기반 추정치로부터 계산된 잔차는 기존 영역별 일정값을 가정하여 계산된 잔차에 비해 상관거리가 상대적으로 길어지고 문턱값도 크게 나타나고 있다. 평균값으로 표현된 폴리곤 자료에 비해, 점 기반 자료는 속성값의 변화폭이 일반적으로 크기 때문에 자료의 변동성을 나타내는 문턱값이 증가하였으며, 지질별 경계를 넘어서는 변화를 설명할 수 있기 때문에 상대적으로 상관거리도 증가한 것으로 판단된다.

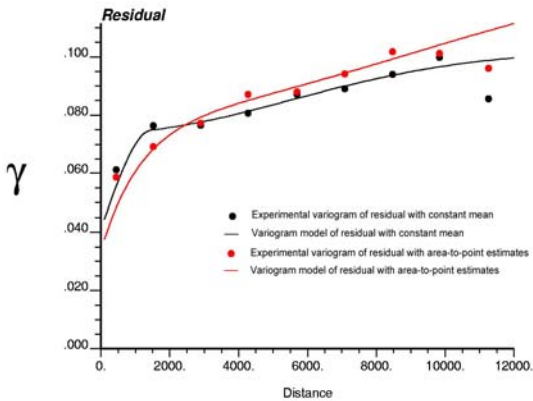


그림 8. 잔차의 실험적 베리오그램과 모델링 결과

영역-점 변환 추정치를 사용하는 가변적 지역 평균 기반의 단순 크리깅의 예측 능력을 정량적으로 판단하기 위해, 기존 단일 지역 평균 기반의 단순 크리깅과 비교하였다. 또한 부가 자료로 지질도를 사용하지 않는 단변량 보간 기법으로 같은 이론적 배경을 가지는 단변량 정규 크리깅을 비교 대상으로 하였다.

예측 능력의 검증에는 leave-one-out 기반 교차 검증을 적용하였다. 즉 각 샘플을 일부러 제거한 후에 나머지 샘플들만을 이용하여 제거한 샘플 지점의 값을 유추하는 과정을 모든 샘플 위치에서 반복하였다. 이러한 교차 검증을 통해 얻어진 추정값으로부터 추정 오차를 정량적으로 계산하기 위해 평균 오차(mean error)와 평균 절대 오차(mean absolute error: MAE)를 계산하였다. 평균 오차를 통해 추정값의 편향 정도를 비교할 수 있으며, 평균 절대 오차를 통해 오차의 크기를 비교할 수 있다. 또한 제안 기법의 기존 보간 기법에 대한 상대적 예측 능력의 향상도를 정량적으로 살펴보기 위해 상대적 향상도를 계산하였다.

교차 검증 결과는 표 1에 제시하였는데, 모든 경우의 평균 오차는 음수로 전반적으로 저추정 양상이 나타나고 있다. 그러나 이러한 양상은 그림 3에 나타난 원 자료가 양으로 치우친 분포로 인해 소수의 값이 아주 큰 샘플 자료의 값을 저추정한 양상이 평균 계산에 반영되어 음의 값을 나타낸 것으로 판단된다. 편향 정도는 단변량 보간 기법에 비해 지질도를 부가 자료로 이용하는 경우가 상대적으로 가장 작았으며, 제안 기법에 가장 우수하였다. 지질도를

표 1. 교차 검증 결과

보간 알고리즘	통계치	평균 오차	평균 절대 오차	정규 크리깅에 대한 평균 절대 오차의 향상도 (%)
정규 크리깅		-19.24	53.64	-
단일 지역 평균 기반의 단순 크리깅		-14.70	47.40	11.64
가변적 지역 평균 기반의 단순 크리깅		-12.62	40.20	25.06

부가 자료로 사용하는 경우와 그렇지 않은 경우의 평균 절대 오차를 비교한 결과, 평균 오차와 마찬가지로 다변량 기법이 단변량 기법에 비해 상대적으로 작은 평균 절대 오차를 나타내었다.

기존 단일 지역 평균 기반의 단순 크리깅은 정규 크리깅에 비해 상대적으로 11.64%의 예측 능력의 향상을 나타내었다. 특히 이 연구에서 제안한 가변적 지역 평균 기반의 단순 크리깅은 기존 정규 크리깅에 비해 약 25% 정도의 예측능력의 향상을 나타냈으며, 기존 단일 지역 평균 기반의 단순 크리깅에 비해 약 15.19%의 예측 능력의 향상을 나타내었다. 이러한 결과는 단변량 기법에 비해 부가 자료를 이용하는 다변량 기법의 예측 능력이 우수하며, 범주형 자료를 이용한 다변량 크리깅을 적용할 때 제안 기법이 유용하게 사용될 수 있음을 정량적으로 뒷받침한다.

그림 9는 제안 기법과 비교 기법들에 의해 작성된 구리 분포도를 나타내고 있다. 샘플 자료만을 사용한 정규 크리깅의 결과에서는 샘플 자료가 많이 존재하는 지역에서는 값의 변이가 크게 나타나지만, 연구지역 하단과 동쪽의 샘플 자료가 적은 지역에서는 평활화 효과가 두드러지는 양상을 보인다.

반면 지질도를 부가 자료로 이용한 경우에는 그림 4와 그림 6의 전반적인 양상과 샘플 자료의 영향이 혼재된 양상을 보인다. 즉 샘플 자료와 가까운 지점에서는 샘플 자료의 영향을 크게 받고, 샘플 자료로부터 멀어지거나 존재하지 않는 지점에서는 지역 평균의 양상을 반영하고 있다. 기존 단일 지역 평균 기반의 단순 크리깅의 결과에서는 전술한 샘플 자료의 영향이 함께 나타나고 있지만, 샘플링 밀도가 높지 않은 대상 지역의 하단과 동쪽 부근에서 지질 속

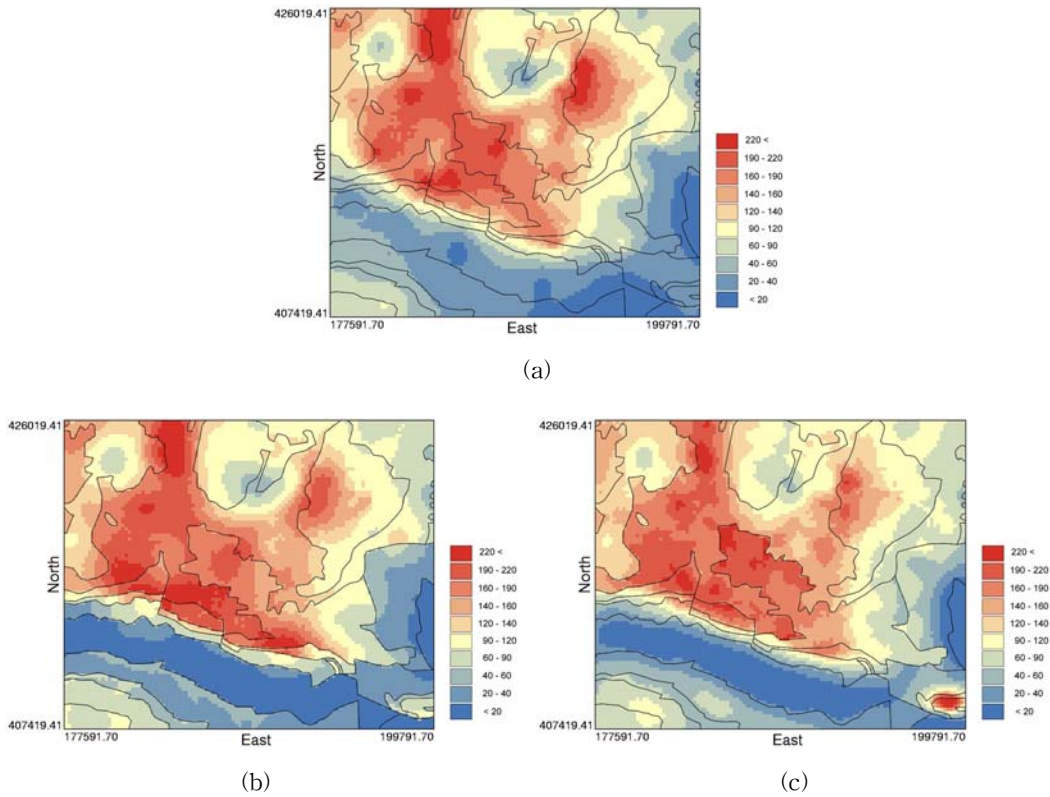


그림 9. (a) 정규 크리깅 적용 결과, (b) 단일 지역 평균 기반의 단순 크리깅 적용 결과, (c) 가변적 지역 평균 기반의 단순 크리깅 적용 결과

성별로 불연속적인 값의 분포가 나타나고 있다. 그러나 이 연구에서 제안한 가변적 지역 평균 기반의 단순 크리깅에서는 지질 속성 경계에서의 불연속적인 분포 양상이 많이 감소하여 보다 연속적인 구리 값의 분포 양상을 나타내고 있다. 실제 공간상에서 연속적으로 변하는 지화학 자료의 특성을 고려할 때, 제안 기법이 범주형 자료를 부가 자료로 사용할 때 성능면과 실제 분포 특성 자체에서도 우수함을 확인할 수 있었다.

4. 결론

공간적으로 산재된 현장 조사 자료와 연관성이 높은 부가 자료를 공간 보간 과정에 통합시킴으로써 보다 신뢰성 높은 주제도의 생성이 가능하다. 지질도와 같은 범주형 주제도는 지화학 자료의 모양 특성을 반영할 수 있기 때문에 공간적으로 소수의 위치에서 획득된 지화학 자료의 공간 보간에 효과적으

로 적용될 수 있다. 이 연구에서는 현장 조사 자료와 범주형 주제도의 통합을 위해 2단계로 구성된 다변량 크리깅 기법을 제안하였다. 우선 범주형 자료의 속성별로 부여되는 평균값에 영역-점 변환 크리깅을 적용하여 크리깅 대상 그리드별로 값이 상대적으로 변화도록 상세화를 수행하였다. 2단계에서는 이 자료를 가변적 지역 평균으로 사용하는 단순 크리깅을 적용하였다.

지화학 원소인 구리 자료의 공간 보간에 지질도를 통합하는 사례 연구를 수행한 결과, 현장 조사 자료만을 이용하는 단변량 크리깅에 비해 지질도를 고려하는 제안 기법이 약 25%의 예측 능력의 향상을 나타내었다. 또한 제안 기법이 속성별로 일정한 평균값을 부여하는 기존 방식에 비해 약 15%의 향상된 예측 능력을 나타내었으며, 지질 경계에서의 불연속적인 속성값의 분포가 완화된 결과를 얻을 수 있었다.

이 연구에서 부가 자료의 통합을 위해 적용한 영

역-점 변환 크리깅 기법은 이러한 목적 이외에 고해상도 자료 생성을 위한 다운스케일링에 유용하게 사용될 수 있다. 예를 들어 저해상도 원격탐사 자료의 다운스케일링과 행정구역과 같이 일정 영역에 걸쳐 대푯값으로 표현되는 인문 사회 자료의 상세화에 적용될 수 있을 것으로 판단된다. 이 연구에서는 2단계로 구성된 다변량 크리깅을 적용하였는데, 앞으로 이 과정들을 한 번에 수행하는 다변량 블록 크리깅 기법의 적용 연구를 수행할 예정이다.

참 고 문 헌

- [1] J.-P. Chiles and P. Delfiner, 1999, *Geostatistics: Modeling Spatial Uncertainty*, p. 720, Wiley-Interscience, New York.
- [2] C. V. Deutsch and A. G. Journel, 1998, *GSLIB: Geostatistical Software Library and User's Guide*, 2nd Ed., p. 369, Oxford University Press, New York.
- [3] P. Goovaerts, 1997, *Geostatistics for Natural Resources Evaluation*, p. 483, Oxford University Press, New York.
- [4] P. Goovaerts, 2000, "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall," *Journal of Hydrology*, vol. 228, no. 1-2, pp. 113-129.
- [5] P. Goovaerts and A. G. Journel, 1995, "Integrating soil map information in modelling the spatial variation of continuous soil properties," *European Journal of Soil Science*, vol. 46, no. 3, pp. 397-414.
- [6] A. G. Journel and C. J. Huijbregts, 1978, *Mining Geostatistics*, p. 600, Academic Press, New Jersey.
- [7] P. C. Kyriakidis, 2004, "A geostatistical framework for area-to-point spatial interpolation," *Geographical Analysis*, vol. 36, no. 3, pp. 259-289.
- [8] E. Pardo-Iguzquiza, 1999, "VARFIT: a Fortran-77 program for fitting variogram models by weighted least squares", *Computers & Geosciences*, vol. 25, no. 3, pp. 251-261.
- [9] N.-W. Park, D.-H. Jang and K.-H. Chi, 2009, "Integration of IKONOS imagery for geostatistical mapping of sediment grain size at Baramarae beach, Korea," *International Journal of Remote Sensing*, vol. 30, no. 21, pp. 5703-5724.
- [10] D. Patriarche, M. C. Castro and P. Goovaerts, 2005, "Estimating regional hydraulic conductivity fields - a comparative study of geostatistical methods," *Mathematical Geology*, vol. 37, no. 6, pp. 587-613.
- [11] 김두식, 원지혜, 김혜인, 김경희, 박관동, 2010, "기상자료 보간에 의한 GPS기반 가강수량 산출 정확도 분석," *한국공간정보학회지*, 제18권, 제4호, pp. 33-41.
- [12] 박노옥, 장동호, 2008, "수치표고모델과 다변량 크리깅을 이용한 기온 및 강수 분포도 작성," *대한지리학회지*, 제43권, 제6호, pp. 1002-1015.
- [13] 오석훈, 2005, "암반등급 해석을 위한 비선형 지시자 변환과 3차원 크리깅 기술의 물리탐사 및 시추 자료에 대한 적용," *한국지구과학회지*, 제26권, 제5호, pp. 429-435.
- [14] 정상용, 이강근, 1995, "난지도 매립지 일대의 지하수위 분포 추정을 위한 복합 크리깅의 응용," *지하수환경*, 제2권, 제2호, pp. 58-63.
- [15] 조홍래, 정종철, 2006, "강우자료에 대한 공간보간 기법의 적용," *한국GIS학회지*, 제14권, 제1호, pp. 29-41.
- [16] 진명식, 이진수, 김성재, 이취영, 1983, *태백산 광화대 호명 도폭 지화학도*, 한국동력자원연구소.

논문접수 : 2011.06.21

수정일 : 1차 2011.08.12 / 2차 2011.08.24

심사완료 : 2011.08.26



박 노 옥

2004년 서울대학교 박사(지구과학)
2000년~2008년 한국지질자원연구원
선임연구원
2008년~현재 인하대학교 지리정보공
학과 조교수

관심분야는 공간통계학 기반 시공간 자료 분석, 원격탐사 등