

라돈 농도 분포도 작성을 위한 단변량 크리깅 기법의 정량적 비교*

곽근호¹ · 김용재² · 장병욱² · 박노욱^{1*}

Quantitative Comparison of Univariate Kriging Algorithms for Radon Concentration Mapping*

Geun-Ho KWAK¹ · Yong-Jae KIM² · Byung-Uck CHANG² · No-Wook PARK^{1*}

요 약

토양, 암석, 지하수로부터 실내에 유입되는 라돈은 인간에게 큰 위험을 끼치는 방사능 가스이다. 라돈 가스의 위해성을 확인하기 위해 실내 라돈 농도를 측정해 오고 있는데, 추가적인 분석 수행을 위해서는 신뢰성 높은 분포도 작성이 매우 중요하다. 본 연구에서는 비대칭 분포를 나타내는 라돈 농도의 공간 분포도 작성을 위해 단변량 크리깅 기법들의 비교를 목적으로 정규 크리깅, 비선형 자료 변환 기반의 로그 정규 크리깅, 다중 가우시안 크리깅과 지시자 크리깅의 예측 능력을 비교하였다. 예측 능력을 비교 분석하기 위해 잭나이프 방법을 이용하여 검증을 수행하였으며, 자료 구간별 오차와 샘플링 밀도의 차이에 따른 오차도 추가적으로 분석하였다. 남한 지역을 대상으로 한 사례 연구 결과에서 전반적으로 정규 크리깅에 비해 비선형 자료 변환 기반 크리깅 기법들이 좋은 예측 능력을 보였으며, 비선형 자료 변환 기반 크리깅은 로그 정규 크리깅, 다중 가우시안 크리깅 순으로 좋게 나타났다. 그러나 공간 패턴과 높은 값의 재생산을 고려할 때, 높은 값의 예측 능력은 정규 크리깅이 가장 우수하였다. 본 연구의 결과는 비대칭 분포 자료의 공간 예측을 위한 크리깅 기법의 선정에 유용하게 사용될 것으로 기대된다.

주요어 : 실내 라돈 농도, 공간 예측, 크리깅, 비대칭 분포 자료

ABSTRACT

Radon, which enters the interior environment from soil, rocks, and groundwater, is a radioactive gas that poses a serious risk to humans. Indoor radon concentrations are measured to investigate the risk of radon gas exposure and reliable radon concentration mapping is then performed for further analysis. In this study, we compared the predictive performance of various univariate kriging algorithms, including ordinary kriging and three

2017년 1월 6일 접수 Received on January 6, 2017 / 2017년 2월 10일 수정 Revised on February 10, 2017 / 2017년 2월 15일 심사완료 Accepted on February 15, 2017

* 본 연구는 원자력안전위원회의 재원으로 한국원자력안전재단의 지원을 받아 수행한 원자력안전연구사업의 연구결과임(No. 1305004).
1 인하대학교 공간정보공학과 Department of Geoinformatic Engineering, Inha University

2 한국원자력안전기술원 생활방사선안전실 Department of Natural Radiation Safety, Korea Institute of Nuclear Safety

* Corresponding Author E-mail : nwpark@inha.ac.kr

nonlinear transform-based kriging algorithms (log-normal, multi-Gaussian, and indicator kriging), for mapping radon concentrations with an asymmetric distribution. To compare and analyze the predictive performance, we carried out jackknife-based validation and analyzed the errors according to the differences in the data intervals and sampling densities. From a case study in South Korea, the overall nonlinear transform-based kriging algorithms showed better predictive performance than ordinary kriging. Among the nonlinear transform-based kriging algorithms, log-normal kriging had the best performance, followed by multi-Gaussian kriging. Ordinary kriging was the best for predicting high values within the spatial pattern. The results from this study are expected to be useful in the selection of kriging algorithms for the spatial prediction of data with an asymmetric distribution.

KEYWORDS : Indoor Radon Concentration, Spatial Prediction, Kriging, Asymmetric Distribution Data

서 론

라돈(^{222}Rn)은 라듐(^{226}Ra)의 딸 핵종으로 자연계에 널리 존재하는 비활성 가스로 인간에게 큰 위해를 끼치는 방사능 물질 중 하나이다 (Baek *et al.*, 2008; Smethurst *et al.*, 2008). 라돈 가스는 토양, 암석으로부터 발생하며 건축물 콘크리트 벽의 갈라진 틈이나 창틈, 배관 등을 통해 실내로 유입된다. 실내로 유입된 라돈 가스는 인간의 폐 조직을 손상시키며, 장기간 노출이 될 경우에는 폐암을 유발할 수도 있다. EPA(1992)에 따르면 라돈에 의한 미국 내 연간 폐암 사망자수는 흡연에 이어 두 번째로 높은 것으로 알려져 있다. 국내외에서는 이러한 라돈 가스의 위해성을 확인하기 위해 현장 조사를 통해 토양 가스 내 라돈 농도와 실내 라돈 농도 등을 측정하고 있다(Dubois, 2005; Green *et al.*, 2009; Kim *et al.*, 2011). 실내 라돈 농도 조사는 일반적으로 가옥이나 학교 등에서 수행되는데, 시간, 비용 등의 한계로 인해 전수 조사가 불가능하므로 일부 지역에서만 획득된다. 실내 라돈 농도는 피복선량 평가나 라돈 권고치 설정 등의 다양한 분석 수행을 목적으로 공간 분포도 제작에 활용될 수 있다. 실내 라돈 농도는 건축 자재나 실내 환기 여부 등과 같은 실내 요인보다 건물 아래에 존재하는 토양 가스에 의해 더 큰 영향을 받는 것으

로 보고되고 있어서(Cha, 2001), 미 획득 지점에 대한 라돈 농도 추정을 위한 공간 예측 기법의 적용이 가능하다.

대상지역에서 소수의 지점에서 얻어진 자료를 이용한 공간 예측 기법으로 거리 역비례 가중치법(Inverse Distance Weighting, IDW), Spline 등과 고급 예측 기법으로 알려진 크리깅(Kriging) 등이 주로 적용되어 왔다(Lee, 2010; Park *et al.*, 2012; Park *et al.*, 2013; Harman *et al.*, 2016). 라돈 농도 분포도 작성과 관련하여 국내에서는 Baek *et al.*(2008)이 서울시의 지하철 역사 내 라돈 농도 자료를 이용한 서울시 라돈 농도 분포 예측에 IDW와 정규 크리깅 기법의 예측 능력을 비교하였는데, IDW에서 전반적으로 우수하지만 자료의 샘플링 밀도가 높을수록 정규 크리깅의 우수하다고 보고하였다. 그리고 외국에서는 실내 라돈 농도 측정 자료를 대상으로 다양한 공간 예측 기법을 적용하여 라돈 농도 분포도를 작성해 왔다(Baek *et al.*, 2008; Kim *et al.*, 2011; Drolet *et al.*, 2013; Bossew *et al.*, 2014; Pásztor *et al.*, 2016).

공간 예측 기법 중에서 크리깅은 표본 자료의 공간 자기상관성 정보를 이용하여 주변값들에 가중치를 부여하고, 미 획득 지점에서의 값을 예측한다(Goovaerts, 1997). 가중치를 부여한다는 점에서 IDW와 동일하나, 거리만을 고려한 가중치 부여가 아닌 거리에 따른 공간 자기상관성을 가중치 결정에 직접 사용하며 베리

오그램 모델링을 통해 이방성(anisotropy) 구조를 공간 예측에 반영할 수 있다는 장점을 가지고 있다(Goovaerts, 1997). 이러한 크리깅 기법은 하나의 변수만을 이용하는 단변량 크리깅 기법과 원 자료와 높은 상관성을 나타내는 부가 자료를 이용하는 다변량 크리깅 기법이 있다. 이 중 단변량 크리깅 기법에는 단순 크리깅(simple kriging, SK), 정규 크리깅(ordinary kriging, OK), 일반 크리깅(universal kriging, UK) 등과 원 자료값에 대해 비선형 변환하여 크리깅을 수행하는 로그 정규 크리깅(lognormal kriging, LK), 다중 가우시안 크리깅(multi-Gaussian kriging, MG), 지시자 크리깅(indicator kriging, IK) 등이 있다(Goovaerts, 1997; Deutsch and Journel, 1998).

크리깅은 기법에 따라 예측 결과가 상이하기 때문에 공간 예측의 목적에 맞게 선택해야 한다. 특히, 크리깅을 적용하고자 하는 원 자료의 분포 특성의 고려가 중요하다. 보통 현장 조사를 통해 얻어진 자료는 큰 이상값을 포함하는 양이나 음으로 치우친 비대칭 분포를 나타내는 경우가 많다(Park, 2009). 이에 따라, 측정 오류에 의한 큰 이상값들은 자료의 95%나 98%만 사용하는 방법으로 해결하거나, 큰 이상값들이 중요한 값으로 여겨질 경우에는 일반적으로 비선형 자료 변환을 통해 정규 분포를 따르게 한다(Webster and Oliver, 2007). Saito and Goovaerts(2000)는 다이옥신 오염 지역에서 획득된 비대칭 분포의 자료에 대해 비선형 자료 변환 기반 크리깅 기법을 적용한 후 예측 능력을 비교하였는데, 로그 정규 크리깅과 지시자 크리깅의 예측 능력이 우수하다고 보고하였다. Buttafuoco *et al.*(2007)은 역시 토양 라돈 가스 분포도 제작을 위해 비선형 자료 변환 기반의 크리깅 기법의 예측 능력을 비교하였는데, 다중 가우시안 크리깅이 가장 우수하다고 보고하였다. 그러나, 공간 예측의 정확도는 샘플링 밀도, 표본 자료의 공간 위치 및 예측 방법 등의 다양한 요소에 의존하기 때문에 단일 사례 연구만으로 일반적인 결론을 제시하기 어렵다. 현재까지 라돈 농도 분포도 작성에는 단변량 크

리깅 기법으로 로그 정규 크리깅이 주로 적용되어 왔는데, 비선형 자료 변환 기반 크리깅 기법인 지시자 크리깅이나 다중 가우시안 크리깅을 함께 비교한 연구는 매우 드문 상황이다.

본 연구에서는 현장 조사를 통해 획득된 실내 라돈 측정 자료를 이용한 라돈 농도 분포도 작성을 위해 일반적인 정규 크리깅 기법과 양으로 치우친 비대칭 분포 특성을 고려한 자료 변환 기반 크리깅 기법들의 예측 능력을 비교하고자 하였다. 이를 위해 남한 전체 지역 내에서 획득된 실내 라돈 농도 자료를 이용한 라돈 농도 분포도 작성 사례 연구를 수행하였다. 크리깅 기법의 예측 능력 비교를 위해 라돈 농도 분포도 작성 결과를 기반으로 공간적인 패턴을 살펴보았으며, 잭나이프 방법을 이용하여 크리깅 기법별로 예측 결과의 정량적 검증을 수행하였다. 또한, 자료의 분포 특성을 고려한 구간별 예측 오차와 자료의 샘플링 밀도에 따른 예측 오차를 추가로 분석하였다.

사용 자료와 연구 방법

1. 연구 대상지역 및 사용 자료

본 연구에서는 단변량 크리깅 기법들의 비교를 위해 한국원자력안전기술원에 의해 1999년부터 2009년까지 총 4번에 걸쳐 가옥과 공공 시설 등 5553개 지점에서 측정된 실내 라돈 농도 자료(Kim *et al.*, 2011)를 이용하여 사례 연구를 수행하였다. 이 중 동일 지점에서 2회 이상 측정된 자료는 평균으로 그 위치 값을 대체하였고, 제주도에서 측정된 자료를 제외하여 총 3899개 지점에서 획득된 자료를 이용하였다. 그림 1에서 자료 조사 위치를 살펴보면, 남한 전역에 걸쳐 대부분의 낮은 값들과 드물게 분포하는 높은 값들이 존재하고 있다. 또한, 전체 표본 자료의 요약 통계값을 살펴보면, 실내 라돈 농도 자료는 일부 큰 이상값의 영향 때문에 62.58Bq/m³의 평균과 67.82Bq/m³의 표준편차가 거의 유사하게 나타났다. 또한, 7.86의 왜도(skewness)값과 평균 보다 낮은 46.25Bq/

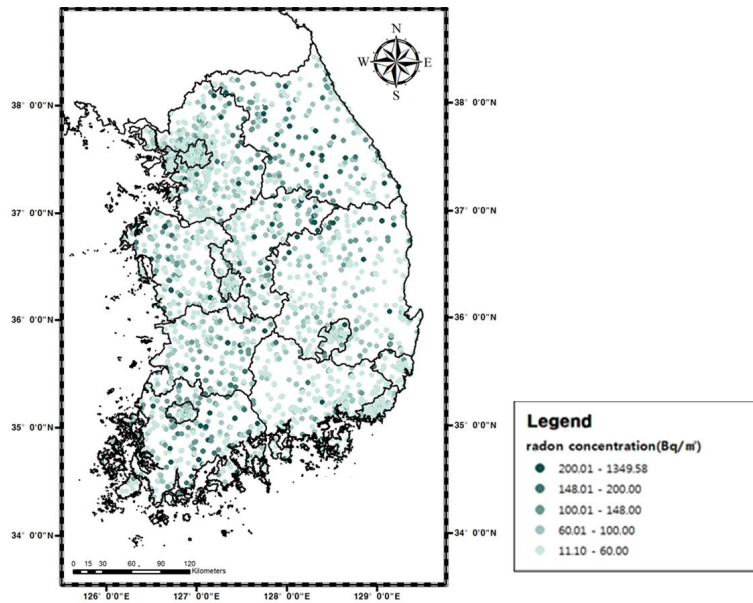


FIGURE 1. Locations of sample indoor radon survey data in the South Korea

m³의 중앙값을 통해 실내 라돈 농도 조사 자료에 큰 이상값들이 분포하는 것을 확인할 수 있다(그림 2(a)). 본 연구에서는 연구지역 내에서 추가적인 검증 자료가 존재하지 않아 전체 3899개의 조사 지점을 각각 50%에 해당하는 1950개 지점의 훈련 자료와 1949개 지점의 검증 자료인 2개의 집단으로 분류하였다. 그러나, 각 집단의 임의의 분류는 각각 다른 분포 구간에서 추출될 수 있으며, 이러한 경우 전체 자료의 특성을 반영하지 못하는 경우가 있다. 이에 따라, 자료를 동일한 분포로 획득하기 위해 백분위수 별로 10개의 구간으로 나누어 훈련 자료와 검증 자료를 각각 분류하였다. 훈련 자료와 검증 자료는 평균, 표준편차, 왜도와 사분위수에서 전체 표본 자료와 값이 유사하게 나타나 전체 표본 자료의 특성을 잘 반영하는 것을 확인하였다(그림 2(b), 그림 2(c)).

2. 단변량 크리깅 기법

지구통계학을 대표하는 크리깅은 관측값들의

선형 조합으로 미 획득 지점에서의 속성값을 예측하는 방법론이다(Goovaerts, 1997). 특정 지점에서의 관측값들은 확률 분포로 존재하고, 확률 분포 상에서 임의의 값이 선택되어지는 확률 함수의 개념을 기반으로 두고 있다. 이때, 확률 함수는 위치 이동에 따라 변하지 않는다는 정상성의 결정을 따르며, 거리만의 함수로 표현된다(Deutsch and Journel, 1998). 지구통계학에서는 이러한 정상성의 결정에 따라 공간 변동성의 정량적 지수로 베리오그램을 이용한다.

본 연구에서는 단변량 크리깅 기법 중에서 대표적인 정규 크리깅과 비선형 자료 변환 기반의 로그 단순 크리깅, 다중 가우시안 크리깅, 지시자 크리깅을 상호 비교하였다. 여기서는 각 크리깅 기법의 일반적인 특성만을 기술하였으며, 자세한 이론적 배경은 Goovaerts(1997)와 Webster and Oliver(2007)를 참조할 수 있다.

1) 정규 크리깅(OK)

정규 크리깅은 지구통계학에서 가장 대표적으로 사용되는 공간 예측 기법으로, 지역 평균

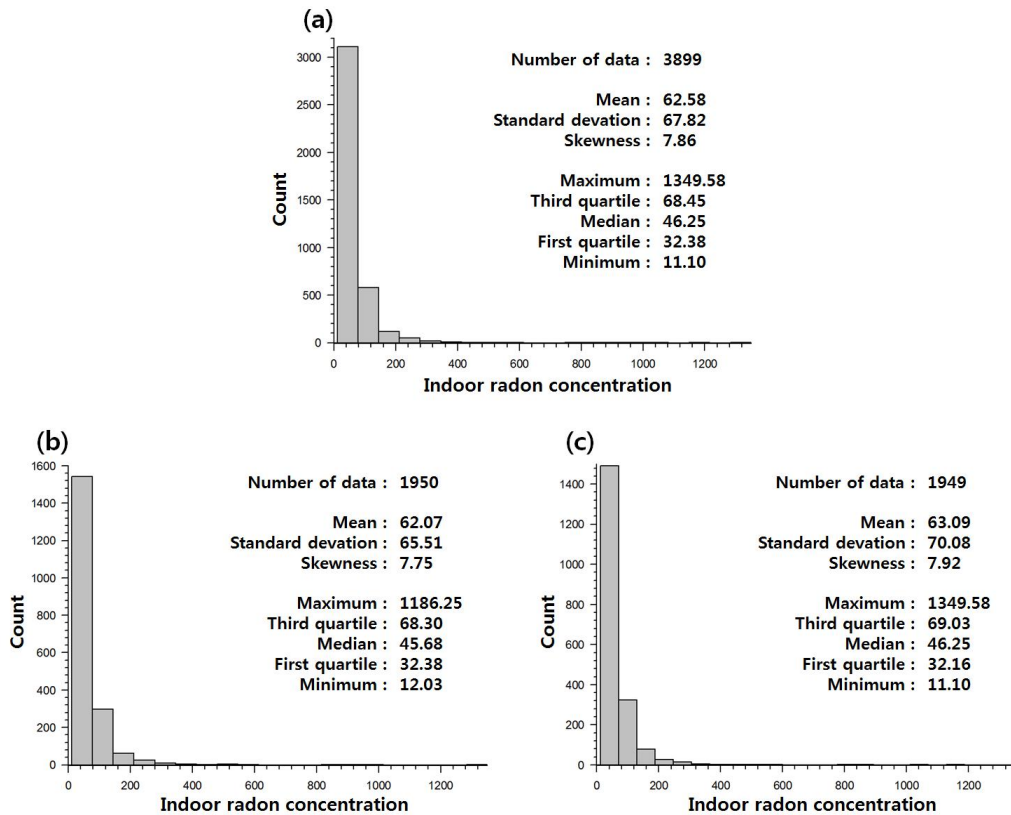


FIGURE 2. Histograms and statistics of sample indoor radon survey data.
 (a)all sample data, (b)estimation set, (c)validation set

이 알려져 있지 않으나 탐색 창 내에서는 일정한 상수로 간주한다.

2) 로그 정규 크리깅(LK)

정규 크리깅을 적용할 때 원 자료의 분포가 정규 분포를 따르지 않는다면, 큰 이상값들은 베리오그램의 불안정성을 가져오기 때문에 예측값에 큰 영향을 미치게 된다(Buttafuoco *et al.*, 2007). 로그 변환은 정규 분포를 따르지 않고, 큰 이상값을 포함하는 원자료의 분포를 대칭 형태로 변환시켜 이러한 왜곡을 제거하는 방법으로 많이 적용되어왔다. 로그 정규 크리깅은 로그 변환된 공간에서의 베리오그램 모델과 단순 크리깅 식에 따라 계산되고, 이를 통해 예측값과 크리깅 분산을 계산한다. 그러나, 로그

변환된 공간상에서 추정된 예측값을 원 자료 공간으로 역변환하기 위해 지수 함수를 적용할 경우 편향된 결과를 나타내는 것으로 알려져 있다(Deutsch and Journel, 1998). 원 자료의 공간에서 비편향된 추정값을 얻기 위해, 로그 변환된 공간상에서의 예측값과 분산값의 합에 지수 함수를 적용하는 방법으로 최종 로그 정규 크리깅 결과를 얻게 된다.

3) 다중 가우시안 크리깅(MG)

이상값을 가진 원 자료에 로그 변환을 적용하더라도 완벽하게 정규 분포를 따르지 않는 경우가 있다. 다중 가우시안 크리깅은 원 자료에 정규 점수 변환(normal score transform)을 적용하여 완벽한 정규 분포를 따르는 자료로 변환한

후에 크리깅을 적용하는 방법이다. 정규 점수 변환은 원 자료를 크기순으로 재배열하고, 이를 통해 얻어지는 누적 분포 함수를 평균 0, 분산 1인 표준 정규 분포에 대응시키는 것을 말한다 (Goovaerts, 1997). 다중 가우시안 크리깅은 정규 점수 변환된 자료를 대상으로 단순 크리깅을 적용하며, 정규 점수로 변환된 관측값의 베리오그램 모델로부터 가중치를 계산하고 관측값의 선형 결합을 통해 정규 분포 공간에서의 예측값을 얻게 된다. 이 결과들을 대상으로 역 정규 점수 변환(normal score back transform)을 적용하여 원 자료 공간에서의 최종 예측값을 얻는다.

4) 지시자 크리깅(IK)

지시자 크리깅은 원 자료에 대해 지시자 변환을 수행한 후에 크리깅을 수행하여 특정 임계치보다 작거나 같게 될 조건부 확률을 계산하여 특정 위치에서의 조건부 누적 확률 분포 함수(conditional cumulative distribution function, ccdf)를 모델링하는 방법론이다(Deutsch and Journel, 1998). 지시자 변환은 연속형 변수의 원 자료값을 여러 개의 특정 임계치에 대해 임계치보다 작거나 같으면 1, 크면 0의 이진값을 가지도록 하는 방법이다. 지시자 변환된 변수들을 대상으로 베리오그램 모델링을 수행한 후에 크리깅을 적용하면, 일련의 임계치에 대한 조건부 확률을 얻게 된다. 여러 개의 임계치들에 대해 크리깅이 모두 수행되면, 이들의 조합을 통해 특정 위치에서의 ccdf를 얻을 수 있다 (Goovaerts, 1997).

지시자 크리깅의 적용 목적은 일반적으로 확률론적 불확실성을 나타낼 수 있는 ccdf 모델링에 있지만, ccdf로부터 계산되는 통계값을 크리깅 추정값으로 사용할 수 있다. 본 연구에서는 전체 자료를 100개로 이산화 시킨 후 얻어지는 분위수의 합으로 추정값을 계산하였다 (Goovaerts, 2009).

3. 예측 능력의 정략적 비교 방법

이미 자료값이 알려진 표본 자료를 이용하여

예측을 수행하는 크리깅 기법은 별도의 검증 자료가 존재하지 않는다면 예측 능력을 검증할 수 없다. 본 연구에서는 각 크리깅 기법별로 예측 능력을 비교하기 위해 잭나이프 방법을 이용하였다. 이 방법은 독립적인 2개의 집단인 훈련 자료와 검증 자료의 집단으로 나누고, 훈련 자료를 이용하여 검증 자료 위치에서의 예측값을 얻은 후 검증 자료의 원 참값과 비교하여 추정 오차를 계산한다.

본 연구에서는 잭나이프 방법의 적용을 통해 얻어지는 추정 오차를 각 크리깅 기법별로 비교하기 위해 평균 오차(mean error, ME), 평균 절대 오차(mean absolute error, MAE)와 평균 제곱근 오차(root mean square error, RMSE)를 계산하였다. 평균 오차는 음 또는 양의 방향으로 나타나는 크리깅 결과값의 편향 정도이며, 평균 절대 오차와 평균 제곱근 오차는 예측값의 오차의 크기를 판단하는데 평균 제곱근 오차는 원자료 규모에서의 오차의 크기를 나타낸다. 또한, 정규 크리깅의 오차값을 기준으로 다른 크리깅 기법들의 상대적인 오차값을 나타내는 상대적 평균 절대 오차(relative mean absolute error, RMAE)와 상대적 평균 제곱근 오차(relative root mean square error, RRMSE)를 계산하여 크리깅 기법별로 상대적인 오차의 크기를 비교하였다.

결과 및 고찰

1. 비선형 자료 변환 및 베리오그램 모델링 결과

각 크리깅 기법별로 크리깅 가중치 결정을 위해 원 자료, 로그 변환된 자료, 정규 점수 변환된 자료와 지시자 변환된 자료에 대해 실험적 베리오그램을 계산하고, 베리오그램 모델링을 수행하였다. 이 때, 지시자 변환된 자료는 임계치가 낮은 경우에 실내 라돈 농도 값이 큰 이상값들에 의해 가까운 샘플 지점 간의 상관성이 더 떨어지는 것으로 나타나, 백분위수에서 40% 이상에 해당하는 6개(40%, 50%, ..., 90%)를 임계치로 설정하였다. 또한, 베리오그

램 모델링은 모든 크리깅 기법에서 이방성이 나타나지 않아 등방성으로 수행하였으며, 각 크리깅 기법별 베리오그램 모델링 수행 결과는 그림 3과 그림 4에 제시하였다.

베리오그램은 모든 자료값이 값 계산에 이용되기 때문에 이상값의 영향에 매우 민감하다. 본 연구에서는 이러한 이상값의 영향을 줄이기 위해 비선형 변환된 자료를 이용한 베리오그램 모델링을 수행하였는데, 오히려 이상값을 포함하는 원 자료의 베리오그램에 비해 상대적 너겟 효과가 더 크게 나타났다. 로그 변환된 자료

는 자료 변환을 통해 이상값의 영향을 일부 줄였지만, 오히려 자료의 대부분을 차지하는 낮은 값의 분포를 넓게 퍼트린 경향인 것으로 판단된다. 또한, 정규 점수 변환된 자료는 본 연구의 자료가 낮은 값에서 동일한 값을 다수 포함하는데 동일한 값들에 대해 주관적으로 순위를 부여하고 표준 정규 분포에 일대일 매칭하는 과정으로 인해 자료 분포가 더 넓게 퍼트려지는 경향이 나타난 것으로 판단된다. 이러한 베리오그램 모델링 결과는 크리깅 가중치 결정 과정에 직접적으로 반영되므로 다중 가우시안

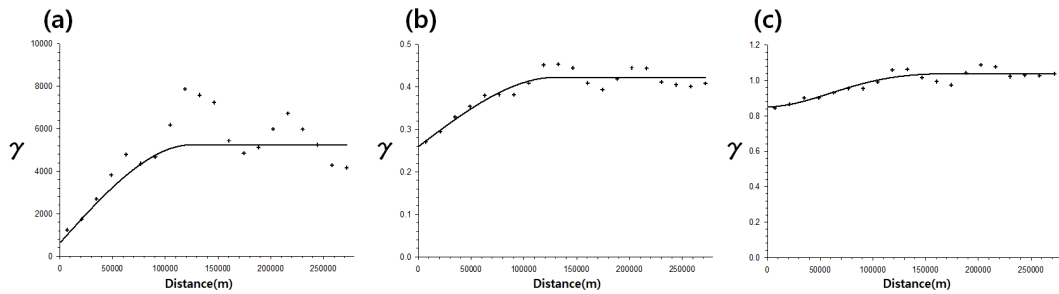


FIGURE 3. Experimental variograms with models fitted for various kriging algorithms. (a)indoor radon concentration data, (b)logarithm data, (c)normal score transform data

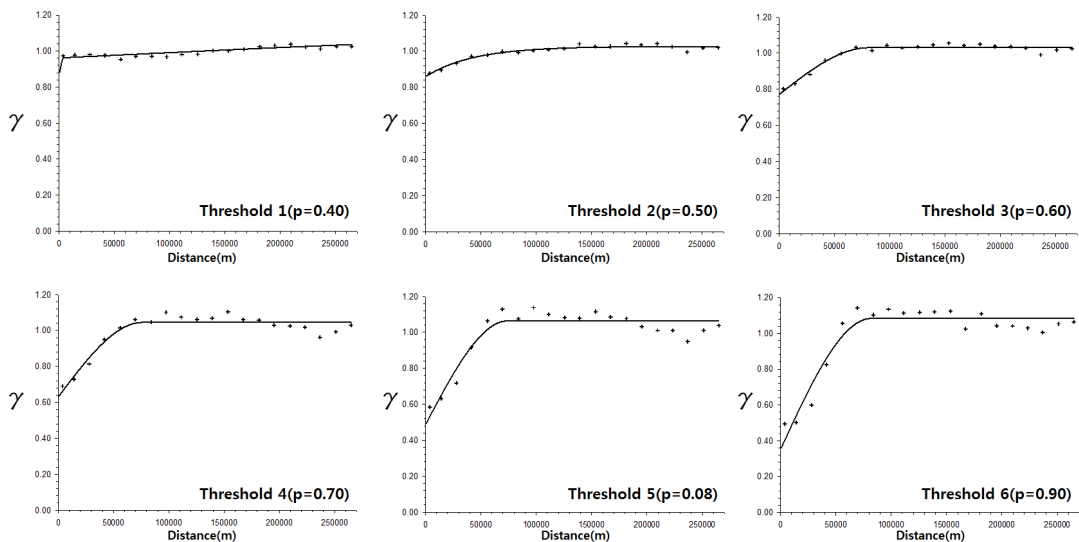


FIGURE 4. Experimental variograms computed for thresholds 1 through 6, with models fitted using indicator transform data

크리깅과 로그 정규 크리깅의 예측 능력에 큰 영향을 미칠 것으로 판단된다.

2. 크리깅 기법별 예측 능력 비교

베리오그램 모델링 후에 각 기법 별로 크리깅을 수행하였으며, 크리깅 수행 시 최대 사용 자료의 개수는 표본 자료의 공간 분포를 고려하여 16개로 설정하였다. 정규 크리깅, 로그 단순 크리깅, 다중 가우시안 크리깅의 적용에는 SpaceStat 4.0(Jacquez *et al.*, 2014)을, 지시자 크리깅의 적용에는 Auto-IK(Goovaerts, 2009)를 이용하였다. 그림 5는 크리깅 기법별 잭나이프 방법을 통해 얻어진 예측값과 참값의 산점도를 나타낸다. 그림 5를 살펴보면, 미관측

지점에서 분산을 최소로 하는 방향으로 값을 추정하기 때문에 낮은 값들은 크게, 높은 값들은 작게 예측하는 평활화 양상이 모든 크리깅 기법에서 확인되고 있다. 그러나, 자연 방사능 물질인 라돈은 위해성이 큰 물질로 높은 값이 더 중요한 의미를 가지므로, 크리깅 기법들의 높은 값에 대한 예측 능력이 중요하다. 각 크리깅 기법별 라돈 농도 분포도를 살펴보면(그림 6), 로그 크리깅, 다중 가우시안 크리깅, 지시자 크리깅에서는 평활화 효과로 인해 높은 값들을 반영하지 못하는 반면 정규 크리깅에서는 다른 기법에 비해 상대적으로 높은 값을 잘 반영하는 것으로 나타났다. 비선형 자료 변환 기반의 크리깅 기법들은 이상값의 영향을 감소시키기

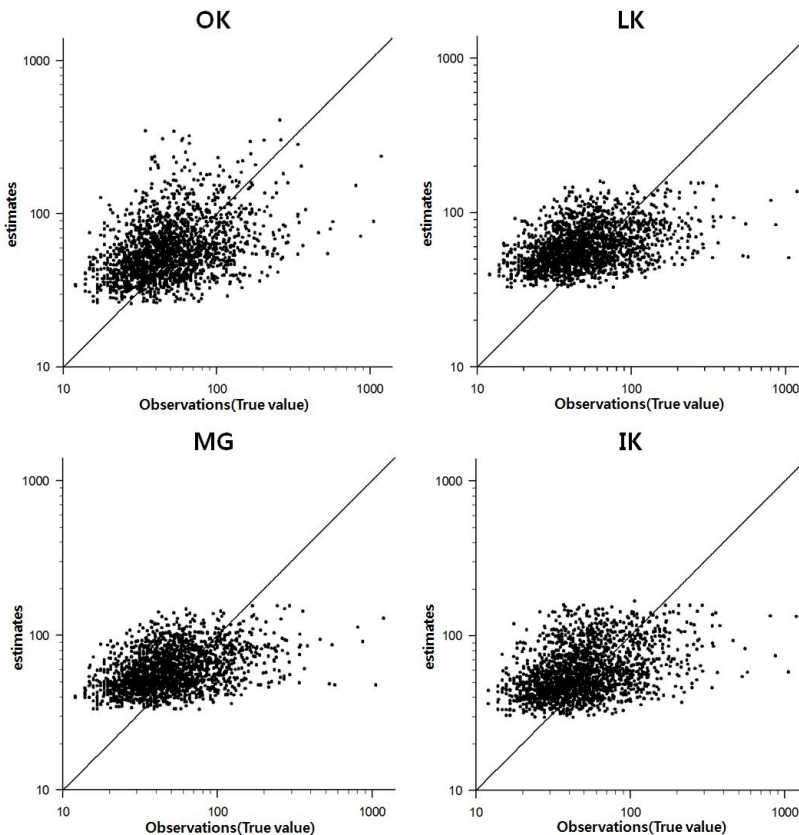


FIGURE 5. Scatterplots of observations versus estimates from various kriging algorithms

위해 변환된 값을 이용하여 공간 예측을 수행한 후에 역변환을 적용하지만, 정규 크리깅은 이상값을 공간 예측에 그대로 적용하기 때문인 것으로 판단된다.

표 1에서는 잭나이프 방법을 통해 각 크리깅 기법별 예측 능력에 대해 정량적으로 비교를 수행하였다. ME는 모든 크리깅 기법에서 0에 가까운 값을 보여 편향이 보이지 않는 것으로 나타났다. MAE와 RMSE는 모든 크리깅 기법에서 그 차이가 두드러지게 나타나지 않았으나, 상대적으로 다중 가우시안 크리깅과 로그 단순

크리깅에서 작게 나타났고, 정규 크리깅에서 가장 크게 나타났다. 또한, RMAE와 RRMSE를 통해서도 다중 가우시안 크리깅과 로그 단순 크리깅에서 오차가 가장 작게 나타나는 것을 확인할 수 있었다. 크리깅 예측값과 원 자료의 참값 간의 선형상관계수에서는 모든 크리깅 기법에서 그 차이가 두드러지지 않았다. 분석 결과를 토대로 편향 정도, 오차의 크기 등의 전체 정확도 측면에서 살펴본 결과, 로그 단순 크리깅, 지시자 크리깅, 다중 가우시안 크리깅과 정규 크리깅 순으로 예측 능력이 좋게 나타났다.

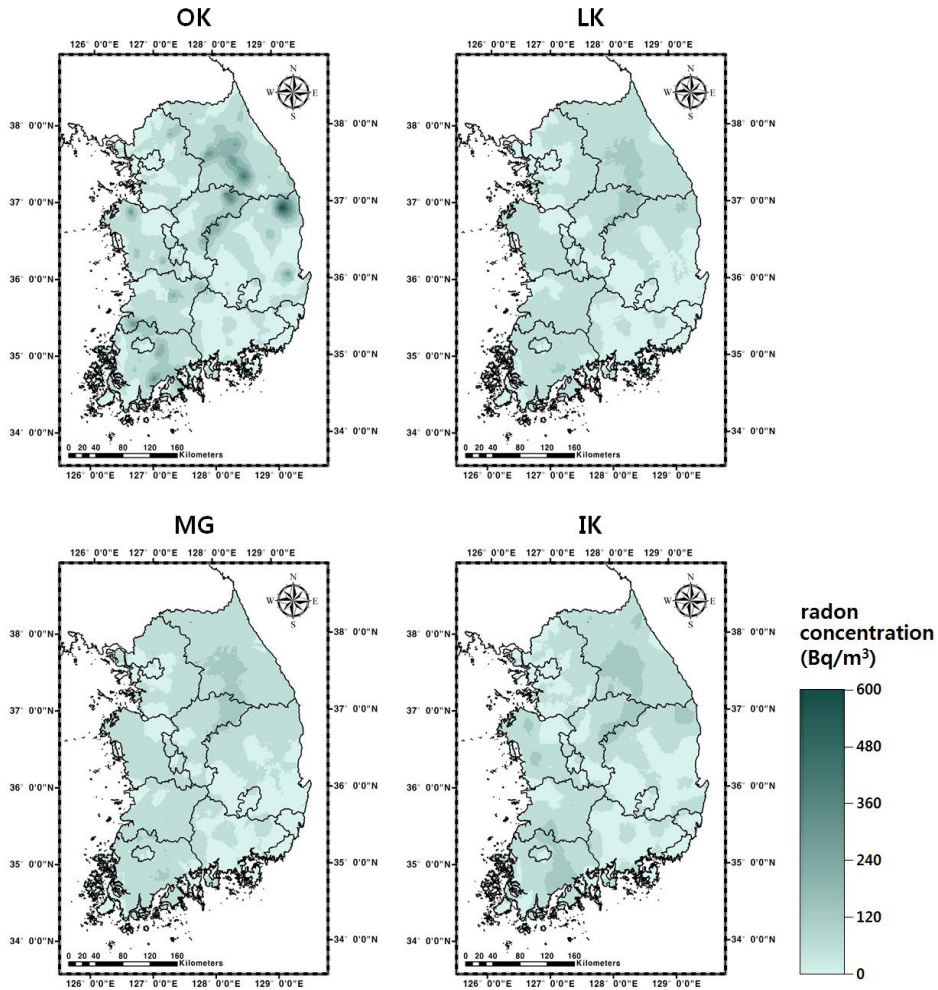


FIGURE 6. Radon concentration maps generated by various kriging algorithms

TABLE 1. Comparison of error statistics of various kriging algorithms

	ME	MAE	RMAE(%)	RMSE	RRMSE(%)	Correlation
OK	1.56	33.84	—	65.83	—	0.283
LK	0.10	31.64	6.49	62.68	4.78	0.298
MG	0.05	31.69	6.35	62.84	4.54	0.290
IK	0.23	32.44	4.14	63.46	3.59	0.279

3. 구간별 예측 능력 비교

표 2에서는 자료값을 백분위수 구간별로 나누어 크리깅 기법별로 예측 능력을 비교하였다. 구간별 ME를 살펴보면, 모든 크리깅 기법들이 70% 백분위수 이하값들에서 과추정 양상을 보였고, 70% 백분위수를 초과하는 값들에서는 저추정 양상을 나타냈다. 정규 크리깅의 ME는 다른 크리깅 기법에 비해 상대적으로 0에 가깝게 나타나, 이러한 편향 정도가 다른 크리깅 기법에 비해 작은 것으로 나타났다. 또한, 정규 크리깅을 제외한 나머지 크리깅 기법에서는 ME값의 차이가 두드러지게 나타나지 않았다. 특히, 모든 크리깅 기법들은 90% 백분위수를 초과하는 값들에서 저추정 양상이 두드러지게 나타났다. 구간별 MAE에서는 30% 백분위수 이하값들에서 정규 크리깅이 가장 작게 나타났고, 나머지 크리깅 기법들은 차이가 두드러지지 않았다. 그러나, 50%에서 90% 백분위수 사이값들에서는 정규 크리깅이 나머지 크리깅 기법에 비해 MAE가 상당히 크게 나타났으며, 다중 가우시안 크리깅과 로그 정규 크리깅의 MAE가 가장 작게

나타났다. 90% 백분위수 이상값들에서는 모든 크리깅 기법에서 MAE가 매우 크게 나타나 큰 이상값들을 반영하지 못하는 것으로 나타났다.

종합적으로 구간별 분석 결과를 살펴보면, 모든 크리깅 기법에서 크리깅의 일반적인 특성인 낮은 값의 과추정과 높은 값의 저추정 양상이 나타났다. 각 크리깅 기법들은 구간별로 예측 능력이 다르게 나타났는데 전반적으로 50%에서 90% 백분위수 사이값들에서는 다중 가우시안 크리깅과 로그 단순 크리깅의 예측 능력에서 좋은 예측 능력을 보였다. 반면, 90% 백분위수를 초과하는 값들에서의 예측 능력은 모든 크리깅 기법에서 그 차이가 두드러지지 않았다. 또한, 표 1을 통해 전체 자료에 대한 기본 통계 분석을 수행하였을 때 평균 오차가 모든 크리깅 기법에 걸쳐 0에 가까운 값을 나타냈는데, 이는 70% 백분위수 이하 값들의 과추정과 70% 백분위수를 초과하는 값들의 저추정이 상쇄된 결과인 것으로 판단된다.

표 2에 따르면 90% 백분위수를 초과하는 값들에서의 예측 능력의 차이가 크게 두드러지지

TABLE 2. Comparison of ME and MAE values of various kriging algorithms for percentile classes of radon concentrations

Unit : %	ME				MAE			
	OK	LK	MG	IK	OK	LK	MG	IK
< 10	25.13	29.11	29.61	27.89	25.13	29.11	29.61	27.89
[10, 20)	22.41	25.03	25.19	24.66	22.43	25.03	25.19	24.66
[20, 30)	21.60	22.84	23.09	22.52	21.83	22.84	23.09	22.53
[30, 40)	23.41	22.48	22.37	22.52	24.95	22.73	22.59	23.25
[40, 50)	17.51	18.86	18.95	17.72	20.45	19.74	19.82	19.61
[50, 60)	15.71	14.68	14.85	14.47	21.84	17.70	17.92	19.40
[60, 70)	10.88	8.16	8.13	8.29	23.62	16.24	16.21	20.07
[70, 80)	1.92	-1.93	-2.44	-2.03	28.33	18.82	17.93	22.90
[80, 90)	-18.44	-22.43	-22.62	-21.24	35.73	28.37	27.91	31.43
≥ 90	-104.57	-115.80	-116.64	-112.45	114.10	115.82	116.64	112.63

TABLE 3. Comparison of ME and MAE values of various kriging algorithms for above 148 and 200Bq/m³ of radon concentrations

		ME	MAE	RMAE(%)	RMSE	RRMSE(%)
above 148Bq/m ³	OK	-157.11	169.05	-	236.66	-
	LK	-173.89	173.89	-2.86	244.95	-3.50
	MG	-175.49	175.49	-3.81	246.10	-3.99
	IK	-170.37	170.37	-0.78	243.18	-2.75
above 200Bq/m ³	OK	-227.21	237.70	-	308.09	-
	LK	-247.87	247.87	-4.28	320.39	-3.99
	MG	-249.50	249.50	-4.96	321.85	-4.47
	IK	-244.57	244.57	-2.89	318.42	-3.35

않았으나, 그림 6의 라돈 농도 분포도를 살펴보면 정규 크리깅이 높은 값을 가장 잘 반영하는 것으로 나타났다. 이에 따라 라돈 농도의 유의값에서의 예측 능력을 비교하기 위해 추가적으로 검증 자료에서 148Bq/m³과 200Bq/m³ 이상의 값들을 추출한 값을 이용하여 각 크리깅 기법별로 정량적 비교를 수행하였다(표 3). 검증 자료에서 라돈 농도 148Bq/m³과 200Bq/m³은 각각 상위 2.58%와 5.12%에 해당하며, 큰 이상값을 대변하고 있다. 표 3을 살펴보면 높은 값의 저 추정 영향으로 인해 모든 크리깅 기법에서 ME는 큰 음의 값을 나타냈고, MAE와 RMSE에서 큰 오차의 크기를 확인할 수 있었다. 평활화 효과로 인해 큰 이상값들은 잘 반영하지 못하지만, RMAE, RRMSE를 살펴보면 그 중에서도 비선형 변환에 기반을 둔 크리깅 기법들보다 정규 크리깅의 예측 능력이 가장 우수한 것을 확인할 수 있었다. 이러한 결과 양상은 148Bq/m³과 200Bq/m³ 이상에서 동일하게 나타났다.

4. 샘플링 밀도에 따른 예측 능력 비교

추가적으로 본 연구에서는 관측 자료의 개수에 따라 크리깅 기법별 예측 능력을 살펴보기 위해 훈련 자료로부터 25%, 50%, 75%를 각각 샘플링한 자료를 이용하여 크리깅을 수행하고, 잭나이프 방법을 통해 기존 검증 자료를 이용하여 검증을 수행하였다. 전체 자료에 대한 임의의 샘플링은 특정 구간의 자료양이 많아지거나 적어질 수 있기 때문에 백분위수 별로 임

의 샘플링을 적용하였다.

표 4는 25%, 50%, 75% 샘플링 밀도에 따른 오차 통계값을 나타낸다. 샘플링 밀도에 따른 ME의 변화 양상을 살펴보면 모든 크리깅 기법에서 그 양상이 일관되지 않게 나타났는데, 이러한 현상은 백분위수 별로 어떤 값을 샘플링 하나에 따라 편향 정도가 변하기 때문인 것으로 판단된다. 샘플링 밀도에 따른 MAE와 RMSE의 차이는 모든 크리깅 기법에서 크게 두드러지게 나타나지는 않았다. 이는 샘플링을 통해 특정 자료값이 제외되더라도 대부분의 자료값이 100Bq/m³ 미만에 밀집되어 분포하고 있기 때문에 제외된 다른 값들의 영향을 대체하였기 때문인 것으로 판단된다. 샘플링 밀도에 따른 차이가 두드러지는 않았으나 RMAE와 RRMSE를 살펴보면, 일관되게 정규 크리깅의 예측 능력이 가장 낮고, 로그 정규 크리깅이 정규 크리깅의 MAE에 비해 가장 향상된 예측 능력을 나타냈다. 상관계수를 살펴본 결과, 모든 크리깅 기법에서 50%, 75%의 샘플링 밀도는 모두 비슷한 양상을 보였으나, 25%의 샘플링 밀도에서는 낮아지는 것을 확인하였고 그 차이는 정규 크리깅에서 가장 크게 나타났다. 샘플링 밀도에 따라 각 크리깅 기법을 비교한 결과를 종합적으로 살펴보면 모든 샘플링 밀도에서 로그 정규 크리깅이 가장 예측 능력이 뛰어났고, 그 다음으로는 지시자 크리깅, 다중 가우시안 크리깅, 정규 크리깅 순으로 예측 능력이 좋은 것으로 나타났다.

TABLE 4. Comparison of error statistics of various kriging algorithms with respect to variations of sampling density

		ME	MAE	RMAE(%)	RMSE	RRMSE(%)	Correlation
25%	OK	-1.43	34.96	-	69.77	-	0.151
	LK	-3.75	30.91	11.59	63.80	8.56	0.237
	MG	1.78	32.78	6.25	63.85	8.48	0.249
	IK	-1.79	32.53	6.95	65.05	6.77	0.196
50%	OK	0.08	33.57	-	66.19	-	0.286
	LK	-3.31	30.50	9.15	62.45	5.66	0.306
	MG	0.92	32.06	4.49	62.58	5.45	0.306
	IK	-1.27	31.73	5.49	63.29	4.38	0.275
75%	OK	2.56	34.62	-	66.79	-	0.273
	LK	-1.43	31.07	10.28	62.44	6.51	0.305
	MG	3.51	33.32	3.78	63.21	5.37	0.298
	IK	0.55	32.10	7.30	62.57	6.33	0.301

결론

현장 조사 자료는 다양한 분석 수행을 위해 분포도 제작이 필수적이다. 일반적으로 현장 조사를 통해 얻어지는 자료는 양이나 음으로 치우친 비대칭 분포를 나타내는 경우가 많고, 낮은 값 혹은 높은 값의 분포가 중요한 역할을 하는 목적 때문에 분포도 작성에 유의해야 한다. 이러한 자료 분포 특성과 분포도 작성의 목적에 따라 적합한 공간 예측 기법의 선택이 필수적이다. 본 연구에서는 비대칭 분포를 나타내는 실내 라돈 농도 자료를 이용한 라돈 농도 분포도 작성을 위해 단변량 크리깅 기법 중에서 정규 크리깅과 비선형 자료 변환 기반의 로그 단순 크리깅, 다중 가우시안 크리깅과 지시자 크리깅을 비교 분석하였다.

각 크리깅 기법들은 잭나이프 방법 기반의 검증을 통해 정량적 비교를 수행한 결과, 전반적인 예측 능력에서는 정규 크리깅에 비해 비선형 자료 변환 기반의 크리깅 기법이 좋은 예측 능력을 보였다. 비선형 자료 변환 기반 크리깅 기법에서는 로그 정규 크리깅, 지시자 크리깅, 다중 가우시안 크리깅 순서로 우수한 예측 능력을 나타냈다. 백분위수의 구간별 예측 능력을 비교한 결과에서도 40%에서 90% 백분위수 사이값들에서 비선형 자료 변환 기반의 크리깅 기법에 비해 정규 크리깅의 오차가 상대적으로

크게 나타났다. 그러나, 라돈 농도는 폐암을 유발하는 위해성이 큰 물질로 높은 값에 유의해야 하는데, 공간분포 측면에서 정규 크리깅이 높은 값을 가장 잘 반영하는 것을 확인하였다. 또한, 라돈 농도의 유의값으로 알려진 148Bq/m³과 200Bq/m³ 이상의 값들을 확인한 결과, 오히려 비선형 자료 변환 기반의 크리깅 기법보다 정규 크리깅에서 좋은 예측 능력을 보였다. 이는 비선형 자료 변환 기반의 크리깅 기법들이 이상값의 영향을 감소시키기 위해 변환된 자료값을 이용하여 공간 예측을 수행하기 때문인 것으로 판단된다.

일반적으로 비대칭 분포를 나타내는 현장 조사 자료에 대해 크리깅 기법을 이용하여 분포도를 작성할 경우, 전반적으로 비선형 자료 변환 기반 크리깅 기법의 예측 능력이 우수할 것으로 인식되어 왔다. 이와 달리 본 연구에서는 높은 값에 유의해야 하는 라돈 농도 자료의 특성을 고려할 때 정규 크리깅이 높은 값에서의 예측 능력이 다른 기법에 비해 우수한 것으로 나타났다. 그러나, 본 연구 결과는 단일 사례 연구를 통해 도출되었기 때문에 다른 비대칭 자료를 통해 정규 크리깅의 적용 가능성을 확인할 필요가 있다. 추후에는 실내 라돈 농도가 지질이나 토양 투수율 등 다양한 요소에 영향을 받기 때문에 이러한 영향을 고려한 분포도 작성을 위해 상관성 높은 부가자료를 사용하는

다변량 크리깅 기법의 적용이 병행되어야 할 것으로 판단된다. 또한, 라돈 가스는 주요 폐암 유발 인자로 알려져 있기 때문에, 폐암 사망률의 분포 특성 분석에 라돈 농도 분포를 다른 인자들과 함께 이용하는 의학 통계와의 연계도 추후 고려되어야 할 것으로 판단된다. **KAGIS**

REFERENCES

- Baek, S.A., T.J. Lee, S.D. Kim, and D.S. Kim. 2008. Studies on the spatial analysis for distribution estimation of radon concentration at the seoul area. *Journal of Korean Society for Atmospheric Environment* 24(5):538-550 (백승아, 이태정, 김신도, 김동술. 2008. 서울지역 라돈농도의 분포예측을 위한 공간분석법 연구. *한국대기환경학회지* 24(5):538-550).
- Bossey, P., Z.S. Žunić, Z. Stojanovska, T. Tollefsen, C. Carpentieri, N. Veselinović, S. Komatina, J. Vaupotič, R.D. Simović, S. Antignani, and F. Bochicchio. 2014. Geographical distribution of the annual mean radon concentrations in primary schools of Southern Serbia-application of geostatistical methods. *Journal of Environmental Radioactivity* 127:141-148.
- Buttafuoco, G., A. Tallarico, and G. Falcone. 2007. Mapping soil gas radon concentration: a comparative study of geostatistical methods. *Environmental Monitoring and Assessment* 131(1-3):135-151.
- Cha, D.W. 2001. A study on mitigation methods of indoor radon concentration in residential buildings(I) - test cell study. *Korea Institute of Ecological Architecture and Environment* 1(2):21-28 (차동원. 2001. 주거용 건축물의 실내 라돈농도 경감방안에 관한 연구(I) - test cell study. *한국생태환경건축학회 논문집* 1(2):21-28).
- Deutsch, C.V. and A.G. Journel. 1998. *Geostatistical software library and user's guide*. Oxford University Press, New York, U.S.. pp.63-86.
- Drolet, J.P., R. Martel, P. Poulin, J.C. Dessau, D. Lavoie, M. Parent, and B. Lévesque. 2013. An approach to define potential radon emission level maps using indoor radon concentration measurements and radiogeochemical data positive proportion relationships. *Journal of Environmental Radioactivity* 124:57-67.
- Dubois, G. 2005. An overview of radon surveys in europe. *Institute for Environment and Sustainability. European Commission*. pp.2-9.
- EPA(Environmental Protection Agency). 1992. *Technical support document for the 1992 citizen's guide to radon*, U.S.. pp.1-3.
- Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York, U.S.. pp.265-328.
- Goovaerts, P. 2009. AUTO-IK: a 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. *Computers and Geosciences* 35(6):1255-1270.
- Green, B.M.R., R. Larmour, J.C.H. Miles, D.M. Rees, and F.K. Ledgerwood. 2009. *Radon in dwellings in Northern Ireland: 2009 review and atlas*. Northern Ireland Environment Agency, Chilton, Oxon, U.K.. pp.21-26.
- Harman, B.I., H. Koseoglu, and C.O. Yigit. 2016. Performance evaluation of IDW, kriging and multiquadric interpolation

- methods in producing noise mapping: a case study at the city of Isparta, Turkey. *Applied Acoustics* 112:147–157.
- Jacquez, G.M., P. Goovaerts, A. Kaufmann, and R. Rommel. 2014. SpaceStat 4.0 user manual: software for the space-time analysis of dynamic complex systems. BioMedware, Washington, U.S., pp.419–440.
- Kim, Y.J., B.U. Chang, H.M. Park, C.K. Kim, and S. Tokonami. 2011. National radon survey in Korea. *Radiation Protection Dosimetry* 146(1–3):6–10.
- Lee, H.S. 2010. Comparison and evaluation of root mean square for parameter settings of spatial interpolation method. *Journal of the Korean Association of Geographic Information Studies* 13(3):29–41 (이형석. 2010. 공간보간법의 매개변수 설정에 따른 평균제곱근 비교 및 평가. *한국지리정보학회지* 13(3):29–41).
- Park, H.J., H.S. Shin, Y.H. Roh, K.M. Kim, and K.H. Park. 2012. Estimating forest carbon stocks in Danyang using kriging methods for aboveground biomass. *Journal of the Korean Association of Geographic Information Studies* 15(1):16–33 (박현주, 신휴석, 노영희, 김경민, 박기호. 2012. 크리깅 기법을 이용한 단양군의 산림 탄소저장량 추정 -지상부 바이오매스를 대상으로-. *한국지리정보학회지* 15(1):16–33).
- Park, J.C. and M.K. Kim. 2013. Comparison of precipitation distributions in precipitation data sets representing 1km spatial resolution over South Korea produced by PRISM, IDW, and cokriging. *Journal of the Korean Association of Geographic Information Studies* 16(3):147–163 (박종철, 김만규. 2013. PRISM, 역거리가중법, 공동크리깅으로 작성한 1km 공간해상도의 남한 강수 자료에서 강수 분포의 비교. *한국지리정보학회지* 16(3):147–163).
- Park, N.W. 2009. Comparison of univariate kriging algorithms for GIS-based thematic mapping with ground survey data. *Korean Journal of Remote Sensing* 25(4):321–338 (박노옥. 2009. 현장 조사 자료를 이용한 GIS 기반 주제도 작성을 위한 단변량 크리깅 기법의 비교. *대한원격탐사학회지* 25(4):321–338).
- Pásztor, L., K.Z. Szabó, G. Szatmári, A. Laborczi, and Á. Horváth. 2016. Mapping geogenic radon potential by regression kriging. *Science of the Total Environment* 544:883–891.
- Saito, H. and P. Goovaerts. 2000. Geostatistical interpolation of positively skewed and censored data in a dioxin-contaminated site. *Environmental Science and Technology* 34(19):4228–4235.
- Smethurst, M.A., T. Strand, A.V. Sundal, and A.L. Rudjord. 2008. Large-scale radon hazard evaluation in the Oslofjord region of Norway utilizing indoor radon concentrations, airborne gamma ray spectrometry and geological mapping. *Science of the Total Environment* 407:379–393.
- Webster, R. and M.A. Oliver. 2007. *Geostatistics for environmental scientists*. John Wiley & Sons, Ltd, West Sussex, U.K.. pp.153–193. **KAGIS**