

다중 자료 변환을 이용한 구성 자료의 지구통계학적 시뮬레이션

박 노 옥*

인하대학교 지리정보공학과, 402-751, 인천광역시 남구 인하로 100

Geostatistical Simulation of Compositional Data Using Multiple Data Transformations

No-Wook Park*

Department of Geoinformatic Engineering, Inha University, Incheon 402-751, Korea

Abstract: This paper suggests a conditional simulation framework based on multiple data transformations for geostatistical simulation of compositional data. First, log-ratio transformation is applied to original compositional data in order to apply conventional statistical methodologies. As for the next transformations that follow, minimum/maximum autocorrelation factors (MAF) and indicator transformations are sequentially applied. MAF transformation is applied to generate independent new variables and as a result, an independent simulation of individual variables can be applied. Indicator transformation is also applied to non-parametric conditional cumulative distribution function modeling of variables that do not follow multi-Gaussian random function models. Finally, inverse transformations are applied in the reverse order of those transformations that are applied. A case study with surface sediment compositions in tidal flats is carried out to illustrate the applicability of the presented simulation framework. All simulation results satisfied the constraints of compositional data and reproduced well the statistical characteristics of the sample data. Through surface sediment classification based on multiple simulation results of compositions, the probabilistic evaluation of classification results was possible, an evaluation unavailable in a conventional kriging approach. Therefore, it is expected that the presented simulation framework can be effectively applied to geostatistical simulation of various compositional data.

Keywords: conditional simulation, compositional data, uncertainty modeling, data transformation

요약: 이 논문에서는 구성 자료의 지구통계학적 시뮬레이션을 위해 다중 자료 변환 기반 조건부 시뮬레이션 틀을 제안하였다. 우선 일반적인 통계 기법의 적용이 가능하도록 구성 자료에 로그비 변환을 적용하였다. 다음 변환들로는 최소/최대 자기상관 인자 변환과 지시자 변환을 순차적으로 적용하였다. 독립적인 새로운 변수의 생성을 위해 최소/최대 자기상관 인자 변환을 적용하였으며, 적용 결과 개별 변수들의 독립적인 시뮬레이션이 가능해진다. 그리고 다중 가우시안 확률 모델을 따르지 않는 변수들의 비모수적 조건부 누적 확률 분포 모델링을 위해 지시자 변환을 적용하였다. 최종적으로는 적용한 변환 방법들의 역순으로 역 변환을 적용하였다. 간석지 표층 퇴적물 성분 자료를 대상으로 제안 시뮬레이션 기법의 적용 가능성을 제시하였다. 모든 시뮬레이션 결과들은 구성 자료의 제한 조건을 만족하면서 샘플 자료의 통계 특성을 잘 반영하였다. 구성 자료의 다수의 시뮬레이션 결과들을 이용한 표층 퇴적물 분류를 통해 기존 크리깅에서는 얻을 수 없는 분류 결과의 확률론적 평가가 가능하였다. 따라서 제안 시뮬레이션 틀은 다양한 구성 자료의 지구통계학적 시뮬레이션에 효과적으로 이용될 수 있을 것으로 기대된다.

주요어: 조건부 시뮬레이션, 구성 자료, 불확실성 모델링, 자료 변환

*Corresponding author: nwpark@inha.ac.kr

Tel: +82-32-860-7607

Fax: +82-32-863-1506

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서론

지구과학의 많은 응용 분야에서 수행되는 현장 조사 자료를 이용하여 특정 연구지역의 분석을 수행하기 위해서는 대부분 공간 내삽(interpolation)을 수행한다. 이러한 공간 내삽과 관련해서 크리깅으로 대표되어온 지구통계학은 다양한 부가 자료를 통합할 수

있고, 불확실성 모델링이 가능한 장점으로 인해 지구 과학의 많은 분야에서 활용되어 왔다(Kyriakidis et al., 2004; Oh, 2005; Goovaerts, 2010; Park, 2010; Park, 2011). 특히 지구통계학적 시뮬레이션은 단일 위치나 동시에 여러 위치에서 고려하고 있는 속성값의 불확실성 추정에 활용될 수 있다(Goovaerts, 1997; Deutsch and Journel, 1998; Chilès and Delfiner, 2012). 예를 들어 특정 지역에서 측정 단위보다 큰 블록 단위에서 오염 임계치를 초과할 확률을 계산하거나, 2차 분석에 사용되는 모델에 입력 자료의 불확실성 전과 등을 정량적으로 모델링하는데 사용될 수 있다(Kyriakidis and Dungan, 2001; Saito and Goovaerts, 2003; Wang et al., 2003; Park and Oh, 2006; Goovaerts et al., 2008; Oh and Han, 2010).

지구과학 자료는 값 자체의 연산이 가능한 연속형 자료와 서로 구별이 되는 몇 개의 범주로 구성된 범주형 자료로 구분할 수 있다(Davis, 2002). 연속형 자료 중에서 퇴적물 혹은 토양의 성분 자료는 전체에 대한 상대적인 비율 정보를 제공하는 구성 자료(compositional data)에 해당된다. 예를 들어 간석지 표층 퇴적물이나 토양 시료의 모래, 실트, 점토 비율 값들이 구성 자료에 해당한다. 이러한 구성 자료는 일반적인 연속형 자료와 달리 몇 가지 고유한 제한 조건을 가진다. 즉 각 성분들은 음의 값을 가질 수 없으며, 전체 합은 일정 상수가 되어야 한다. 예를 들어 상대적인 비율로 표현할 경우, 전체 합은 1 혹은 100%이어야 한다. 이러한 구성 자료를 대상으로 일반적인 통계 분석 방법론을 적용할 경우, 전술한 제한 조건을 만족시키지 않는 분석 결과를 나타낼 수 있다. 대표적인 공간 내삽 기법인 크리깅은 샘플 자료 범위 안에서 속성값을 예측하는 것이 아닌 non-convex 내삽 기법이기 때문에 물리적인 의미가 없는 음의 값이 나타날 수 있다. 또한 전체 합이 일정 상수인 제한 조건으로 인해 특정 성분들의 절대적인 값이 양의 상관관계를 나타내더라도 특정 성분의 비율값이 증가하면 다른 성분 중 일부는 반드시 감소하는 음의 상관관계를 나타내게 된다. 일반적인 연속형 자료와 다른 이러한 특성들은 구성 자료들이 심플렉스(simplex)라는 제한된 공간에 존재하기 때문이다(Aitchison, 1986). 일반적으로 다변량 통계 기법은 특정 공간에 제한되지 않는 자료에 적용이 가능하기 때문에, 제한된 공간 안에서 값을 가지는 구성 자료에 직접적으로 적용할 수 없는 것으로 알려져 있다

(Pawlowsky-Glahn and Buccianti, 2011).

소수의 지점에서 획득된 구성 자료의 공간 내삽 혹은 시뮬레이션을 위해서는 전통적인 지구통계학 기법을 직접적으로 적용할 수 없기 때문에, 특화된 기법의 적용이 필요하다. 이와 관련하여 구성 자료에 자료 변환 기법을 적용하여 심플렉스 공간에서 자료 값의 제한이 없는 공간으로 변환함으로써 일반적인 다변량 통계기법을 적용이 가능한 것으로 알려져 있다(Pawlowsky-Glahn and Olea, 2004; Pawlowsky-Glahn and Buccianti, 2011). 대표적인 자료 변환 기법으로 로그비(log-ratio) 변환이 주로 사용되었는데, 대상 자료에 공동 크리깅(cokriging)과 같은 다변량 크리깅 기법을 적용한 후에 적절한 역변환을 수행하게 되면 구성 자료의 제한조건을 모두 만족시킬 수 있다. 이러한 로그비 변환과 결합된 크리깅 기법은 주로 토양학의 토성 자료, 연안 퇴적물 입도 자료, 저류층 퇴적물 특성 자료 등에 적용되어 왔다(Odeh et al., 2003; Lark and Bishop, 2007; Lark et al., 2012). 그러나 국내에서는 구성 자료의 제한 조건을 고려하지 않고 개별 성분에 독립적으로 공간 내삽 기법을 적용한 연구가 대부분이었으며(Oh and Kum, 2001; Shin et al., 2004), 이러한 구성 자료에 특화된 방법론을 공간 내삽에 적용한 사례는 최근 Park and Jang (2014)에 의해 수행되었을 뿐 매우 드문 상황이다. 그리고 로그비 변환 자료에 대해 공동 크리깅을 적용한 결과를 최종적으로 원 자료 스케일로 변환할 경우, 편향된 결과를 나타내는 것으로 알려져 있다(Pawlowsky-Glahn and Olea, 2004). 크리깅은 최소 자승 관점에서 최적의 선형 비편향 예측자(best linear unbiased estimator, BLUE)이지만, 로그비 변환 자료에 대해 적용할 경우 로그의 특성상 선형 결합 즉 산술 평균이 아닌 기하 평균이 비편향 예측값이기 때문이다. 공간 내삽 이외에 지구통계학적 시뮬레이션을 적용한 사례로는 최근 Emery and Gálvez (2012)이 plurigaussian 모델을 적용한 사례와 토양 유실 인자의 공간 불확실성 모델링에 적용한 사례(Buttafuoco et al., 2012) 이외에는 국외와 국내 모두 드문 상황이다. 또한 전술된 연구들은 적용이 상대적으로 간편한 다중 가우시안 확률 모델을 적용하였는데, 이 모델에 사용된 다점 가우스 분포 가정의 적합성은 이론적으로 검증이 힘들고 양이나 음으로 치우친 분포를 가진 자료에 대해서는 적합하지 않은 경우가 많다.

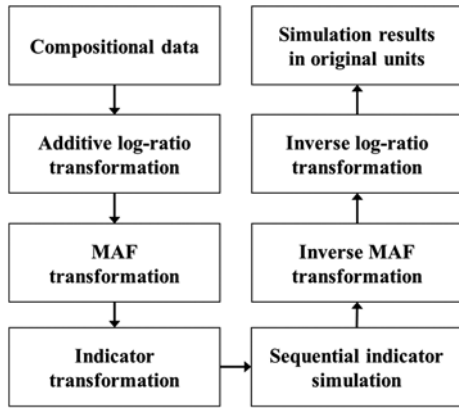


Fig. 1. The work flows applied for conditional simulation of compositional data.

이 연구에서는 구성 자료의 특성을 고려한 공간 예측과 더불어 공간 불확실성 모델링이 가능한 지구통계학적 시뮬레이션 기법을 제안하고자 한다(Fig. 1). 우선 구성 자료의 특성을 고려하기 위해 로그비 변환 방법으로 가산 로그비(additive log-ratio, alr) 변환을 적용하였다. 그리고 로그비 변환된 변수들을 직접적으로 시뮬레이션에 적용하지 않고, 로그비 변환 자료에 여전히 존재할 수 있는 상관성을 없애기 위해 최소/최대 자기상관 인자(minimum/maximum autocorrelation factor, MAF) 변환을 적용하였다. MAF 변환을 적용할 경우, 모든 이격거리에서 상관성이 없는 변수가 생성되기 때문에 공동 시뮬레이션이 아닌 독립적인 시뮬레이션을 적용할 수 있는 장점이 있다(Rondon, 2012). 그리고 다중 가우시안 확률 모델을 따르지 않는 경우를 고려해서 비모수적 지시자 변환(indicator transformation) 기반 순차적 지시자 시뮬레이션(sequential indicator simulation)을 개별 MAF 변수들에 적용하였다. 순차적 지시자 시뮬레이션 결과들은 MAF 변수들의 일종의 분위수를 나타내므로, 이들을 대상으로 자료 변환 적용 순서의 반대로 역 MAF 변환 및 역 가산 로그비 변환을 적용하면 원 구성 자료 스케일에서의 시뮬레이션 결과를 얻게 된다. 시뮬레이션 결과들은 일종의 분위수로 간주할 수 있으므로 이들로부터 구성 자료의 예측값, 분산 등과 더불어 특정 임계값을 초과할 확률 등의 계산이 가능하다. 이를 위한 사례연구로는 안면도 남단의 간석지에서 획득된 표층 퇴적물 입도성분 비율 자료를 활용하며, 제안 시뮬레이션 기법의 적용 절차와 적용 가능성을 제시하였다.

다중 자료 변환 기반 시뮬레이션

로그비 변환

구성 자료에 로그비 변환을 적용하면 심플렉스 공간은 자료값의 제한이 없는 공간으로 변환된다. 로그 변환을 적용하면 $(-\infty, +\infty)$ 범위의 자료값을 얻게 되므로 자료값의 제한 없이 일반적인 다변량 통계 분석 기법의 적용이 가능하다. 그리고 구성 자료의 특정 성분들이 로그비 변환시 모두 사용되므로 자료 변환을 통한 정보 손실이 없는 장점이 있다(Aitchison, 1986). 이 연구에서는 구성 자료의 변환에 가산 로그비 변환을 적용하였다.

M 차원 심플렉스 S^M 는 다음과 같은 제한 조건으로 표현될 수 있다.

$$S^M = [x_1, x_2, \dots, x_m]: x_i > 0; i=1, \dots, M; \sum_{i=1}^M x_i = C \quad (1)$$

여기서 C 는 일정 상수로, %로 표현할 경우 100이 된다.

위의 M 차원 구성 자료의 가산 로그비 변환은 식 (2)로 표현할 수 있으며, 가산 로그비 변환을 적용하면, M 차원 심플렉스 S^M 는 $M-1$ 차원으로 변수 1개가 줄어든 실수 공간으로 변환된다.

$$alr_i = \ln \left[\frac{x_i}{x_M} \right], i=1, \dots, M-1 \quad (2)$$

이렇게 가산 로그비 변환 자료를 대상으로 공간 내삽 혹은 시뮬레이션 기법을 적용한 후에, 초기 구성 자료 공간으로 재변환 시키기 위해서는 구성 자료의 합이 상수가 되는 제한 조건을 만족시키기 위해 식 (3)과 같이 역 가산 로지스틱 변환을 적용한다.

$$alr_i^{-1} = \frac{\exp(alr_i)}{\sum_{i=1}^{M-1} \exp(alr_i) + 1}, i=1, \dots, M-1 \quad (3)$$

$$alr_M^{-1} = \frac{1}{\sum_{i=1}^{M-1} \exp(alr_i) + 1}$$

이러한 가산 로그비 변환 이외에도 Egozcue et al. (2003)에 의해 제안된 직교정규(orthonormal) 좌표계 기반 변환을 이용하여 심플렉스 공간과 실수 공간 모두에서 기하학적 특성(거리와 각도)을 유지하는 등

거리 로그비(isometric log-ratio) 변환도 적용할 수 있다. 가산 로그비 변환과 등거리 로그비 변환은 서로 독립적이지 않으며, 직교정규 단위벡터로의 투영 행렬과 $M-1$ 차원의 단위 행렬과의 곱으로 표현할 경우 상호 변환이 가능하다(Egozcue et al., 2003). 이 두 가지 로그비 변환을 상호 비교하면, 우선 가산 로그비 변환은 변환 수식이 특정 고정된 성분에 대한 비율로 표현되기 때문에 상대적으로 변환된 변수의 해석이 용이한 장점이 있다. 그러나 등거리 로그비 변환과 달리 기하학적 특성이 유지되는 등거리 속성을 가지지 않기 때문에 분모에 해당하는 성분의 선택에 영향을 받을 수 있다. Aitchison (1986)에 의해 논의된 것처럼 크리깅과 같은 선형 통계 분석 기법을 적용할 경우에는 분모 성분의 선택이 이후 분석에 영향을 미치지 않는 것으로 알려져 있다. 반면 등거리 로그비 변환은 등거리 속성을 유지하지만, 변환 식 자체에 여러 성분이 혼재되어 있기 때문에 변환된 변수에 대한 해석이 용이하지 않은 단점이 있다. 이러한 이유로 이 연구에서는 가산 로그비 변환을 적용하였다.

MAF 변환

구성 자료는 전체 합이 일정 상수인 제한 조건 때문에 특정 성분값이 증가(감소)하면 다른 성분 중 일부는 반드시 감소(증가)하는 spurious 상관관계를 보인다. 구성 자료에 가산 로그비 변환을 적용하더라도, 때에 따라서는 변환된 자료들의 일부에서는 상관관계가 나타날 수 있다. 상관관계를 보이는 자료의 시뮬레이션에는 공동 시뮬레이션을 적용해야 하는데, 이 경우 다수의 베리오그램 모델링과 교차 베리오그램 모델링을 수행해야 하는 번거로움이 있다. 만약 N 개의 변수들이 존재할 경우, $N(N+1)/2$ 개의 베리오그램과 교차 베리오그램값의 모델링 과정이 필요하다. 그리고 일반적으로 가장 널리 사용되는 순차적 시뮬레이션을 적용할 경우, 크리깅과 시뮬레이션에 사용되는 자료수가 점차 증가하므로 많은 계산량을 요구된다. 따라서 서로 상관성이 존재하는 변수들에 대한 시뮬레이션에는 원 자료들을 서로 상관성이 없는 변수로의 변환을 적용한 후에 개별적으로 시뮬레이션을 적용하는 경우가 많다(Desbarats and Dimitrakopoulos, 2000; Boucher and Dimitrakopoulos, 2009).

다변량 자료들을 서로 상관성이 없는 자료로의 변환 방법으로 가장 널리 사용되어온 방법론은 주성분

변환(principal component transformation)이다. 그러나 주성분 변환은 이격 거리가 '0'인 경우에만 상관성이 없는 변수들로 변환을 수행하고, 다른 이격 거리에서는 상관성이 존재하는 단점이 있다(Goovaerts, 1993). 이러한 단점을 보완하기 위해 이격 거리가 '0'인 경우와 또 다른 이격 거리에서 상관성이 없는 변수로의 선형 변환을 수행하는 MAF 변환이 제안되었고(Switzer and Green, 1984), 이후 지구통계학적 시뮬레이션에 많이 적용되어 왔다(Desbarats and Dimitrakopoulos, 2000; Vargas-Guzmán and Dimitrakopoulos, 2003; Boucher and Dimitrakopoulos, 2009).

MAF 변환은 고려하고 있는 변수들의 분산-공분산 행렬을 이용하여 표준 주성분 변환을 적용한 후에 주성분 요소들을 고유벡터를 이용하여 추가적으로 직교 좌표로 회전시키는 2단계를 거친다(Rondon, 2012). MAF 변환의 자세한 사항 및 적용 절차는 Rondon (2012)를 참조할 수 있으며, 여기서는 주요 사항만을 기술하고자 한다. N 개의 변수들 $Y(u)=\{Y_1(u), \dots, Y_N(u)\}$ 가 있을 때, MAF 변환 변수들 $M(u)=\{M_1(u), \dots, M_N(u)\}$ 은 아래와 같이 직교 계수로 구성된 행렬 A 를 가중치로 하는 Y 의 선형결합으로 정의된다.

$$M(u)=A^T Y(u) \quad (4)$$

직교화 행렬 A 는 다음과 같이 고유치 분석에 의한 고유벡터들로 구성된다(Boucher and Dimitrakopoulos, 2009).

$$2\Gamma_Y(h)B^{-1}=A^T A A \quad (5)$$

$$\text{with } B=\text{cov}[Y(u), Y(u)],$$

$$2\Gamma_Y(h)=\text{cov}[Y(u)-Y(u+h), Y(u+h)] \quad (6)$$

여기서 B 는 $Y(u)$ 의 분산-공분산 행렬을 나타내며, $\Gamma_Y(h)$, A 및 cov 는 $Y(u)$ 의 베리오그램 행렬, 고유값으로 구성된 대각 행렬 및 공분산 함수를 각각 나타낸다.

만약 고려하고 있는 변수 Y 의 베리오그램과 교차 베리오그램이 2개의 모델로 구성된 a linear model of coregionalization (LMC)로 표현된다면 모든 이격 거리에서 MAF 변환된 변수들 사이의 상관관계가 사라진다. 그리고 3개 이상의 모델로 구성되었을 경우에는 이격거리가 0인 경우와 적어도 다른 하나의 이격 거리에서 상관관계가 사라지게 된다(Vargas-Guzmán and Dimitrakopoulos, 2003; Boucher and

Dimitrakopoulos, 2009). 역변환은 변환 행렬 A 와 B 를 이용하여 계산되며, $Y(u)$ 가 정규 분포를 따르지 않을 경우에는 변수 자체의 평균과 표준편차를 이용하여 원 자료 스케일로 변환한다.

순차적 지시자 시물레이션

공간 자료의 지구통계학적 조건부 시물레이션은 크리깅과 같이 최소자승 관점에서의 최적의 단일 해를 제공하는데 목적이 있는 것이 아니라, 몬테카를로 시물레이션을 적용하여 미지의 속성값에 대한 대안적인 다량의 값들을 제시하는데 목적이 있으며 이 과정에서 통계 특성과 공간 패턴을 재생산한다(Goovaerts, 1997).

이 연구에서는 기존 조건부 시물레이션 방법으로 널리 사용되어온 순차적 시물레이션을 적용하였으며, 조건부 누적 확률 분포(conditional cumulative probability distribution function) 모델링에는 지시자 크리깅을 이용하였다. 구성 자료는 특정 성분값이 매우 우세하여 다른 성분들이 매우 작은 값을 나타내거나, 전처리 과정에서 탐지 한계 미만의 값들이 다수 존재하는 경우가 있다. 이 경우 로그비 변환과 MAF 변환을 적용해도 극단적으로 큰 값이나 작은 값이 많이 나타날 수 있다. 만약 이러한 자료들을 대상으로 순차적 가우시안 시물레이션을 적용할 경우, 정규 점수 변환(normal score transformation)을 적용하여 정규 분포를 강제로 따르게 하는 과정에서 같은 자료값들이 변환 결과 서로 다른 값으로 나타나게 된다. 또한 정규 분포의 대칭 특성으로 인해 매우 작거나 큰 값들이 서로 유사한 공간 패턴을 보인다는 내재적인 가정으로 인해 실제 자료의 공간 패턴과는 다른 양상을 보일 수 있어서 매우 높거나 낮은 값들의 시물레이션이 제대로 수행되지 못할 수도 있다. 이러한 다중 가우시안 확률 모델의 문제점을 고려해서 이 연구에서는 순차적 지시자 시물레이션을 적용하였다. 순차적 지시자 시물레이션에서는 조건부 누적 확률 분포 모델링에 비모수적 접근으로 지시자 크리깅을 이용한다(Goovaerts, 1997; Deutsch and Journel, 1998).

지시자 크리깅에 의한 조건부 누적 확률 분포 모델링을 위해서는 우선 원 자료에 대해 지시자 변환을 수행한 후에 크리깅을 수행하여 특정 임계치보다 작거나 같을 조건부 확률을 계산한다. 지시자 변환은 연속형 자료값을 K 개의 특정 임계치 z_k 에 의해 0과

1의 이진 값을 가지게 하는 변환이다.

$$i(u_\alpha; z_k) = \begin{cases} 1 & \text{if } z(u_\alpha) \leq z_k \quad k=1, \dots, K \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

원 자료의 범위를 이산화시킨 일련의 임계치 z_k 에 대한 조건부 확률은 지시자 정규 크리깅을 통해 얻게 되며, 이로부터 원자료의 조건부 누적 확률 분포를 모델링한다.

$$[i(u; z_k)]^* = [F\{u; z_k | (n)\}]^* = [Prob\{Z(u) \leq z_k | (n)\}]^* = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(u; z_k) i(u_\alpha; z_k) \quad (8)$$

여기서 (n) 은 특정 위치 u 주변의 자료값을 의미하며, $F(\bullet)$ 는 특정 임계치 z_k 에서의 조건부 누적 확률 분포를 나타낸다. λ_α 는 탐색창 내의 샘플 자료에 부여되는 정규 크리깅 가중치를 나타낸다. 그리고 지시자 크리깅을 적용하기 위해서는 K 개의 임계치별 로 지시자 변환된 값에 대한 베리오그램 모델이 필요하다.

순차적 지시자 시물레이션의 실제 적용 절차를 요약하면 다음과 같다. 우선 시물레이션을 수행할 각 예측 위치를 한번씩 순차적으로 방문할 임의의 경로를 정의한다. 그리고 임의의 경로를 따라 개별 예측 위치에서 주변 자료를 이용하여 지시자 크리깅을 적용하여 조건부 누적 확률 분포를 모델링한다. 이 조건부 누적 확률 분포를 대상으로 몬테카를로 시물레이션을 적용하여 임의로 생성한 난수에 대응되는 분위수를 계산한다. 이 분위수값을 자료에 포함시킨 후에 다른 예측 위치를 임의의 경로에 따라 방문하고, 전술한 과정을 모든 예측 위치에서 시물레이션을 마칠 때까지 반복한다. 새로운 시물레이션값은 임의의 경로를 다르게 지정하여 앞의 과정을 반복하여 얻어지며, 미리 정해놓은 시물레이션 회수에 따라 확률적으로 등가인 다량의 시물레이션 결과들을 얻게 된다.

이 연구에서는 MAF 변환된 변수들을 대상으로 순차적 지시자 시물레이션을 적용하기 때문에, 변수들 개별적으로 시물레이션을 적용할 수 있다. 개별 변수들의 지시자 시물레이션 결과는 개별 예측 위치에서의 분위수에 해당하므로, 이 값들을 대상으로 역 MAF 변환을 수행하면 로그비 스케일에서의 해당 분위수를 얻게 되며 마지막으로 식 (3)의 역 가산 로그

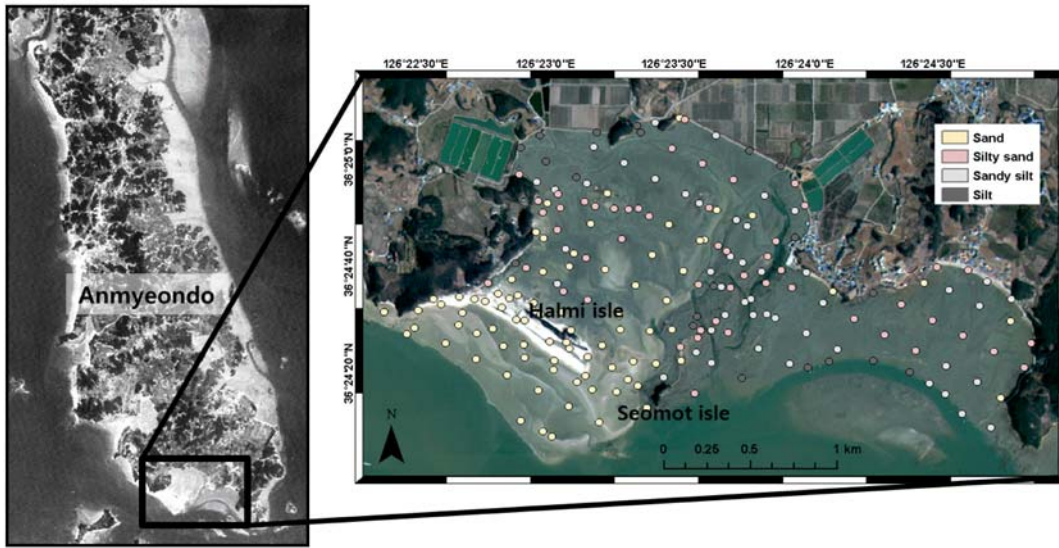


Fig. 2. The location of the sample data and sediment classes by Shepard' classification rule.

비 변환을 적용하면 최종적으로 원 구성 자료 스케일에서의 시뮬레이션 결과를 얻게 된다.

사례연구

대상 지역 및 사용 자료

이 연구에서 제한한 구성 자료의 지구통계학적 시뮬레이션의 적용 절차를 예시하기 위해 간석지 표층 퇴적물 입도 비율 자료를 대상으로 사례연구를 수행하였다. 대상 지역은 충청남도 안면도 남단의 바람아래 해변으로부터 남서쪽으로 발달한 바람아래 간석지로, 사례연구에는 2010년 1월에 수행한 총 215개 지점의 현장 조사 자료를 이용하였다(Fig. 2). 실내 실험에서는 우선 퇴적물 입도 시료에 과산화수소수(15%)와 염산(10%)을 차례로 넣어 유기물과 탄산염을 완전히 제거한 후에 자연 건조를 하였다. 그리고 표준망체(63 μm)를 사용한 습식 분석을 통해 사질과 니질을 분리시켰으며, 분리된 사질 시료는 건조기로 건조후 진탕기를 이용하여 0.5 ϕ 간격으로 체질하여 입도별 중량 백분율을 구하였다. 4 ϕ 이상의 니질 시료는 확산계 용액에 넣고 초음파 자기진동기로 균일하게 분산시킨 후 자동입도분석기(Mastersizer 2000)를 이용하여 입도분석을 실시하였다(Jang et al., 2010; Lee et al., 2012). 이러한 실내 실험을 거쳐 모래, 실트 및 점토의 함량 비율을 얻었고, 이 자료를 이후 지구통계학적 분석에 사용하였다.

Table 1. Summary statistics of original compositional data

| Statistics | Sand | Silt | Clay |
|-------------|--------|---------|---------|
| Minimum | 0.0036 | 0.0001 | 0.0001 |
| Maximum | 0.9998 | 0.9235 | 0.1393 |
| Mean | 0.6488 | 0.3204 | 0.0308 |
| Std. Dev. | 0.3127 | 0.2832 | 0.0316 |
| Correlation | Sand | Silt | Clay |
| Sand | 1.0000 | -0.9993 | -0.9437 |
| Silt | - | 1.0000 | 0.9309 |
| Clay | - | - | 1.0000 |

탐색적 자료 분석 및 베리오그램 모델링 결과

215개 표층 퇴적물의 성분들을 대상으로 가산 로그비 변환을 적용하기 위해서는 함량에 0%가 없어야 한다. 원 성분 자료에서 모래 함량이 매우 높은 지점에서 실트와 점토의 함량이 0%인 지점들이 나타나서 이러한 지점들은 Martin-Fernandez and Thio-Henestrosa (2006)의 제안 방법을 따라 0%를 제외한 가장 작은 함량값인 0.02%의 1/2인 0.01%를 측정의 최소 분해능으로 간주하여 0% 대신 대입시킨 후 전체 합이 100%가 되도록 정규화시켰다. Table 1은 이렇게 전처리를 거친 215개 입도 자료의 각 성분별 요약 통계치를 나타내고 있다. 연구 지역은 모래 성분의 함량이 매우 높으며, 상대적으로 점토의 함량은 낮은 것으로 나타났다. 또한 모래 성분은 실트와 점토 성분과 매우 강한 음의 상관관계를 나타내고, 실트와 점토 성분은 강한 양의 상관관계를 보이는 등 전형

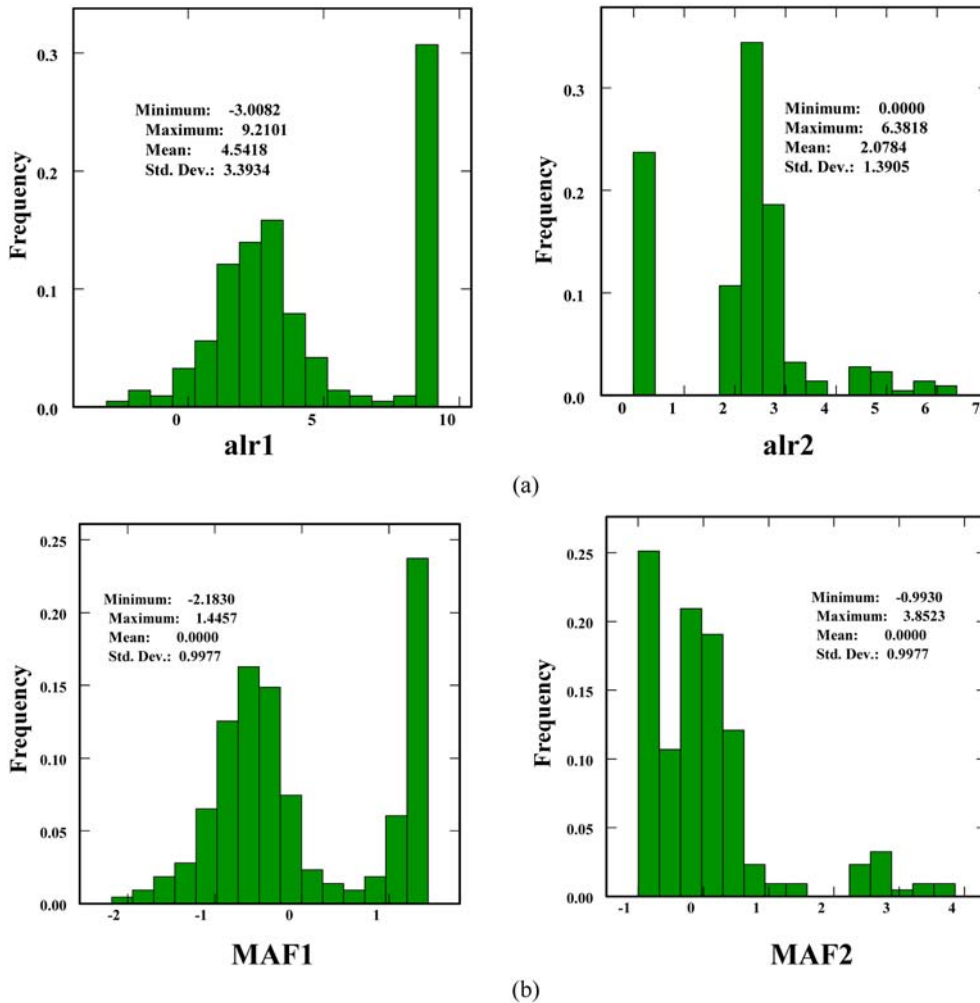


Fig. 3. Histograms and summary statistics of transformed variables: (a) alr transformed variables and (b) MAF transformed variables.

적인 구성 자료의 특성을 보이고 있다. 이러한 상관관계는 전체 합이 100%가 되어야 하는 상대적인 비율 계산에서 유발된 것으로, 상호 비교시에만 제한적인 의미를 가진다.

가산 로그비 변환은 식 (2)를 적용하여 수행하였고, 이 때 분모 성분은 실험적으로 점토 성분으로 하였다. Fig. 3a는 가산 로그비 변환의 적용 결과 얻어진 두 개의 성분들의 히스토그램과 요약 통계값을 나타내고 있다. 두 변수에서 각각 매우 큰 값과 매우 작은 값이 독립적으로 존재하고 있는데, 이 값들은 모두 실트와 점토 성분이 0.01%로 치환된 지점으로 매우 큰 모래 성분을 보인 지점에 해당한다. 이러한 값들이 나타나는 지점은 모두 51개 지점으로 전체

자료의 23.7%에 해당한다. 만약 이 자료들을 다중 가우시안 확률 모델로 가정할 경우, 정규 점수 변환 과정에서 이 51개 지점들의 값이 같음에도 불구하고 다른 값으로 변환되어 실제 매우 높거나 낮은 값의 시뮬레이션 결과는 신뢰성이 떨어질 수 있다.

가산 로그비 변환된 변수들을 대상으로 분산-공분산 행렬을 이용하여 MAF 변환을 수행하였다. 이 때 필요한 이격 거리는 샘플 사이의 거리를 고려하여 120 m로 설정하였다. Fig. 3b는 MAF 변환된 변수들의 히스토그램과 요약 통계치를 나타내고 있다. 가산 로그비 변환된 변수들 사이의 상관계수는 -0.338 이었는데, MAF 변환을 통해 모든 이격 거리에서 상관계수가 0이 되는 독립적인 변수들이 생성된다. 개별

변수들의 히스토그램을 살펴보면 평균 0, 분산 1인 분포를 나타내지만 Fig. 3a와 마찬가지로 여전히 매우 크거나 매우 작은 값의 빈도수가 높은 분포를 보이고 있다. 이러한 탐색적 자료 분석 결과는 해당 자료에 대해 다중 가우시안 확률 모델의 적용이 어렵다는 것을 입증하고 있으며, 순차적 지시자 시뮬레이션의 적용이 이론적으로 타당함을 지시한다.

순차적 지시자 시뮬레이션의 적용을 위해서는 개별 변수별 특정 임계값에 따른 베리오그램 모델이 필요하다. 이 연구에서는 MAF 변환된 변수들을 대상으로 백분위수를 기준으로 9개의 임계값을 설정하였는데, Fig. 3b에서 확인하였듯이 일부 치우친 분포를 나타내어 이를 고려하여 일부 백분위수를 수정한 임계값을 설정하였다(Table 2). 지시자 변환 변수들의 실험적 베리오그램 계산과 베리오그램 모델링에는 Goovaerts (2009)의 AUTO-IK Fortran 프로그램을 이용하였다. Fig. 4는 두 번째 MAF 변수의 지시자 베리오그램들을 나타내고 있다. 임계값과 상관없이 서로 다른 상관거리를 가지는 구형 혹은 지수 모델의 결합으로 모델링하였으며, 임계값이 커질수록 첫 번째 모델의 상관거리가 짧게 나타났다. 이러한 임계치 별로 상이한 상관거리를 보이는 구조는 지시자 변환을 통해 탐지할 수 있는 결과로, 특히 첫 번째 임계치와 마지막 임계치를 기준으로 지시자 베리오그램 모델이 서로 다른 것은 다중 가우시안 확률 모델의 적용이 적절치 않음을 지시한다.

시뮬레이션 결과

순차적 지시자 시뮬레이션에는 2개의 MAF 변환 변수들을 대상으로 각각의 지시자 베리오그램 모델들을 이용하였으며, 실제 적용에는 GSLIB (Deutsch and Journel, 1998)을 이용하였다. 전체 시뮬레이션 개수는 시범적으로 50개로 설정하였으며, 시뮬레이션에 필요한 파라미터들은 지시자 베리오그램 모델을 기반으로 설정하였다. 50개의 MAF 변환 변수들의 시뮬레이션 결과를 얻은 후에는 Fig. 1에 제시한 것처럼 다중 변환의 적용의 역순으로 즉 역 MAF 변환, 역 가산 로그비 변환을 적용하여 최종적으로 원래 구성 자료의 스케일에 해당하는 성분별 50개의 시뮬레이션 결과를 생성하였다.

Fig. 5는 각 성분별 10번째 시뮬레이션 결과를 나타내고 있다. 역 가산 로그비 변환을 통해 모든 위치에서 세 성분들의 합은 100%를 만족한다. 시뮬레이

Table 2. Thresholding values of MAF transformed variables for indicator simulation

| No | MAF1 | | MAF2 | |
|----|--------------------|---------------------------|--------------------|---------------------------|
| | thresholding value | corresponding probability | thresholding value | corresponding probability |
| 1 | -1.069 | 0.1 | -0.993 | 0.2 |
| 2 | -0.813 | 0.2 | -0.719 | 0.25 |
| 3 | -0.653 | 0.3 | -0.477 | 0.3 |
| 4 | -0.519 | 0.4 | -0.258 | 0.4 |
| 5 | -0.392 | 0.5 | -0.142 | 0.5 |
| 6 | -0.146 | 0.6 | 0.029 | 0.6 |
| 7 | 0.936 | 0.7 | 0.147 | 0.7 |
| 8 | 1.192 | 0.75 | 0.391 | 0.8 |
| 9 | 1.446 | 0.8 | 0.922 | 0.9 |

션의 일반적인 특성인 공간 변동성이 크게 나타나고, 크리깅 결과에서 두드러진 평활화 효과가 적음을 확인할 수 있다.

지구통계학적 시뮬레이션의 주된 목적은 샘플 자료의 통계값의 재생산에 있기 때문에, 시뮬레이션 결과의 검증은 크리깅처럼 교차 검증 등을 통한 예측 능력을 제시하지 않고, 이러한 통계값의 재생산 여부를 대상으로 한다. 이 연구에서는 최종 변환된 각 성분별 시뮬레이션 결과들을 대상으로 요약 통계값과 베리오그램을 계산하여 샘플 자료와 비교하였다. Fig. 6은 10번째와 40번째 시뮬레이션 결과와 샘플 자료와의 분위수-분위수 그림을 나타내고 있다. 2번의 자료 변환을 거쳤음에도 불구하고 샘플 자료의 통계값을 잘 재생산하고 있으며, 특히 높거나 낮은 값들에 대한 편향 정도가 심하지 않은 것을 확인할 수 있다.

Fig. 7은 각 성분별 10번째 시뮬레이션 결과의 베리오그램과 샘플 자료의 베리오그램을 비교한 결과이다. 요약 통계치와 마찬가지로 샘플 자료의 베리오그램과 교차 베리오그램 모두를 잘 반영하고 있다. 이러한 요약 통계치와 베리오그램의 비교를 통해 이 연구에서 제안한 시뮬레이션 기법은 두 번의 자료 변환 과정을 거쳤음에도 불구하고 조건부 시뮬레이션의 특성을 잘 반영하고 있음을 확인할 수 있다.

50개의 시뮬레이션 결과들로부터 다양한 통계값 계산이 가능한데, 이 연구에서는 평균과 분산값을 계산하였다. Fig. 8은 50개의 시뮬레이션 결과들로부터 계산한 평균값을 나타내고 있다. 조건부 누적 확률 분포의 일종의 분위수에 해당하는 시뮬레이션 결과들을 평균한 값은 최소 자승의 관점에서 최적의 해를 나타내며, 분위수에 일련의 역변환을 적용하였기 때

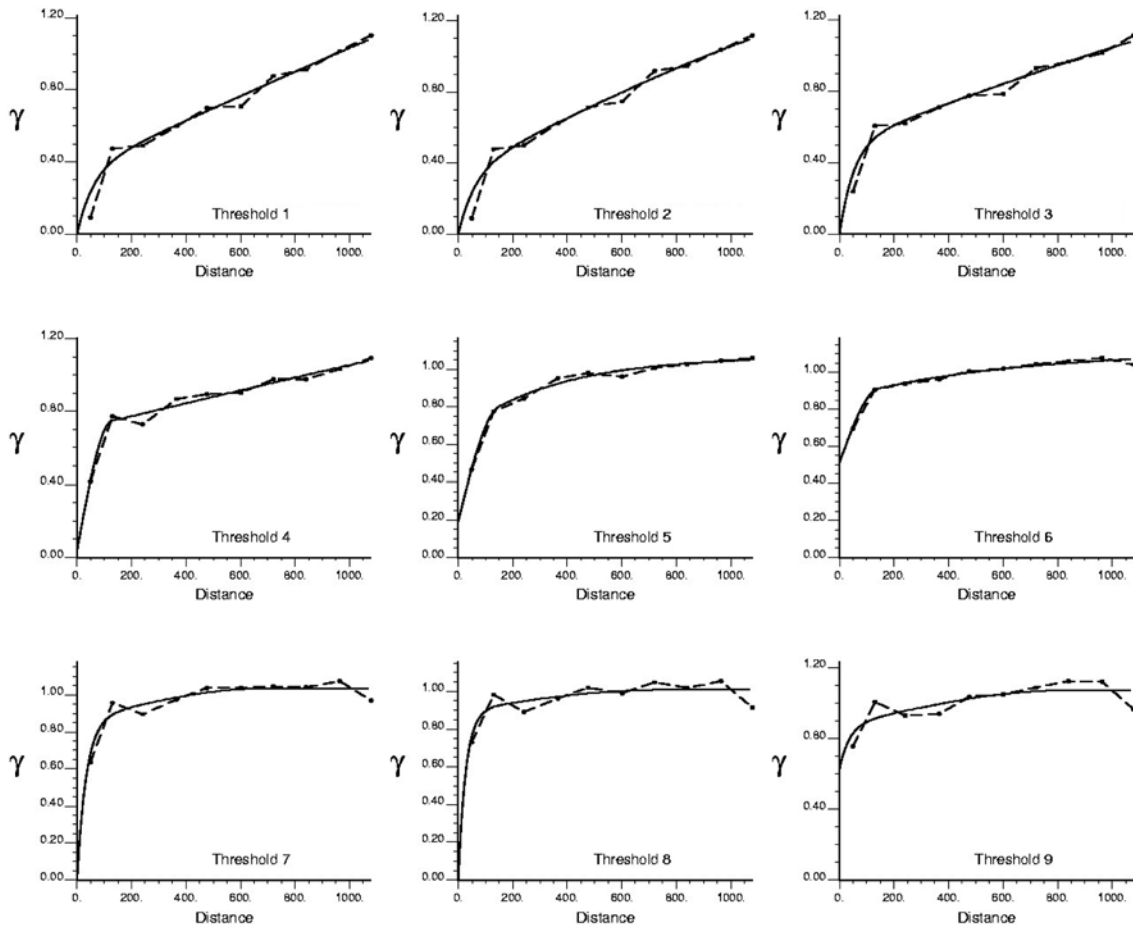


Fig. 4. Indicator variograms of the second MAF transformed variable.

문에 비편향된 예측 결과에 해당한다. 연구지역내 표층 퇴적물 성분들의 공간 분포를 살펴보면(Fig. 8), 모래 성분의 함량이 전반적으로 높으며 연구지역 남서쪽에 위치한 할미섬을 중심으로 주변부에 광범위하게 분포하며, 만입지와 연구지역 동쪽 남단에서는 함량이 상대적으로 낮게 나타나고 있다. 반면 실트와 점토 성분들은 모래 성분과의 음의 상관관계를 보이기 때문에 모래 함량이 높은 지점에서 상대적으로 낮은 함량값을 나타내며, 모래의 함량이 낮은 만입지와 섬웃섬 동쪽 남단에서 상대적으로 높은 함량을 보이고 있다. 또한 실트 성분에 비해 점토 성분은 상대적으로 매우 낮은 함량을 보이며, 함량이 높은 지역은 상대적으로 협소한 특징을 보인다.

평균값 이외에 50개의 시뮬레이션 결과들로부터 분산을 계산하여 불확실성의 척도로 사용하였다. 만약 분산값이 큰 경우에는 50개 시뮬레이션 값들의

차이가 크게 나타난 것이므로 불확실성이 상대적으로 큰을 의미한다. 분산 분포를 살펴보면(Fig. 9), 상대적으로 모래 함량이 높은 지역에서는 모든 성분들에서 분산값이 작게 나타나고 있는데, 이는 50개 시뮬레이션 결과에서 일관되게 값들이 나타남을 지시한다. 이 지역들은 모래 함량이 매우 우세하여 실트와 점토 함량이 0%에 가까운 지역에 해당된다. 반면 연구지역 중앙 부분과 만입지와 연구지역 동편에서 상대적으로 분산값이 높게 나타나고 있는데, 이들 지역은 상이한 샘플 자료값들이 존재하는 지역에 해당되어 자료의 변동성으로 인해 분산이 상대적으로 높게 나타난 것으로 판단된다. 점토 성분은 원래 함량값이 상대적으로 낮은 관계로 분산값의 범위가 매우 작게 나타난다. 이렇게 국소적으로 시뮬레이션 결과들의 분산이 상대적으로 크게 나타나는 지역은 보강 현장 조사가 필요한 지역에 해당한다.

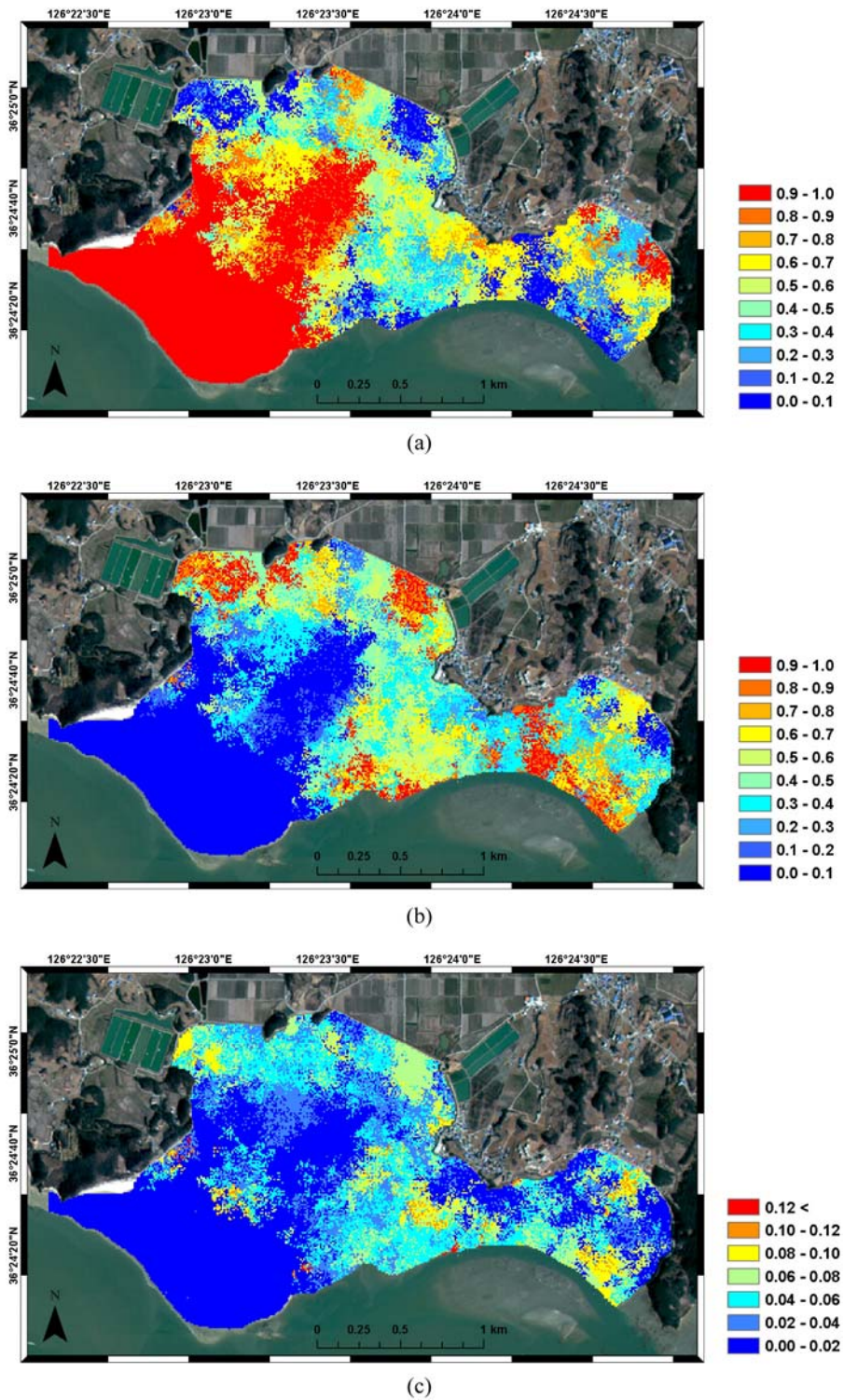


Fig. 5. 10th simulation results: (a) sand, (b) silt, and (c) clay.

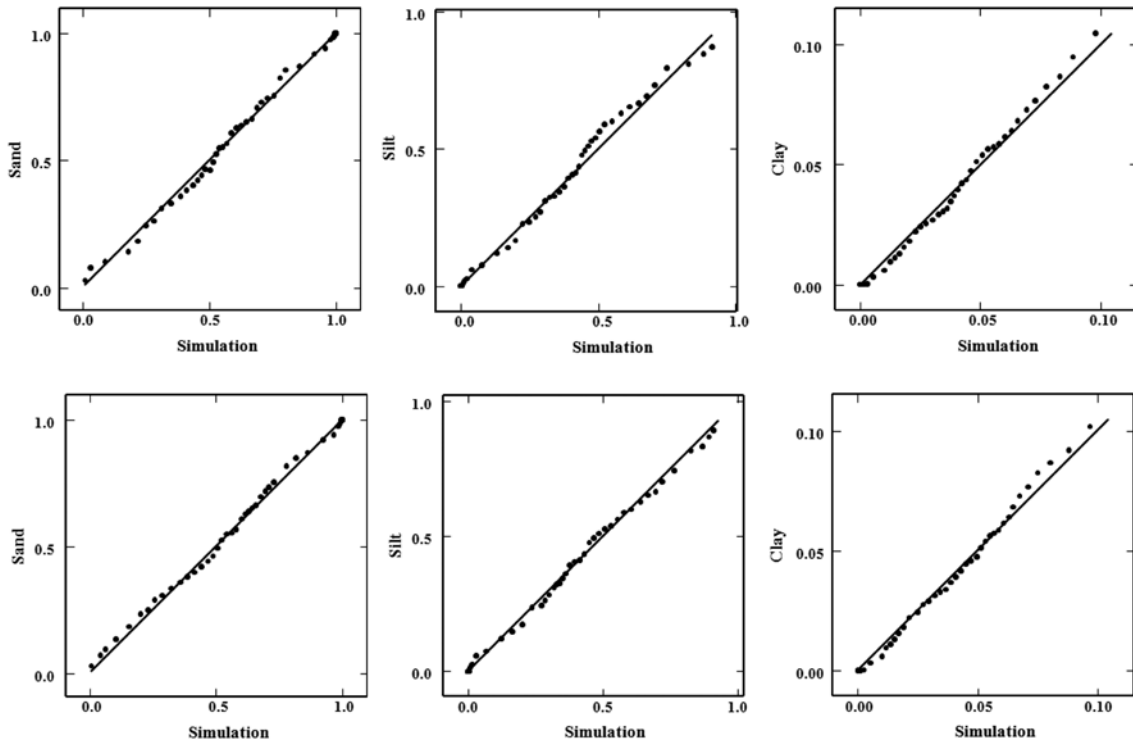


Fig. 6. Quantile-quantile plots of simulations for each component. The first row is for the 10th simulation the second row for the 40th simulation.

여기서 유의할 점은 시뮬레이션의 목적은 여러 시뮬레이션 결과들로부터 특정 시뮬레이션 결과를 선택하거나, 평균과 같은 요약 통계값의 계산에만 있지 않다는 점이다. 즉 개별 시뮬레이션 결과 자체는 확률 함수 관점에서 고려하고 있는 구성 자료의 성분별 미지의 참값을 나타내는 대안적인 결과로 간주할 수 있어 개별적으로 의미가 있다는 점이다. 구성 자료의 공간 예측은 연구지역내 각 성분들의 공간 분포 파악을 위해 수행되지만, 그 자체가 연구의 주된 목적이 아니라기보다는 일반적으로 성분별 공간 예측 결과를 2차 분석의 입력 자료로 사용한다. 표층 퇴적물 성분들은 특정 분류 규칙에 따라 표층 퇴적물 분류에 사용되는데(Poppe et al., 2003), 이 연구에서는 Shepard의 분류 규칙(Shepard, 1954)을 적용하여 퇴적물 분류도를 작성하였다. 만약 크리깅 등의 공간 내삽방법에 의해 퇴적물 분류도를 작성할 경우, 단일 퇴적물 분류도를 얻게 되어 각 성분별 예측에 수반되는 불확실성 혹은 오차가 퇴적물 분류에 미치는 영향을 분석할 수 없다. 반면 이 연구에서는 각 성분별 시뮬레이션 결과들을 생성했기 때문에, 시뮬레이

션 개수만큼 퇴적물 분류도를 얻을 수 있으며 이를 통해 퇴적물 분류의 불확실성을 정량적으로 표현할 수 있다. 이 연구에서는 개별적인 시뮬레이션 결과들을 대상으로 Shepard의 분류 규칙을 적용하여 50개의 퇴적물 분류 결과를 생성하였으며, Fig. 10은 10번째와 40번째 시뮬레이션 결과로부터 생성된 퇴적물 분류 결과를 나타내고 있다. 여기서 유의할 점은 시뮬레이션의 목적은 시뮬레이션된 퇴적물 분류 결과 중에서 실제 퇴적물 분포를 가장 잘 반영하는 시뮬레이션 결과를 찾는 것이 아니며, 그 결과물들이 확률론적으로 등가(equi-probable)인 미지의 퇴적물 분포를 나타내기 때문에 결과물 그 자체가 의미가 있다는 것이다.

연구지역내 모든 위치에서 50개의 퇴적물 분류결과로부터 특정 퇴적물로 분류되는 회수를 카운팅하여 확률 분포도를 작성하였다. Fig. 11은 연구지역 내 존재하는 4개 퇴적물에 대한 확률 분포도를 나타내는데, 이 확률 분포도는 연구지역내에서 특정 퇴적물이 우세하게 나타날 확률을 나타낸다. 연구지역내 존재하는 4개 퇴적물 중에서, 모래 성분이 75%이상인

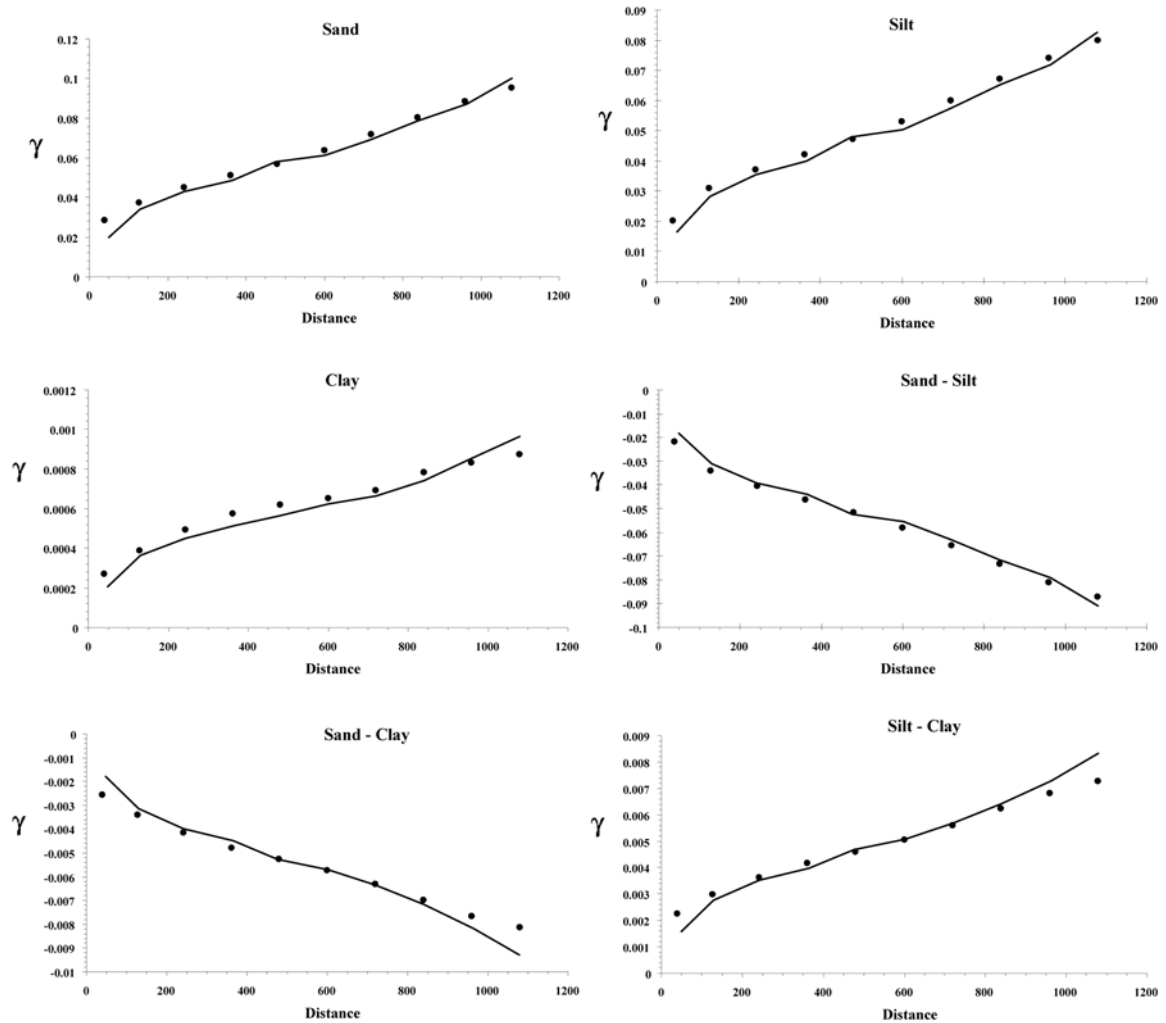


Fig. 7. Comparison of variograms of sample data and the 10th simulation. Lines and dots denote the experimental variogram of sample data and that of the simulation result, respectively.

경우에 해당하는 모래질 퇴적물은 연구지역의 북서쪽에 광범위하게 확률값이 매우 높게 나타나므로, 이 지역은 확실하게 모래질 퇴적물로 분류됨을 지시한다. 반면, 다른 퇴적물들은 분류 확률값이 상대적으로 낮으며 분포 범위 역시 상대적으로 좁게 나타나고 있다. 실트 성분이 75% 이상인 실트 퇴적물은 만입지 일부와 연구지역의 남동쪽에 우세하게 존재하는 것으로 나타났다.

Fig. 11에 제시한 각 퇴적물의 분류 확률로부터 최대 사후 확률 규칙을 적용하여 연구지역의 최종 퇴적물 분류도를 작성하였다(Fig. 12). 즉 특정 위치에서 개별 퇴적물에 속할 확률이 가장 큰 퇴적물로 할

당하는 과정을 적용하였다. Fig. 11에서 살펴본 것처럼, 연구지역의 북서쪽과 만입지 북쪽 일부 지역에는 모래질 퇴적물로 분류되며, 만입지로 갈수록 실트질 모래 퇴적물로 전이되었다가 만입지에서는 실트질 퇴적물이 나타나고 있다. 섬웃섬 동편과 가경주 마을 동쪽에는 모래질 실트 퇴적물이 넓게 나타나며, 바다에 인접한 연구지역 동쪽 남단과 만 안쪽의 만입지에서는 미립질인 실트질 퇴적물이 존재하고 있다. 이러한 다양한 2차 분석 결과는 시뮬레이션을 적용하였기 때문에 얻을 수 있는 부가 정보에 해당하며, 크리깅과 대별되는 시뮬레이션의 장점과 동시에 제안 시뮬레이션 기법의 유용함을 나타낸다.

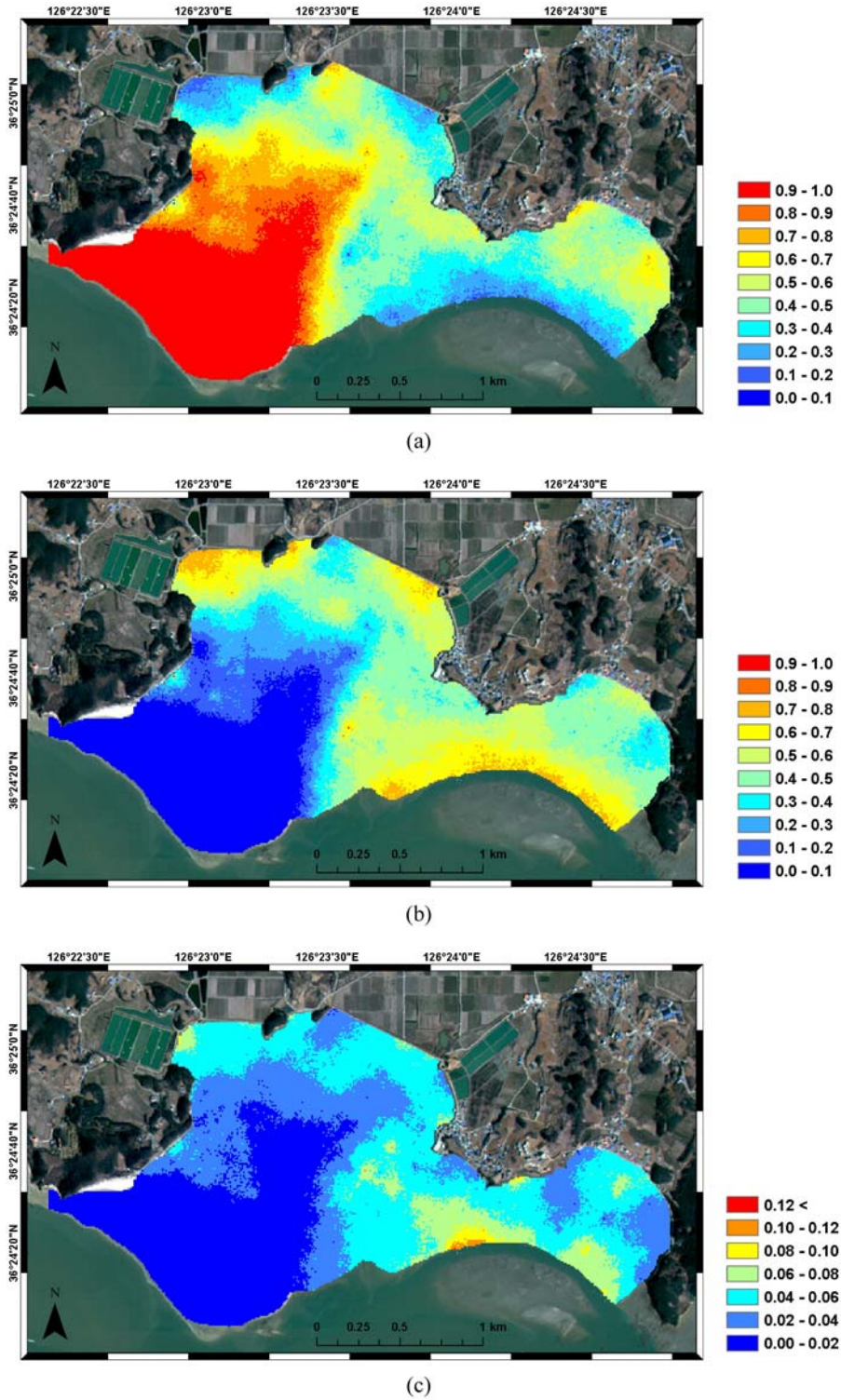


Fig. 8. Average values of 50 simulation results: (a) sand, (b) silt, and (c) clay.

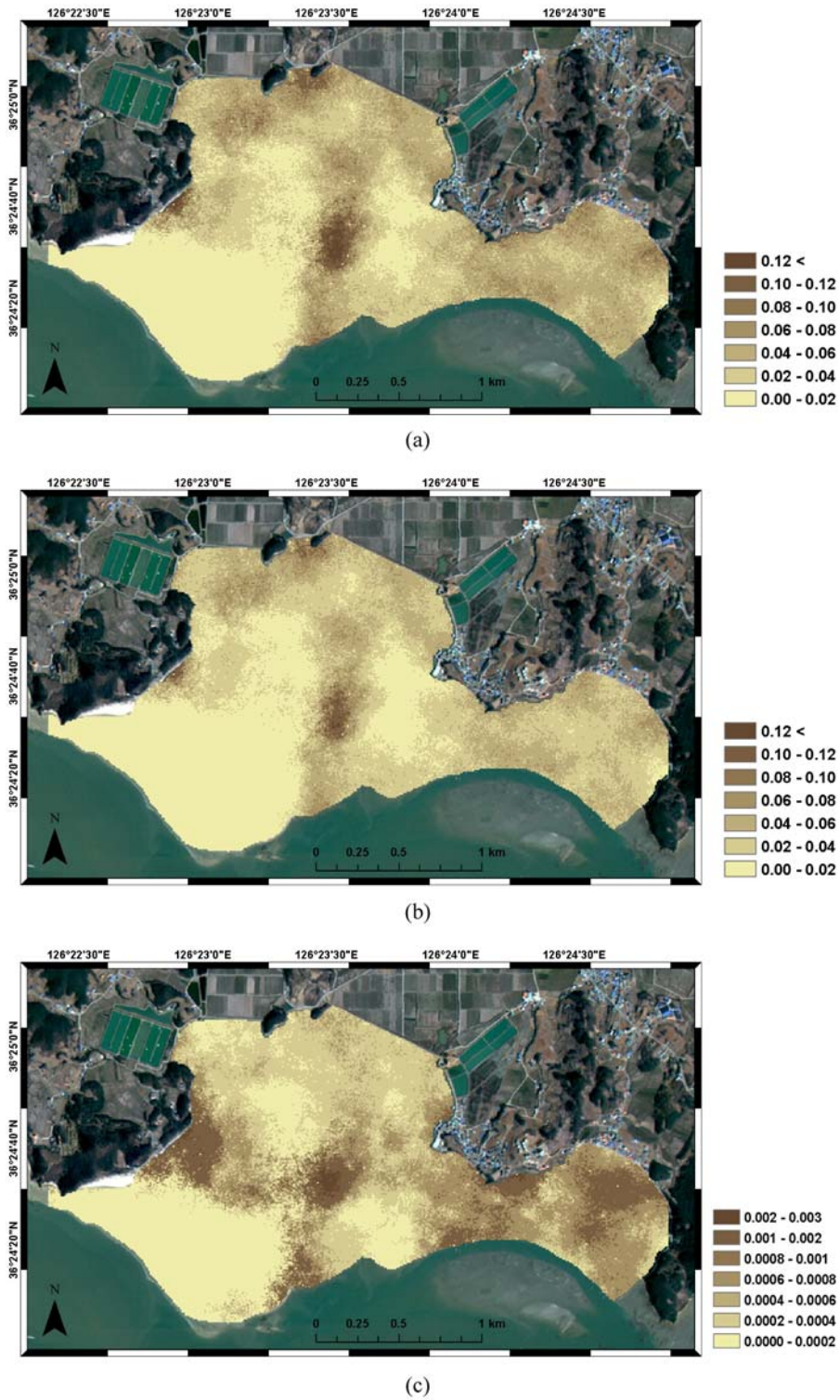


Fig. 9. Conditional variance values of 50 simulation results: (a) sand, (b) silt, and (c) clay.

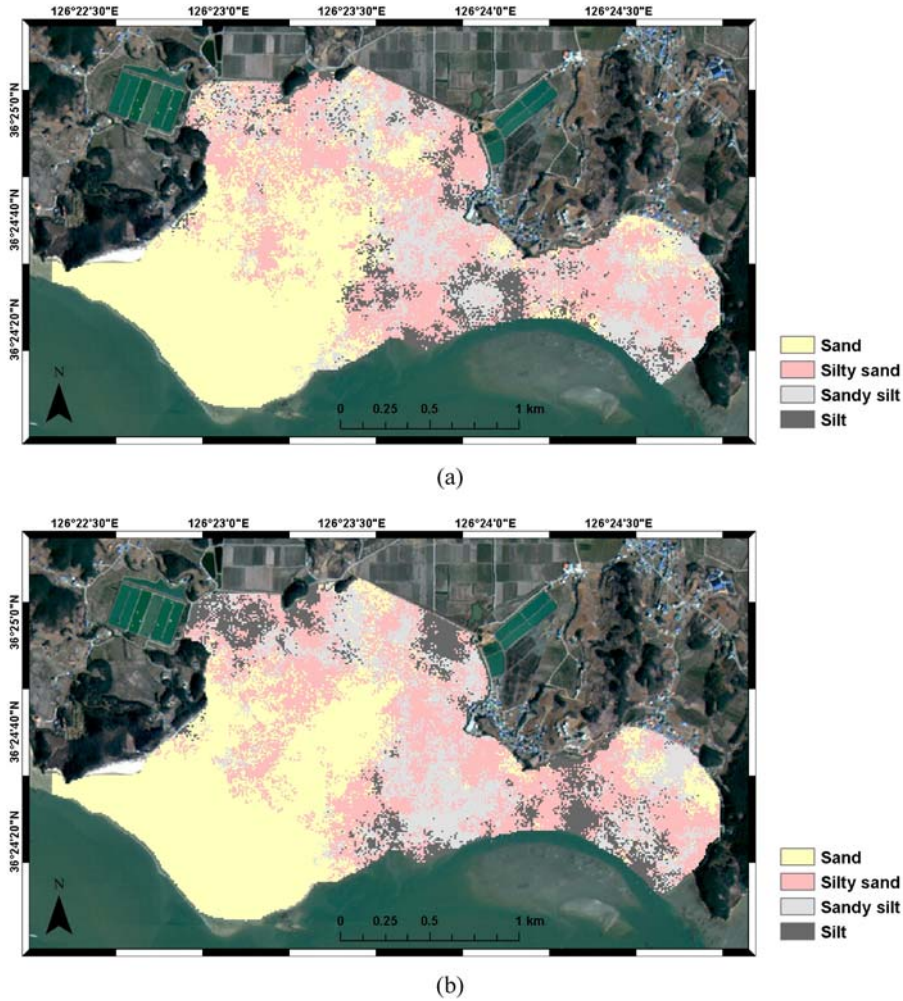


Fig. 10. Surface sediment classification results: (a) 10th simulation and (b) 40th simulation.

결론

토양 혹은 퇴적물의 성분 자료는 특정 제한조건을 만족해야하는 구성 자료에 해당하며, 이러한 구성 자료의 공간 예측에는 구성 자료에 특화된 자료 처리 방식을 적용해야 한다. 이 연구에서는 구성 자료의 지구통계학적 시뮬레이션을 목적으로 일련의 다중 자료 변환을 적용하는 시뮬레이션 틀을 제안하였다. 우선 simplex 공간내에 위치하는 구성 자료의 특성을 고려하여 일반적인 다변량 통계 분석 기법의 적용을 위해 가산 로그비 변환을 적용하였다. 가산 로그비 변환된 변수들을 대상으로 조건부 시뮬레이션을 적용하고 역 변환을 수행하면 구성 자료의 제한 조건을 모두 만족시킬 수 있다. 두 번째 고려한 자료 변환은

MAF 변환으로, 가산 로그비 변환된 변수들 사이에 상관성이 존재할 경우 공동 시뮬레이션을 적용시 요구되는 계산량을 완화하기 위해 적용하였다. MAF 변환이 적용된 변수들은 서로 상관성이 없어지므로, 독립적인 시뮬레이션의 적용이 가능하다. 세 번째 고려한 자료 변환은 지시자 변환으로 MAF 변환된 변수들이 다중 가우시안 확률 모델을 따르지 않는 경우를 고려해서 적용하였다. 지시자 변환이전에 MAF 변환을 적용하지 않을 경우, 서로 다른 지시자 변환 변수들의 임계치별로 공동 크리깅을 적용해야하는 번거로움이 있다. 따라서 지시자 변환을 조건부 누적 확률 분포 모델링에 사용하기 위해 두 번째 자료 변환인 MAF 변환은 반드시 수행해야한다. 이러한 자료 변환을 기반으로 조건부 시뮬레이션 기법으로 순

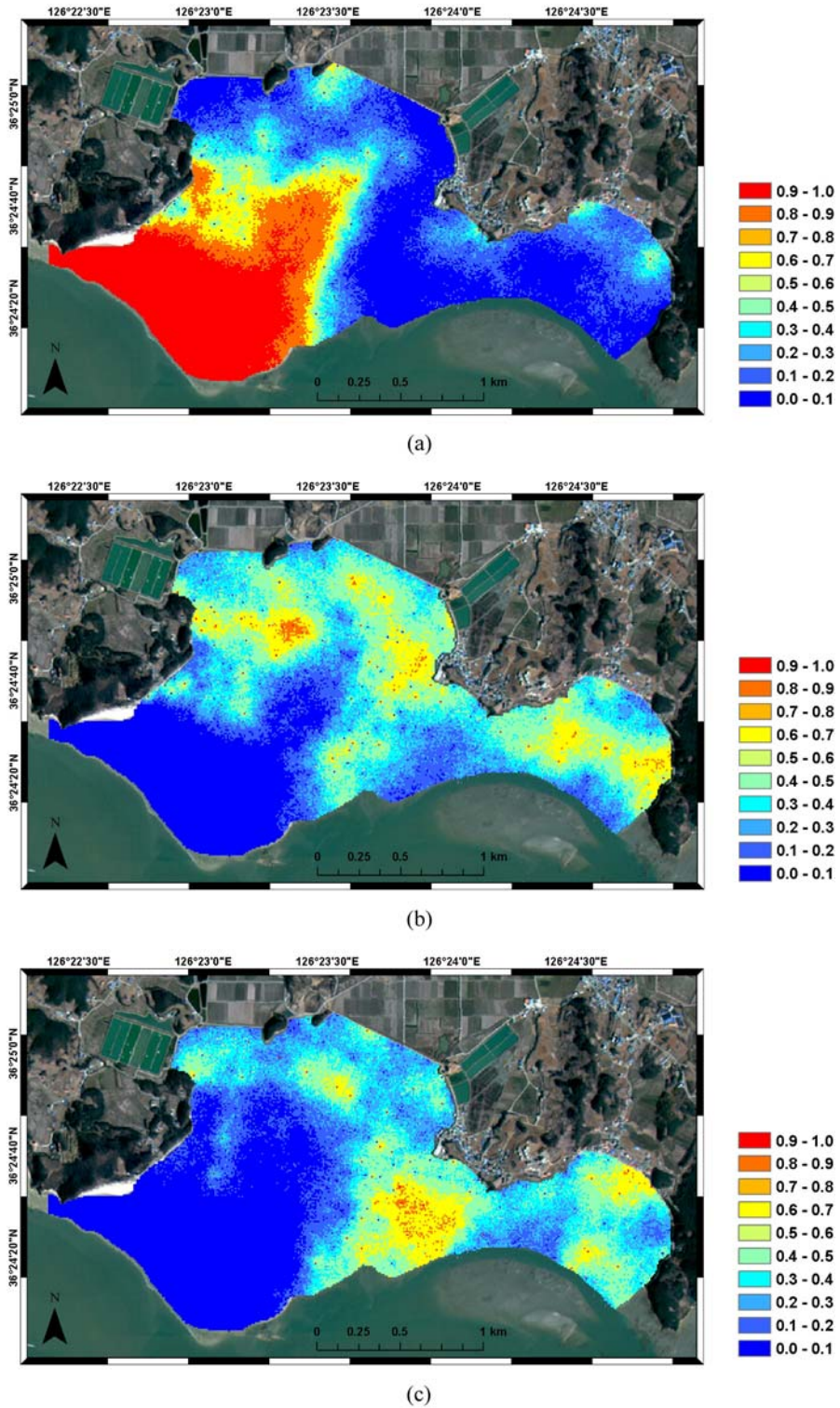


Fig. 11. Conditional probability that a certain sediment class prevails: (a) sand, (b) sandy silt, (c) silty sand, and (d) silt.

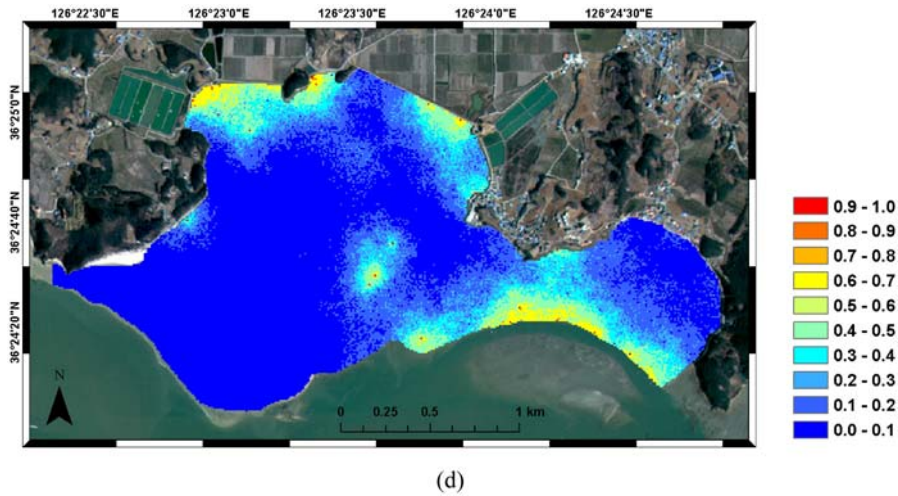


Fig. 11. Continued.

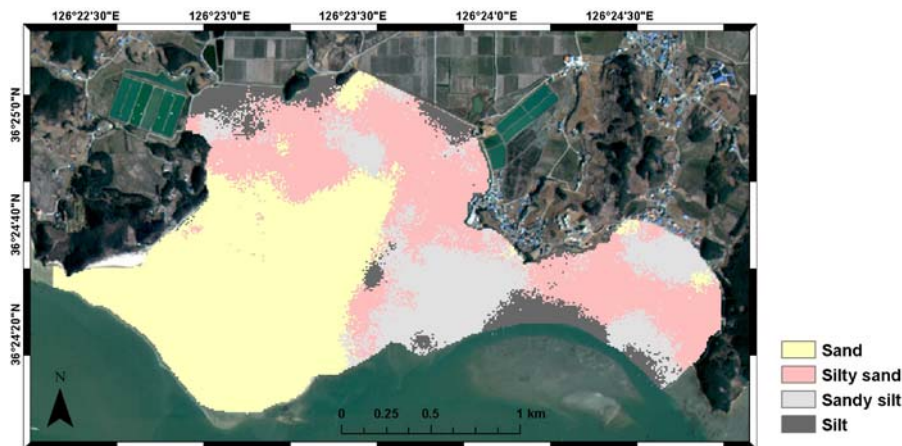


Fig. 12. Surface sediment class with the maximum probability.

차적 지시자 시뮬레이션을 적용하였고, 이후 자료 변환의 역순으로 역 MAF 변환과 역 가산 로그비 변환을 적용하면 다수의 원 구성 자료의 성분별 시뮬레이션 결과를 얻을 수 있다. 특히 탐지 분해능의 한계로 구성 자료에 대부분 포함되는 탐지 한계 이하의 값들을 고려할 때, 다중 가우시안 확률 모델에 비해 계산량은 많더라도 비모수적 지시자 접근이 더 적합하다.

간석지 표층 퇴적물의 성분 자료를 대상으로 제안 기법의 적용성을 평가한 결과, 제안 기법에 의한 시뮬레이션 결과들은 지구통계학적 시뮬레이션의 특징인 샘플 자료의 통계값과 공간 패턴을 잘 재생산하면서 구성 자료의 제한조건을 모두 만족하는 것을

확인할 수 있었다. 시뮬레이션 결과들은 각 성분별 일종의 분위수에 해당하므로 이들로부터 계산되는 시뮬레이션의 평균값은 비편향된 성분 분포도를 나타낼 수 있다. 또한 시뮬레이션은 크리깅과 같이 단일 주제를 생성하는 것이 아니라 다수의 대안적 주제도 생성이 가능하므로, 이후 성분 자료를 입력으로 하는 2차 분석인 표층 퇴적물 분류에 미치는 영향을 확률론적으로 표현이 가능하다. 즉 다수의 성분 자료를 대상으로 특정 분류 체계에 따라 표층 퇴적물 분류를 수행할 경우, 시뮬레이션 회수만큼 다수의 표층 퇴적물 분류 결과를 얻을 수 있으며 이를 통해 표층 퇴적물 분류 결과의 불확실성을 정량적으로 표현할 수 있었다.

이 연구에서는 로그비 변환 및 MAF 변환 자료가 다중 가우시안 확률 모델을 따르지 않는 경우를 고려해서 순차적 지시자 시물레이션을 적용하였는데, 지시자 접근의 특성상 다수의 베리오그램 모델링을 요구한다. 직접 순차적 시물레이션(direct sequential simulation)은 지시자 변환 등을 수행하지 않고 원 자료를 대상으로 시물레이션을 수행하기 때문에 상대적으로 적용이 간편하다. 다만 이 시물레이션 기법은 입력으로 사용한 베리오그램 모델을 재생산할 수 있지만 히스토그램과 같은 요약 통계값을 완전히 재생산하지 못하는 단점이 있다. 그러나 Oz et al. (2003) 이 제안한 히스토그램 재생산이 가능한 직접 순차적 시물레이션의 이용이 가능하므로, 추후 순차적 지시자 시물레이션 대신 이 기법의 적용성을 평가할 예정이다. 이 연구에서는 구성 자료 자체만을 이용하여 시물레이션을 적용하였는데, 구성 자료와 상관성이 존재하면서 보다 많은 위치에서 자료가 획득된 부가 자료의 이용이 가능하다면 보다 신뢰성 높은 시물레이션이 가능하다. 표층 퇴적물 구성 자료의 경우에는 고해상도 원격탐사 자료가 효과적으로 적용될 수 있는데(Lee et al., 2012), 다중 자료 변환을 거친 상태에서 다변량 크리깅을 이용하여 조건부 누적 확률 분포의 모델링 과정에 이러한 부가 자료를 통합할 수 있다. 앞으로 이 연구에서 제안한 단변량 구성 자료의 시물레이션 기법을 다변량 시물레이션 기법으로 확장한 연구를 수행할 예정이다.

감사의 글

이 연구는 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업 지원(NRF-2012 R1A1A1005024)을 받아 수행되었으며, 부분적으로 인하대학교의 지원을 받았습니다. 이 연구에 사용된 현장 조사 자료를 제공해 주신 공주대학교 장동호 교수님께 감사드립니다. 그리고 논문의 표현을 명확히 하는데 도움을 주신 익명의 심사자분들과 강원대학교 오석훈 교수님께 감사드립니다.

References

- Aitchison, J., 1986, The statistical analysis of compositional data. Chapman and Hall, London, UK, 416 p.
- Boucher, A. and Dimitrakopoulos, R., 2009, Block-support simulation of multiple correlated variables. *Mathematical Geosciences*, 41, 215-237.
- Buttafuoco, G., Conforti, M., Aucelli, P.P.C., Robustelli, G., and Scarciglia, F., 2012, Assessing spatial uncertainty in mapping soil erodibility factor using geostatistical stochastic simulation. *Environmental Earth Sciences*, 66, 1111-1125.
- Chilès, J.-P. and Delfiner, P., 2012, *Geostatistics: Modeling spatial uncertainty*. Wiley, New York, USA, 734 p.
- Davis, J.C., 2002, *Statistics and data analysis in geology*. Wiley, New York, USA, 656 p.
- Desbarats, A.J. and Dimitrakopoulos, R., 2000, Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. *Mathematical Geology*, 32, 919-942.
- Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, USA, 369 p.
- Egozcue, J.J., Pawłowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C., 2003, Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35, 279-300.
- Emery, X. and Gálvez, I., 2012, A plurigaussian model for simulating regionalized compositions. In Abrahamsen, P., Hauge, R., and Kolbjørnsen, O. (eds.), *Geostatistics Oslo 2012*. Springer, New York, USA, 39-50.
- Goovaerts, P., 1993, Spatial orthogonality of the principal components computed from coregionalized variables. *Mathematical Geology*, 25, 281-302.
- Goovaerts, P., 1997, *Geostatistics for natural resources evaluation*. Oxford University Press, New York, USA, 483 p.
- Goovaerts, P., 2009, AUTO-IK: A 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. *Computers & Geosciences*, 35, 1255-1270.
- Goovaerts, P., 2010, Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. *Mathematical Geosciences*, 42, 535-554.
- Goovaerts, P., Trinh, H.T., Demond, A., Franzblau, A., Garabrant, D., Gillespie, B., Epkowski, J., and Adriaens, P., 2008, Geostatistical modeling of the spatial distribution of soil dioxins in the vicinity of an incinerator. 1. theory and application to Midland, Michigan. *Environmental Science & Technology*, 42, 3648-3654.
- Jang, D.-H., Kim, J.-S., and Park, N.-W., 2010, Characteristics variation of the sedimentary environment in winter season around the Baramarae beach of Anmyondo using surface sediment analysis. *Journal of the Korean Geomorphological Association*, 17, 15-27. (in Korean)
- Kyriakidis, P.C. and Dungan, J.L., 2001, A geostatistical

- approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environmental and Ecological Statistics*, 8, 311-330.
- Kyriakidis, P.C., Miller, N.L., and Kim, J., 2004, A spatial time series framework for simulating daily precipitation at regional scales. *Journal of Hydrology*, 297, 236-255.
- Lark, R.M. and Bishop, T.F.A., 2007, Cokriging particle size fractions of the soil. *European Journal of Soil Science*, 58, 763-774.
- Lark, R.M., Dove, D., Green, S.L., Richardson, A.E., Stewart, H., and Stevenson, A., 2012, Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. *Sedimentary Geology*, 281, 35-49.
- Lee, S.-W., Park, N.-W., Jang, D.-H., Yoo, H.Y., and Lim, H., 2012, Surface sediments classification in tidal flats using multivariate kriging and KOMPSAT-2 imagery. *Journal of the Korean Geomorphological Association*, 19, 37-49.
- Martin-Fernandez, J.A. and Thio-Henestrosa, S., 2006, Rounded zeros: Some practical aspects for compositional data. In Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional data analysis in the geosciences: From theory to practice*. Geological Society of London, UK, 191-201.
- Odeh, I.O.A., Todd, A.J., and Triantafyllis, J., 2003, Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, 168, 501-515.
- Oh, J.-K. and Kum, B.-C., 2001, Depositional environments and characteristics of surface sediments in the nearshore and offshore the mid-western coast of the Korean peninsula. *Journal of the Korean Earth Science Society*, 22, 377-387. (in Korean)
- Oh, S., 2005, RMR evaluation by integration of geophysical and borehole data using non-linear indicator transform and 3D kriging. *Journal of the Korean Earth Science Society*, 26, 429-435. (in Korean)
- Oh, S. and Han, S.-M., 2010, Downscaling of geophysical data for enhanced resolution by geostatistical approach. *Journal of the Korean Earth Science Society*, 31, 681-690. (in Korean)
- Oz, B., Deutsch, C.V., Tran, T.T., and Xie, Y., 2003, DSSIM-HR: A FORTRAN 90 program for direct sequential simulation with histogram reproduction. *Computers & Geosciences*, 29, 39-51.
- Park, N.-W., 2010, Exemplifying the potential of indicator geostatistics for probabilistic uncertainty and risk analyses of geochemical data. *Journal of the Korean Earth Science Society*, 31, 301-312. (in Korean)
- Park, N.-W., 2011, Time-series mapping and uncertainty modeling of environmental variables: A case study of PM10 concentration mapping. *Journal of the Korean Earth Science Society*, 32, 249-264. (in Korean)
- Park, N.-W. and Jang, D.-H., 2014, Comparison of geostatistical kriging algorithms for intertidal surface sediment facies mapping with grain size data. *The Scientific World Journal*, (in press).
- Park, N.-W. and Oh, S., 2006, Application of geostatistical simulation to assessment of the effects of uncertainty of spatial data in mineral potential prediction. *Journal of the Korean Society of Mineral and Energy Resources Engineers*, 43, 213-223. (in Korean)
- Pawlowsky-Glahn, V. and Buccianti, A., 2011, *Compositional data analysis: Theory and applications*. Wiley, Chichester, UK, 400 p.
- Pawlowsky-Glahn, V. and Olea, A., 2004, *Geostatistical analysis of compositional data*. Oxford University Press, New York, USA, 304 p.
- Poppe, L.J., Eliason, A.H., and Hastings, M.E., 2003, A visual basic program to classify sediments based on gravel-sand-silt-clay ratios. *Computers & Geosciences*, 29, 805-809.
- Rondon, O., 2012, Teaching aid: Minimum/maximum autocorrelation factors for joint simulation of attributes. *Mathematical Geosciences*, 44, 469-504.
- Saito, H. and Goovaerts, P., 2003, Selective remediation of contaminated sites using a two-level multiphase strategy and geostatistics. *Environmental Science & Technology*, 37, 1912-1918.
- Shepard, F.P., 1954, Nomenclature based on sand-silt-clay ratios. *Journal of Sedimentary Petrology*, 24, 151-158.
- Shin, D.-H., Kum, B.-C., Park, E.Y., Lee, H.-I., and Oh, J.-K., 2004, Seasonal sedimentary characteristics and depositional environments after the construction of seawall on the Iwon macrotidal flat. *Journal of the Korean Earth Science Society*, 25, 615-628. (in Korean)
- Switzer, P. and Green, A.A., 1984, Min/Max autocorrelation factors for multivariate spatial imagery. Technical Report No. 6, Department of Statistics, Stanford University, Stanford, USA, 14 p.
- Vargas-Guzmán, A. and Dimitrakopoulos, R., 2003, Computational properties of min/max autocorrelation factors. *Computers & Geosciences*, 29, 715-723.
- Wang, G., Gerner, G.Z., Fang, S., and Anderson, A.B., 2003, Mapping multiple variables for predicting soil loss by joint sequential co-simulation with TM images and slope map. *Photogrammetric Engineering & Remote Sensing*, 69, 889-898.

2014년 1월 29일 접수

2014년 2월 7일 수정원고 접수

2014년 2월 12일 채택