

지구통계학적 시뮬레이션을 이용한 지화학 자료의 공간통합에서의 불확실성 추정

박노욱*¹, 지광훈¹, 권병두²

¹한국지질자원연구원 지질자원정보센터

²서울대학교 지구과학교육과

요약

이 논문에서는 지구통계학적 시뮬레이션을 이용하여 자료 표현에서의 불확실성이 최종적인 공간통합에 미치는 영향을 정량적으로 분석하고자 하였다. 광물자원 탐사를 위한 공간통합 사례연구를 통해 시뮬레이션 결과에 따라 예측 능력의 차이가 나타남을 확인할 수 있었으며, 결론적으로 지구통계학적 시뮬레이션이 공간 자료의 불확실성 모델링에 효율적으로 이용될 수 있을 것으로 판단된다.

주요어 : 지구통계학, 시뮬레이션, 교차검증

1. 서론

대부분의 지구과학 자료는 공간적으로 분포하면서 고유의 속성 정보를 가지며, 이는 각종 지도나 수치 형태로 저장 및 관리된다. 그러나 조사시 비용, 시간 및 인력 등의 문제로 원하는 지역의 모든 부분을 전면적으로 조사하지 않고, 부분적인 샘플링을 수행하게 된다. 대부분의 분석은 특정 영역의 모든 위치에 대한 정보를 필요로 하기 때문에 이렇게 공간적으로 sparse하게 샘플링된 자료로부터 exhaustive한 자료를 얻기 위해 각종 내삽 등의 모델링을 수행하게 된다. 지화학 자료의 경우, 수계, 암석 혹은 토양 등으로부터 샘플링을 수행하게 되며, 항공 지구물리탐사의 경우 특정 간격을 두고 항공기가 비행하면서 얻어진 자료로부터 전 지역의 자료를 얻게 된다. 기존의 연구에서는 각종 내삽 기법을 이용하여 전 지역의 자료를 얻은 것과 같은 자료를 얻고 이를 분석에 이용하여 왔다. 그러나 샘플링된 자료로부터 내삽을 수행하는 과정에서 필연적으로 자료의 불확실성이 내포하게 되며 이러한 불확실성은 분석 결과에 영향을 미치게 된다. 또한 실세계를 수치 자료로 표현하는데 있어서 단순화하는 경향이 있게 되며, 종이 지도로부터 수치화

하는 과정에서 불확실성 혹은 오차가 수반되게 된다.

이러한 공간자료의 불확실성 분석과 관련하여 통계학 혹은 GIS 공간분석 분야와 연계된 연구가 최근 활발하게 이루어지고 있다. 이 연구에서는 sparse하게 샘플링된 자료로부터 exhaustive한 자료를 얻는 내삽과정에서의 불확실성 혹은 오차가 공간 자료 통합에 미치는 영향을 분석하기 위해 지구통계학적 시뮬레이션을 적용하였다. 광물자원 탐사를 목적으로 여러 지화학 자료를 통합하는 과정에서의 불확실성을 추정하는 사례연구를 통해 지구통계학적 시뮬레이션의 적용 가능성을 알아보고자 하였다.

2. 지구통계학

1960년대 이후에 광상의 위치를 추정하기 위해 개발된 지구통계학은 기존에 내삽 방법으로 알려진 크리깅으로 대표되어 왔다. 지구통계학과 통계학의 가장 큰 차이는 기본적으로 통계값을 추출하는데 있어서 공간적인 분포를 고려한다는 것이다. 자료의 변이를 나타내는 통계치로 분산 혹은 상관계수 등이 주로 사용되지만, 지구통계학에서는 거리와 방향을 고려한 배리오그램(variogram)을 기본적으로 이용한다. 1980년대 말까지는 내삽 방법으로 많이 활용되어 왔으나, 1990년대 들어 시뮬레이션 기법이 개발됨으로써 공간적 불확실성 분석에 많이 활용되고 있다(Goovaerts, 1997).

크리깅과 시뮬레이션 모두 배리오그램으로부터 얻어지는 공간적 상관관계를 기반으로 하지만, 계산과정과 최종 결과물에 대해서는 차이가 크다. 크리깅은 일종의 최소오차 회귀분석 방법으로 자료 근처에서는 원 자료값과 예측값 사이의 분산을 작게 하여 예측값이 원 자료값과 비슷하게 나타나도록 한다. 따라서 자료의 분포 편차가 큰 경우에 예측값은 그 분산을 현저히 감소시키는 방향으로 나타난다(Goovaerts, 1997). 결과적으로 전반적으로 국소적인 특성이 사라진 평활화된 결과를 얻게 되며, 원래 자료의 배리오그램이나 히스토그램을 재생산할 수 없다. 또한 추정오차로 제시되는 크리깅 분산은 자료의 배치 및 구성에 관한 정보를 제공하지만 예측과 관련된 불확실성의 척도가 될 수는 없다. 이에 반해 시뮬레이션은 크리깅에 기반을 두지만 국소적인 특성을 나타내면서 원래 자료의 값과 분포, 배리오그램 등의 통계치를 재생산할 수 있다(Fig. 1). 또한 몬테카를로 기법을 이용함으로써 동일한 시뮬레이션 조건하에서도 확률적으로 등가(equi-probable)인 다른 결과를 나타내므로 이들의 비교를 통해 공간적인 불확실성 분포를 추정할 수 있다(Goovaerts, 1997).

이 연구에서는 여러 시뮬레이션 알고리즘 중에서 순차적 가우시안 시뮬레이션

(Sequential Gaussian Simulation: SGS)방법을 적용하였다. 이 방법은 주어진 자료와 공분산 함수를 만족하면서 순차적으로 원하는 모든 그리드 위치에서의 값을 순차적으로 추정하는 기법으로 매단계에서 생성된 자료는 다음 단계에서는 알려진 자료값을 가정하게 된다. 이 방법에서는 주어진 자료가 가우시안 분포를 이룬다는 가정하에 연산을 수행하며 새로운 값을 유추해내는 과정에서 역시 가우시안 분포를 이용하게 된다.

SGS의 실제 수행단계는 다음과 같다.

- ① 주어진 샘플 자료에 대한 배리오그램을 계산한다.
- ② 배리오그램에 기반하여 주위 값들을 고려하여 크리깅을 수행한다.
- ③ 크리깅으로부터 구한 예측값과 분산을 이용하여 누적확률분포로부터 난수를 이용하여 하나의 값을 생성한다.
- ④ 생성된 값을 다음 단계의 연산에서는 주어진 값으로 가정하여 임의의 경로를 따라 ②와 ③의 과정을 반복하여 모든 그리드 위치에서의 값을 계산한다.

미리 정해놓은 시뮬레이션 회수 N(각기 다른 임의의 경로를 가짐)에 따라 시뮬레이션을 수행하면 각 결과는 원 자료의 값과 배리오그램과 같은 통계치를 그대로 재생산하게 된다. 또한 이러한 N개의 결과들은 확률적으로 등가인 서로 다른 결과를 나타내므로 이 결과들을 입력으로 하는 모델에 적용하면 모델의 결과값의 분포를 얻게 되며 이러한 분포를 통해 불확실성 분포를 추정할 수 있다.

3. 사례 연구

3.1 연구 지역 및 사용 자료

사례 연구 지역은 광물자원 탐사를 목적으로 지화학 탐사가 이루어진 태백산 지역인 옥동지역이다. 이 지역은 행정구역상 강원도 영월군 영월읍, 상동읍, 하동면, 충청북도 단양군 영춘면, 경상북도 영주군 단산면, 부석면 태화군 물야면에 달하는 지역이다. 연구지역에는 철, 납, 아연 등의 금속광상, 석회석, 장석 등의 비금속광상, 그리고 흑연을 수반하는 석탄광의 세 가지 광종이 등록되어 있다. 연구지역은 옥동광산을 중심으로 맥상으로 발달하는 철광상의 북서부지역, 이목화강암체를 중심으로 맥상 및 교대광상으로 발달하는 연, 아연 및 철광상의 북동부 지역 2개소가 가장 활발한 광화작용을 이루고 있음이 보고된 바 있다(서정률, 1985). 이 연구에서는 이목, 유전, 도화동, 옥동, 오양, 정양, 가리 등 7개를 고려하였다. 사용자료로는 옥동 도폭(1:50,000)내에 분포하는 200개의 하상퇴적

물 시료를 채취하여 얻어진 지화학자료(이진수 외, 1984) 중 Ag, Cu, Pb 및 Zn의 총 4개의 원소를 사용하였다.

3.2 분석 방법

이 연구에서는 우선 각 지화학 원소의 배리오그램 모델링을 수행한 후에 이로부터 순차적 가우시안 시뮬레이션을 수행하였다. 시범적으로 20번의 시뮬레이션을 수행하였다. 그리고 광물자원 탐사도 작성을 위해 앞에서 얻어진 지화학 자료와 기존 알려진 7개의 광상 위치와의 연관성을 고려한 공간 통합 기법으로 로지스틱 회귀분석을 이용하였다. 개별적인 시뮬레이션 기법의 영향을 정량적으로 알아보기 위해 교차 검증에 기반을 둔 예측비율곡선(Chung and Fabbri, 1999)를 작성하였다. 이 연구에서는 비교적 적은 수의 광상이 존재하기 때문에 one-leave-out 기반의 교차 검증을 수행하였다. 이 방법은 다음 순서에 따라 수행되게 된다.

- ① 7개의 광상 위치 중에서 한 개를 제외한 나머지 6개와 지화학 자료에 대해 로지스틱 회귀분석을 적용하여 공간 통합 결과를 생성한다.
- ② 제외된 한 개의 광상위치에서의 공간 통합 결과값을 저장한다.
- ③ ①과 ②의 과정을 7개의 광상 모두에 대해서 적용한다.
- ④ 각 광상 위치에서의 공간 통합 결과값으로부터 예측비율곡선을 작성한다.
- ⑤ ①부터 ④의 과정을 20개의 시뮬레이션 결과에 대해 반복한다.

Fig. 1은 Zn에 대한 정규 크리깅 결과와 2번의 시뮬레이션 결과를 비교한 것이다. 그림에서 확인할 수 있듯이, 크리깅 결과는 전반적으로 평활화된 양상을 나타내며, 시뮬레이션 결과는 원자료값을 보전하면서 자료값이 없는 곳에서는 예측값의 양상이 다소 다르게 나옴을 확인할 수 있다. 이러한 시뮬레이션 결과의 공간적 분포 양상의 차이는 공간 통합 결과에도 영향을 미치게 된다.

교차검증을 통해 얻어지게 되는 예측비율곡선의 차이를 보다 명확하게 나타내기 위해 상위 5%별로 20번의 시뮬레이션에 따른 예측비율곡선의 오차 곡선을 작성하였다(Fig. 2). 비교를 위해 정규 크리깅 결과를 통합하여 얻어진 예측비율곡선을 함께 도시하였다. 오차 곡선에서 오차 막대의 크기는 위 아래로 각각 표준편차를 나타낸다. 정규 크리깅 결과는 상위 15%까지는 평균 이하의 값을 나타내다가 그 이상의 범위부터는 평균 이상의 값을 나타내고 있다. 이러한 결과는 기존 광상의 위치와 유사한 값을 나타내는 지역에서

평활화가 더 크게 나타나고, 다른 지역에서는 분산을 최소화 하는 과정에서 값이 커진것을 나타내는데, 이는 샘플링 위치와 연관이 있다. 앞에서 살펴본 바와 같이 부존 가능성이 높은 상위 15% 정도까지 오차 막대의 크기가 큼을 확인할 수 있다. 만약 입력자료 표현에 의한 불확실성 영향이 적다면, 오차 막대의 크기는 작아야 한다. 그러나 이렇게 오차 막대의 크기가 큰 것은 지화학 자료만을 이용한 공간통합의 결과에서 탐사 가능성이 높다고 생각할 수 있는 상위 부분의 불확실성이 큼을 나타내어 탐사의 위험도가 큼을 간접적으로 나타낸다.

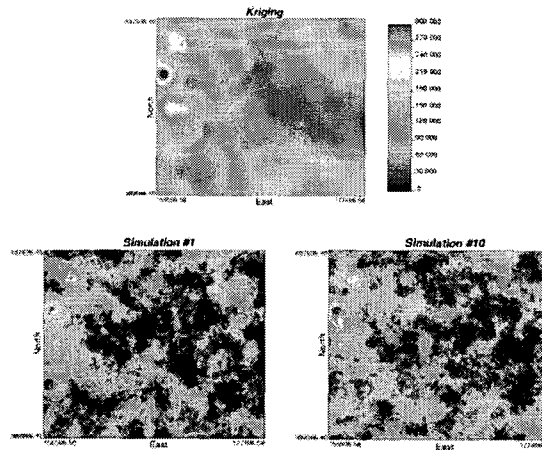


Fig. 1. Zn의 크리깅과 시뮬레이션 적용 결과.

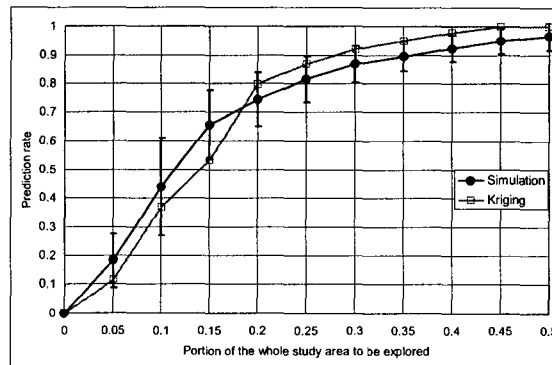


Fig. 2. 시뮬레이션에 따른 예측비율곡선의 오차 표시.

3. 결론

이 연구에서는 지구통계학적 시뮬레이션을 이용하여 지화학 자료의 내삽과정에서의 불확실성이 광물자원 탐사를 위한 공간통합에 미치는 영향을 알아보려고 하였다. 옥동 지역의 사례 연구를 수행한 결과, 시뮬레이션 결과에 따라 공간통합의 양상이 다르게 나타남을 확인할 수 있다. 이번 연구를 통해 지구통계학적 시뮬레이션은 공간 자료의 불확실성을 모델링하는데 효율적으로 이용될 수 있음을 확인하였으며, 앞으로 원격탐사 자료나 다른 공간자료의 불확실성 모델링에도 이용이 가능할 것으로 판단된다. 추후 연구로 다른 공간통합 방법의 적용, 다른 시뮬레이션 기법의 적용, 항공지구물리탐사 자료와의 통합 등의 비교 연구를 수행할 예정이다.

4. 참고문헌

- 서정률, 1985, 옥동도폭내의 광화작용에 대한 지질·광상학적 연구, 금속광상 조사연구 85-15, 한국동력자원연구소, p.75-121.
- 이진수, 홍영국, 김성재, 윤현수, 진만식, 이취영, 1984, 한국 지화학도 : 옥동도폭 (1:50,000), 한국동력자원연구소.
- Goovaerts, P., 1997, Geostatistics for natural resources evaluation, Oxford University Press.