한국지하수토양환경학회지 Vol. 10, No. 5, pp. 52~60, 2005 ≪原著≫

수정 연쇄 말콥체인을 이용한 2차원 공간의 추계론적 예측기법의 개발

박 은 규*
경북대학교 지질학과

A Development of Generalized Coupled Markov Chain Model for Stochastic Prediction on Two-Dimensional Space

Eungyu Park*
Department of Geology, Kyungpook National University

ABSTRACT

The conceptual model of under-sampled study area will include a great amount of uncertainty. In this study, we investigate the applicability of Markov chain model in a spatial domain as a tool for minimizing the uncertainty arose from the lack of data. A new formulation is developed to generalize the previous two-dimensional coupled Markov chain model, which has more versatility to fit any computational sequence. Furthermore, the computational algorithm is improved to utilize more conditioning information and reduce the artifacts, such as the artificial parcel inclination, caused by sequential computation. A generalized 2D coupled Markov chain (GCMC) is tested through applying a hypothetical soil map to evaluate the appropriateness as a substituting model for conventional geostatistical models. Comparing to sequential indicator model (SIS), the simulation results from GCMC shows lower entropy at the boundaries of indicators which is closer to real soil maps. For under-sampled indicators, however, GCMC under-estimates the presence of the indicators, which is a common aspect of all other geostatistical models. To improve this under-estimation, further study on data fusion (or assimilation) inclusion in the GCMC is required.

Key words : Generalized coupled Markov chain (GCMC), Soil map, Geostatistics

요 약 문

본 연구에서는 기존 연쇄 말콥체인(Coupled Markov Chain, CMC) 확률식의 연산 경직성을 개선하기 위하여 일반화된 2차원 연쇄 말콥체인(Generalized Coupled Markov Chain, GCMC) 확률식이 개발되었다. 또한 개발된 확률식에 근거하여 평면상에서 무작위적으로 분포하는 참조정보를 효율적으로 활용하는 연산 알고리듬이 개발되었다. 개발된 모델은 대안적 지구통계 기법으로의 새로운 가능성을 제시한다. 본 연구를 통해 새롭게 개발된 GCMC 확률식은 기존 CMC 확률식에 비해 보다 유연한 참조 정보 활용 가능성을 가지며 특수한 경우로 기존 CMC 확률식이 유도되었다. 또한 순차적 연산의 인위적 오류 발생 가능성 및 실제 야외 데이터의 낮은 빈도를 고려하여 무작위로 추출된 위치에서 각 범위를 이용한 연산 알고리듬이 제안되었다. 개발된 모델은 가상의 2차원 토양도에 적용 되었으며 기존 지구통계 기법인 SIS에 비하여 손색이 없는 새로운 지구통계 기법으로 토양 및 지질을 포함한 다양한 예측에 이용될 수 있는 가능성을 보였다. 낮은 빈도로 샘플링 된 지시자에 대해서는 기존 지구통계 기법과 마찬가지로 저평가되는 현상을 보였으며 이를 보완하기 위하여 다양한 소스의 데이터 융합 등을 바탕으로 한 계속적인 연구가 요구된다.

주제어 : GCMC, 말콥체인, 지구통계, 토양도

*Corresponding author : egpark@knu.ac.kr
원고접수일 : 2005. 5. 9 게재승인일 : 2005. 6. 16
질의 및 토의 : 2005. 12. 31 까지

1. 서 론

지하의 불균질성은 불연속적 경계에 의해 규정되는 개개의 지하구조 및 이들의 무작위적 공간배열에 기인하는 복잡한 상호관계로 특징지어진다. 이는 지하를 대상으로 하는 모든 연구 및 실용 분야를 막론하여 가장 큰 관심사이며 동시에 장애가 되어왔다. 지하 불균질성에 대한 근본적인 규명은 현대 과학기술의 비약적인 발전에도 불구하고 요원한 과제임이 분명하다. 지하의 불균질적 특성과 함께 지하 데이터 획득의 한계성은 지하구조 예측과 이에 근거한 제반 연구에 있어 더욱 큰 불확실성을 내포하도록 한다.

데이터 획득의 한계를 극복하고 우리가 모르는 지하정보에 대해 정확하게 예측하는 것은 지하를 대상으로 하는 모든 분야에서 공통적으로 요구되는 사항이다. 불확실성은 결정론적 방법으로는 다루기 힘든 문제이다. 결정론적 방법은 예측의 불확실성에 대한 정보를 배제하므로, 위험요소를 다루는 예측모델 또는 자원광물 분포에 따른 경제성 분석 등에 이용하기에는 부적합할 경우가 많다. 추계론적 예측 모델의 적용을 통해 이러한 한계성을 극복할 수 있으며 현재로서는 위험요소에 대한 예측 모델 중 가장 효과적인 접근 방법이라 할 수 있다. 이러한 추계론적 예측에 내포된 기본 가정은 우리가 얻지 못한 지하 데이터에 기인하는 불확실성은 야외로부터 측정한 데이터의 통계적 분포에 의해 예측될 수 있다는 것이다. 그러나 추계론적 방법의 한계는 실제 우리가 얻는 데이터의 제한성 및 통계적 처리에 있어 지나친 간결화 그리고 컴퓨터 자원의 제한성이라 할 수 있다(Li et al., 2003).

전통적 지구통계학에서 이용되어 온 대표적 추계론 예측모델로는 순차 가우스 모사법(sequential Gaussian method, SGS)과 순차 지시자 모사법(sequential indicator method, SIS)을 들 수 있다(Deutsch and Journel, 1992). SGS 방법은 우리가 대상으로 하는 야외변수의 분포가 근원적 정상분포(intrinsic stationary)라는 가정으로부터 출발한다. 이는 야외에서 측정된 값들을 대표하는 값이 존재하며 공간적으로 일정 거리에 떨어져 있는 두 위치에서의 변수 분산은 세미베리오그램(semi-variogram)을 통해 설명될 수 있다고 가정한다. 그러나 이러한 가정은 실제 야외에서 관측되는 현상에 비해 지극히 이상적이라는 점에서 한계성을 가지며 불연속적 변화를 보이는 토양이나 지질현상에 부합하지 않는다. 이를 보완하기 위한 모델이 SIS로, 이 모델은 기본적으로 야외 변수의 분포가 비정상(non-stationary)이라는 관점을 가진다. 또한 SGS와 달리 지시자에 의존함으로써 야외변수의 분포가 연속적이라는 가정을 하지 않기 때문에 상호 배타적일 수 있는 비수치적 지질 및 토양 현상에 좋은 적용성을 보인다. 그러나 SIS 방법 역시 다양한 지시자 간의 코크리깅 (cokriging)을 하는데 있어 수학적 복잡성의 문제를 지니고 있으며 평면상의 데이터가 적은 경우에는 세미베리오그램을 구성하는데 한계성을 보인다. 또한 기존 지구통계기법의 근간인 세미베리오그램에 의존하기 때문에 지질현상에서 많이 관찰되는 지향적 이방성(directional anisotropy)을 예측하는 데에는 한계가 있다(Carle and Fogg, 1996; Carle and Fogg, 1997; Carle et al., 1998).

기존 지구통계기법의 문제점을 극복하기 위하여 최근에 대두되고 있는 방법은 지질 지시자의 전이확률(transition probability)에 근거하는 Markov Chain(Markov Chain) 모델이다. Markov Chain 모델은 과거 퇴적학 분야에서 많이 활용되었고 공간적으로는 1차원적 연구가 주류를 이루었다(Krumbein, 1967). 근래에 들어서 Carle and Fogg (1996, 1997)와 Carle et al. (1998)은 기존의 SIS와 Markov Chain을 수정-연계하여 세미베리오그램에 의존하지 않는 지구통계 기법을 제시하였다. 또한 Elfeki(1996) 및 Elfeki and Dekking(2001)은 지시자에 의존하며, 양함수법(explicit method)에 근거하여 전이확률을 직접적으로 예측 알고리듬에 이용하는 2차원 CMC 모델을 소개하였다.

기존 지구통계학에서 이용되어 온 추계론적 예측 모델들은 크리깅 (kriging)을 기초로 하는 음함수법(implicit method)을 근간으로 하여왔기 때문에, 지질학적 아이디어나 지구물리학 탐사자료 등과 같은 소프트정보(soft information)와 융합시키는 것이 쉽지 않았다. Elfeki and Dekking(2001)에 의해 개발된 CMC 모델은 전이확률을 기초로 하는 양함수법을 이용함으로써 기존 지구통계에서 어려웠던 소프트정보와의 데이터 융합(data fusion or assimilation)에 있어 용이성을 보일 것으로 예측된다. 또한 전이확률을 이용함으로써 지질학적 통찰이 직접적으로 연산에 개입할 수 있도록 조절하는 것이 가능하다.

그러나 기존 CMC 모델의 한계는 확률식에 기인하는 연산 알고리듬 적용의 경직성에 있다. 기존 CMC 확률식은 반드시 순차적인 연산(row/column-wise sequential path) 알고리듬과 병행하여 사용되며 이로 인한 참조 정보(conditional information)의 단-방향(uni-directional) 전달에 의해 경우에 따라서 지시자 파셀(indicator parcel)의 비의도적 오류(artifact)가 발생할 수 있다(Li et al., 2004). 따라서 본 연구의 목적은 보다 유연성 있는 연산 알고리듬 적용을 위해 보다 일반화된 2차원 연쇄 말콥체인

(GCMC) 확률식을 개발하며 이를 이용한 효율적 연산 알고리듬의 개발에 있다. 또한 개발된 모델 적용을 통해 가상의 2차원 토양도를 복원 함으로써 GCMC 모델을 기존 지구통계 기법의 대안으로 제시하는데 있다. 이를 위해 본 연구에서는 기존 CMC 조건부 확률식에서 출발하여 GCMC 확률식을 유도하고 개발된 확률식을 효율적으로 활용할 수 있는 연산 알고리듬을 개발하며 최종적으로 가상의 2차원 토양도에 적용 함으로써 일반화된 2차원 GCMC 기법의 효용성 및 가능성을 고찰할 것이다.

2. 실험방법

2.1. 기존 CMC 이론

말코비언(Markovian) 틀에서는 공간-미래의 상(state)은 공간-과거 히스토리와는 무관하며 오직 공간-현재의 상에만 의존한다. 이를 수식으로 표현하면, 개개의 추계론적 과정 $\{Z_n,\ n=0,1,2,\ldots\}$이 상-공간(state space) $\{S_1,S_2,\ldots,S_n\}$으로부터 값을 취한 연속적 무작위 변수(random variable)라 할 때, 말코비언이라 불리는 연쇄과정은 아래의 식,

$$\begin{aligned}&\mathrm{P}(Z_i=S_k\,|\,Z_{i-1}=S_l,\ Z_{i-2}=S_n,\ Z_{i-3}=S_r,\ \cdots,\mathrm{L},\ Z_0=S_p)\\&=\mathrm{P}(Z_i=S_k\,|\,Z_{i-1}=S_l)\,P_{lk},\end{aligned}\tag{1}$$

으로 표현될 수 있다(Ross, 2000). 여기서 p_{lk}는 S_l에서 S_k로 전이되는 빈도 확률로써 1차원 문제에서 전이확률 행렬(transition probability matrix)의 한 요소로 표현될 수 있으며 무작위변수, Z의 아래첨자는 위치를 그리고 상변수, S의 아래첨자는 주어진 위치에서의 상을 지시한다. 전이확률은 한 상에서 다른 상으로 전이되는 각 상들의 상대빈도(relative frequency)이며 다음과 같은,

$$\mathbf{P}=\begin{bmatrix}p_{11} & \cdots & p_{1n}\\ \vdots & \ddots & \vdots\\ p_{n1} & \cdots & p_{nn}\end{bmatrix},\tag{2}$$

의 $n\times n$의 요소를 가진 정방행렬로 표현될 수 있다. 만일 상의 출현이 비주기적(aperiodicity)이라 가정하면,

$$\lim_{N\to\infty}P_{lk}^{N}=w_k,\tag{3}$$

와 같이 표현할 수 있다. 여기서 w_k는 종말확률(marginal probability)로써 초기 상인 S_l에 의존하지 않으며 오직 S_k의 출현 빈도에만 의존한다. 또한 행렬 $\mathbf{p}$는 다음과 같은 조건을 만족한다: (1) 각 요소는 양의 수이다; (2) 각 행의 합은 1이다. Elfeki and Dekking(2001)은 이웃하는 공간-과거 및 일정거리 떨어져 있는 공간-미래의 상에 의해 참조되는 조건부 확률식을

$$\mathrm{P}\,(Z_i=S_j|Z_{i-1}=S_l,Z_N=S_q)=\frac{P_{jq}^{N-1}P_{lj}}{P_{lq}^{N-i+1}},\tag{4}$$

와 같이 유도하였다. 여기서 i는 현재 그리고 N은 미래 상의 위치를 의미하며 S_l, S_j, 및 S_q는 각각 과거, 현재, 및 미래의 상을 의미한다. 또한 두 개의 독립적 상-공간 $\{S_1,S_2,S_3,\ldots,S_n\}$이 교차하는 위치에서 개개의 두 상-공간이 전이되는 상이 같아짐을 강제하여 2차원 상-공간, $\{S_1,S_2,S_3,\ldots,S_n\}\times\{S_1,S_2,S_3,\ldots,S_n\}$,에서 CMC 확률식,

$$\begin{aligned}&\mathrm{P}\,(Z_{i,j}=S_k|Z_{i-1,j}=S_p,\ Z_{N,j}=S_q,\ Z_{i,j-1}=S_f,\ Z_{i,P}=S_g)\\&=C\frac{p_{pk}p_{kq}^{N-i}P_{fk}P_{kg}^{P-j}}{P_{pq}^{N-i+1}P_{fg}^{P-j+1}},\end{aligned}\tag{5}$$

을 개발하였으며, 여기서 비례상수, C, 는

$$C=\frac{P_{pq}^{N-i+1}P_{fg}^{P-j+1}}{\sum_{l=1}^{n}p_{pl}p_{lq}^{N-i}P_{fl}P_{lg}^{P-j}},\tag{6}$$

로 주어진다(Elfeki and Dekking, 2001).

2.2 기존 CMC 확률식의 한계성

기존 CMC 확률식의 한계는 공간-과거의 참조되는 (conditioned) 상이 항상 현재와 인접하여야 한다는 것이다. 이와 같은 조건은 첫째, 전체 연산을 열 또는 행 차순(line-by-line sequence)으로 계산할 수 밖에 없다는 경직성을 내포한다. 이에 의해 발생하는 정보의 단-방향(uni-directional) 전달에 의해 차분의 크기(Δx or Δy)가 작을 시에 또는 참조 정보 (conditioning information)의 빈도가 낮을 때 $\tan^{-1}(\Delta y/\Delta x)$ 방향으로 지시자 파셀(indicator parcel)의 인위적 기울기가 발생할 수 있다. 또 다른 문제점은 보다 응용적 면에 있어 보다 치명적일 수 있다. 많은 경우 야외에서의 데이터 획득은 공간적으로 연속적이지 않을 가능성이 크다. 이렇게 참조 데이터의 분포가 산재(random sparse distribution)하여 있을 때 기존의 2차원 연쇄 Markov Chain 확률식의 열 또는 행 차순 연산은 참조 정보가 충분하지 않을 경우 일부 참조 정보를 과평가(overestimate)하게 되어 적용하기에 적합하지 않으며, 무작위 차순 연산의 적용이 보다 타당하다 (Isaak, 1990).

Fig. 1. Schematic drawing for one-dimensional coupled Markov Chain calculation.

2.3 2차원 GCMC 모델

CMC 모델의 가장 큰 장점은 정보 참조가 직접적으로 이루어짐으로 인해서 연산 효율이 우수하다는 것이다. Fig. 1에서 보는 바와 같이 공간-과거에 위치하는 셀 M의 상이 S_p이고 공간-미래에 위치하는 셀 N의 상이 S_q일 때, 셀 i의 상이 S_j라 가정한다면 이는 확률 식으로,

$$P(Z_i=S_j|Z_M=S_p,Z_N=S_q)=\frac{P(Z_i=S_j,Z_M=S_p,Z_N=S_q)}{P(Z_M=S_p,Z_N=S_q)}, \quad (7)$$

와 같이 표현될 수 있다. 이러한 확률 식은 말코비언 특성과 조건부 확률 식을 이용하여

$$P(Z_i=S_j|Z_M=S_p,Z_N=S_q)$$
$$=\frac{P(Z_N=S_q|Z_i=S_j)P(Z_i=S_j|Z_M=S_p)}{P(Z_N=S_q|Z_M=S_p)}, \quad (8)$$

와 같이 전개할 수 있으며 이는,

$$P(Z_i=S_j|Z_M=S_p,Z_N=S_q)=\frac{p_{pj}^{|M-i|}\,p_{jq}^{|N-i|}}{p_{pq}^{|N-M|}}, \quad (9)$$

와 같은 일반화된 1차원 말콥체인 확률 식으로 재유도 될 수 있다. 여기서 셀 N이 셀 i로부터 무한히 멀리 떨어져 있을 때 $p_{jq}^{|N-i|}$와 $p_{pq}^{|N-M|}$은 모두 종말 확률, w_q에 수렴하며 조건부 확률 식은 오직 공간적 과거 정보에만 의존하며, 그와 반대의 경우에 조건부 확률 식은 공간적 미래 정보에만 의존하게 된다. 또한 셀 N과 M이 모두 셀 i로부터 무한히 멀리 떨어져 있을 때 조건부 확률 식은 w_j가 되며 이는 식이 어떠한 참조 정보에도 의존하지 않으며 단지 관찰된 지시자의 빈도만 따름을 의미한다. 기존 1차원 말콥체인 확률 식과 비교해 보았을 때, 유도된 식은 참조 정보의 상대 위치에 무관하게 무작위의 위치에서 연산될 수 있으며 공간적 과거와 미래에 있는 참조 정보를 동등하게 대우(honoring) 한다는 장점을 지니고 있다. 특수한 경우로, 만약 M의 위치가 i-1일 때 식은,

$$P(Z_i=S_j|Z_M=S_p,Z_N=S_q)=\frac{P_{pj}P_{jq}^{N-1}}{P_{pq}^{N-i+1}}, \quad (10)$$

와 같이 Elfeki and Dekking(2001)에 의해 유도된 1차원 식과 같아지며 이는 GCMC 확률식의 특수형태이다.

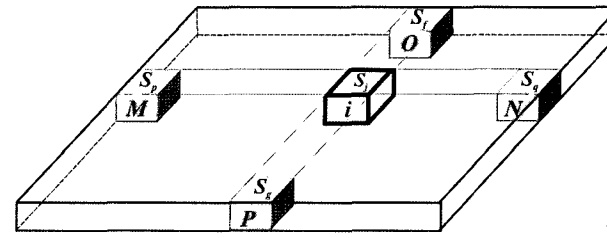

Fig. 2. Schematic drawing for two-dimensional coupled Markov Chain calculation.

1차원과 마찬가지 방법으로 2차원 확률식 역시 개발될 수 있으며, Fig. 2와 같은 참조 정보의 배열이 있을 경우 이는,

$$P(Z_i=S_j|Z_M=S_p,Z_N=S_q,Z_O=S_f,Z_P=S_g)$$
$$=\frac{p_{pj}^{\delta_w|M-i|}\;p_{jq}^{\delta_e|N-i|}\;p_{fj}^{\delta_n|O-i|}\;p_{jg}^{\delta_s|P-i|}}{\sum_k^n p_{pk}^{\delta_w|M-i|}\;p_{kq}^{\delta_e|N-i|}\;p_{fk}^{\delta_n|O-i|}\;p_{kg}^{\delta_s|P-i|}}, \quad (11)$$

와 같이 주어진다. 여기서 δ_x는

$$\delta_x=\begin{cases}0 & \text{distance to conditioning information} > \text{searching range}\\ 1 & \text{distance to conditioning information} < \text{searching range}\end{cases}, \quad (12)$$

와 같이 주어지며 δ의 값이 모두 1일 경우 정규식 그렇지 않을 경우 비정규식이라 정의한다.

2.4. 개선된 연산 알고리듬

기존 2차원 CMC 연산 알고리듬은 확률식의 한계에 의하여 이를 이용한 연산에 많은 제약을 받았다. 그러나 2차원 GCMC 확률 식을 이용한 연산에서는 확률식의 경직성에 의한 연산의 제약이 없으며 임의의 위치에서 주변에 위치한 참조 정보를 이용한 연산이 가능하다. 그러므로 2차원 GCMC 모델의 연산은 참조 정보가 순차적 또는 무작위 적으로 주어졌을 때 모두를 다룰 수 있는 장점을 가지고 있다.

순차적 데이터로부터 전이확률을 얻는 방식은 기존 CMC 모델의 연산 알고리듬을 따른다(Elfeki and Dekking, 2001). 실제 야외에서 얻어지는 데이터의 이산성을 고려하여 데이터가 무작위의 위치에서 주어졌을 때에는 개개 참조 정보의 동등성을 보존하기 위하여 역거리 가중치 방법(inverse distance weighting method, IDW)으로 보간된 트레이닝 맵에 기초하여 전이확률을 계산한다. 단, 본 연구에 이용된 역거리 가중치 보간에서는 통계적 등방성

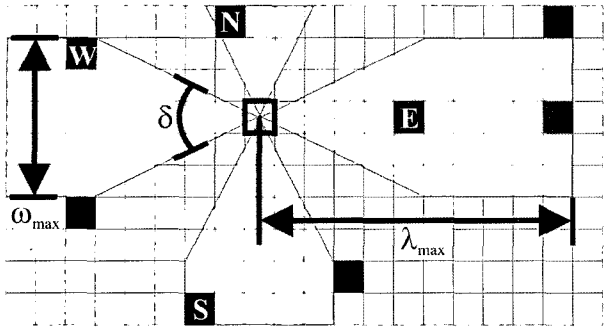

Fig. 3. Schematic drawing of calculation using generalized two-dimensional coupled Markov Chain probability equation.

을 가정하였으며 이는 여타 지구통계기법과 마찬가지로 데이터의 수가 충분하지 않으며 실제 토양의 통계적 등방성이 보장되지 않을 시에는 예측에 오류가 발생할 가능성이 있다. 본 연구에서 사용된 이산적 데이터 분포는 Fig. 4와 같으며 총 2개의 샘플링 빈도를 사용하였다.

연산 위치는 기존 2차원 연쇄 Markov Chain의 순차적 방식과 달리 난수발생에 의해 무작위적으로 결정되며 주어진 연산 위치에서 N, E, W, 및 S의 네 방향으로 참조 정보를 찾게 된다. 각 지시자의 상관거리 중 최소 상관거리는 셀 크기의 스케일링 팩터(scaling factor)로 사용 되며 최대 상관거리는 참조 정보를 찾는 최대탐색거리, λ_{max} (search range), 및 최대탐색폭, w_{max}(band width)의 스케일링 팩터로 이용된다. 보다 효율적으로 참조 정보를 찾기 위하여 탐색각, δ(angle tolerance),가 이용되며, 참고로, 기존 2차원 CMC 모델의 연산에서는 탐색각의 허용치가 0으로 한정되었다.

2.5. 연산 알고리듬

2차원 GCMC 모델을 이용한 도메인의 추계론적 예측은 다음과 같은 과정을 따른다.

Step 1: 2차원 연산 도메인이 수직 및 수평 샘플링 간격 또는 최소 상관거리에 근거하여 차분한다.

Step 2: 야외관찰을 통해 얻은 참조 정보는 연산 도메인에 저장된다.

Step 3: 순차적 데이터의 경우, 주어진 참조 정보에 기초하여 수직 및 수평 전이확률이 방향 별로 계산되며, 무작위적 위치에서 주어진 데이터의 경우, 역거리 가중치 방법에 의해 보간된 데이터에 기초하여 전이확률이 계산된다.

Step 4: 난수발생에 의해 무작위로 연산 위치를 선정한다.

Step 5: 도메인의 경계를 제외한 구역에서는 정규 2차원 확률식(Eq. 11)이 적용되며 그 외의 지역에서는 비정규 2차원 확률식 (Eq. 11)이 적용되며 모든 셀을 방문할 때까지 무작위 위치를 선정하여 연산을 진행한다. 참조 정보의 수가 1 이하일 경우에는 연산 순위를 마지막으로 미룬다.

3. 결과 및 고찰

Fig. 5a와 같은 임의의 토양도가 2차원 GCMC 모델을 검토하기 위한 트레이닝 맵으로 사용되었다. 이 토양도는 그 토양의 특성에 의해 다섯 개의 지시자로 구성되며, 전체 토양도를 통해 대표되는 지시자가 위치에 따라 지속적으로 변화하는 통계적 경향을 보이는 비정상 분포(non-stationary distribution)를 갖는다.

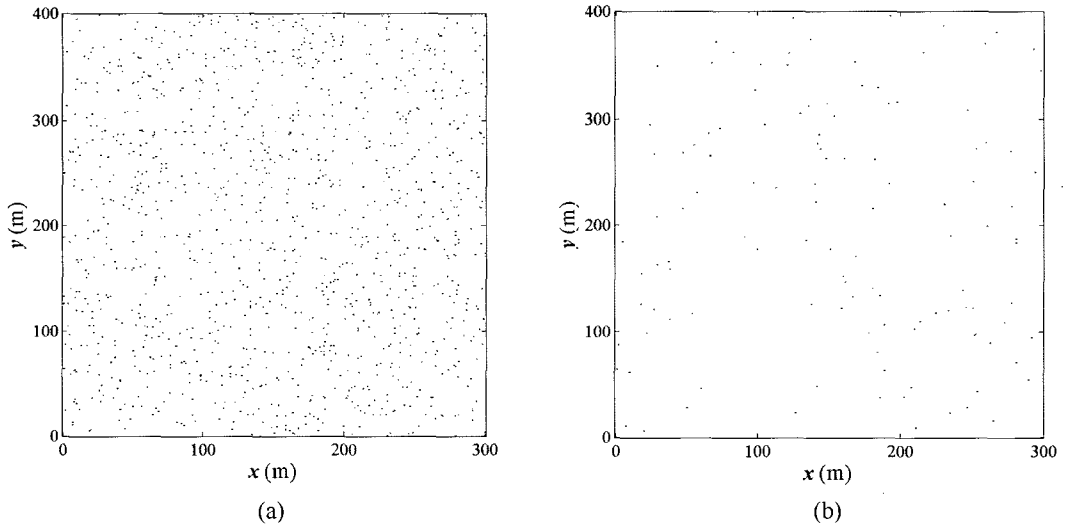

Fig. 4. Randomly chosen sampling location maps of each frequency: (a) 1/100 m^2 frequency; (b) 1/1000 m^2 frequency.

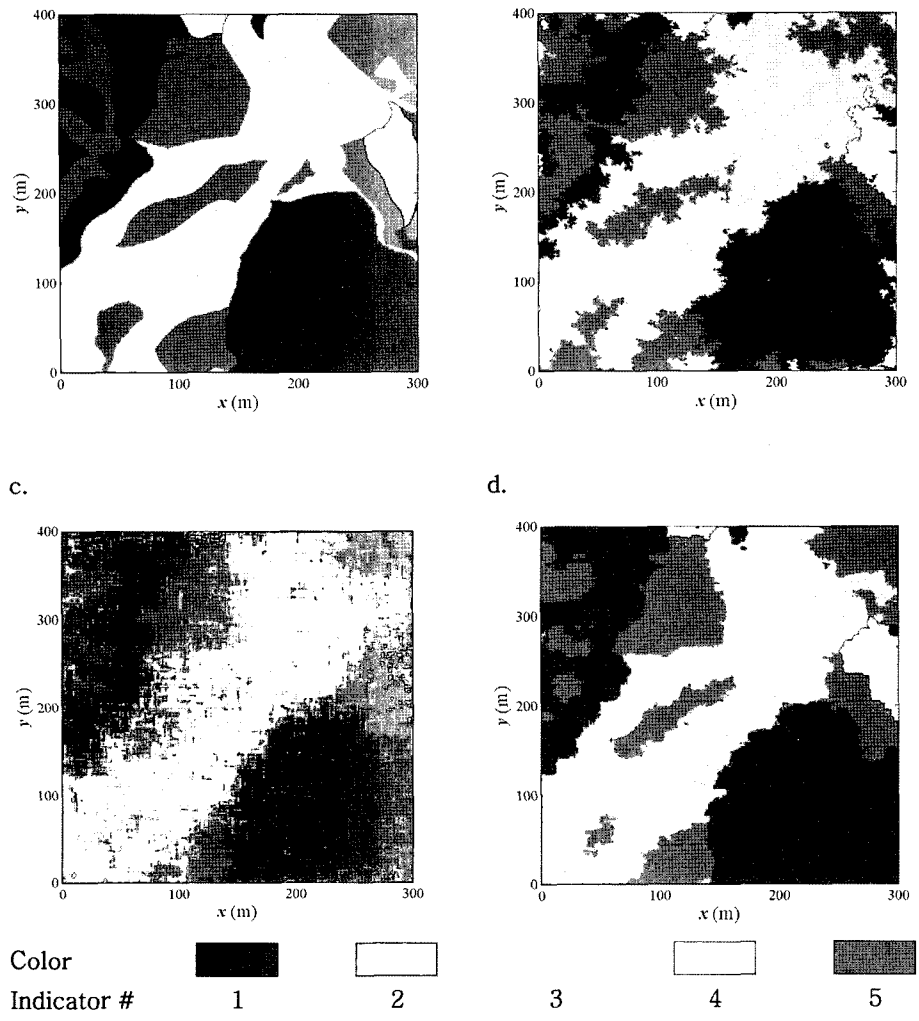

Fig. 5. Reconstructed map out of the data from the original soil map with 1/100 m^2 sampling frequency at randomly chosen locations: (a) original soil map; (b) reconstruction of the map using sequential indicator simulation algorithm; (c) reconstruction of the map using zero angle tolerance coupled Markov Chain algorithm; (d) reconstruction of the map using generalized two-dimensional coupled Markov Chain algorithm (each indicators does not need to match with a particular lithology or soil type).

Fig. 5b-d는 1개/100 m^2의 빈도로 샘플링한 데이터를 이용하여 각 알고리듬을 통해 재구성한 토양도이다. 두 가지 데이터 빈도(1개/100 m^2와 1개/1000 m^2)로 샘플링한 데이터를 이용하여 원 토양도를 예측하였을 경우, 연산 속도 면에서 볼 때 2차원 GCMC 모델이 SIS에 비해 20% 이상 빠름을 보였다. 또한 다차원 가변거리법 (simulated annealing, SA)을 실시하지 않은 SIS의 재현 결과와 일반화된 연쇄 Markov Chain의 재현 결과를 비교해 보았을 때, 경계부의 엔트로피 면에서 2차원 GCMC 모델에 의한 모사 결과가 보다 원 토양도와 가까운 것을 볼 수 있다(Fig. 5b와 5d). 연산의 정확도를 정량적으로 비교하기는 힘들지만 일반화된 연쇄 Markov Chain이 SIS를 대체할 수 있는 가능성을 보여주고 있다. Fig. 5c는 기존 2차원 CMC 모델에 의해 복원된 토양도이며 모사 결과가 2차원 GCMC 모델이나 SIS에 의한 모사결과에 비해 현저하게 부정확한 것을 볼 수 있다.

Fig. 6는 공간적 거리 변화에 대한 각 지시자들의 상호관계 변화를 전이확률을 이용하여 분석한 결과 예이다. 이를 위해 1개/1000 m^2의 빈도로 샘플링한 데이터를 이용하였다. 그림에서 보는 바와 같이 각 지시자들의 상관관계는 거리에 따라 각각 지수적으로 감소하는 경향을 보인다. 대각선요소들은 자체 상관관계(self correlation)의 거리에 따른 감쇄를 보여주며 비대각선 요소들은 상호 상관관계의 거리에 따른 기울기의 감쇄를 보여주며 크로스 베리오그램(cross-variogram)에 상응한다.

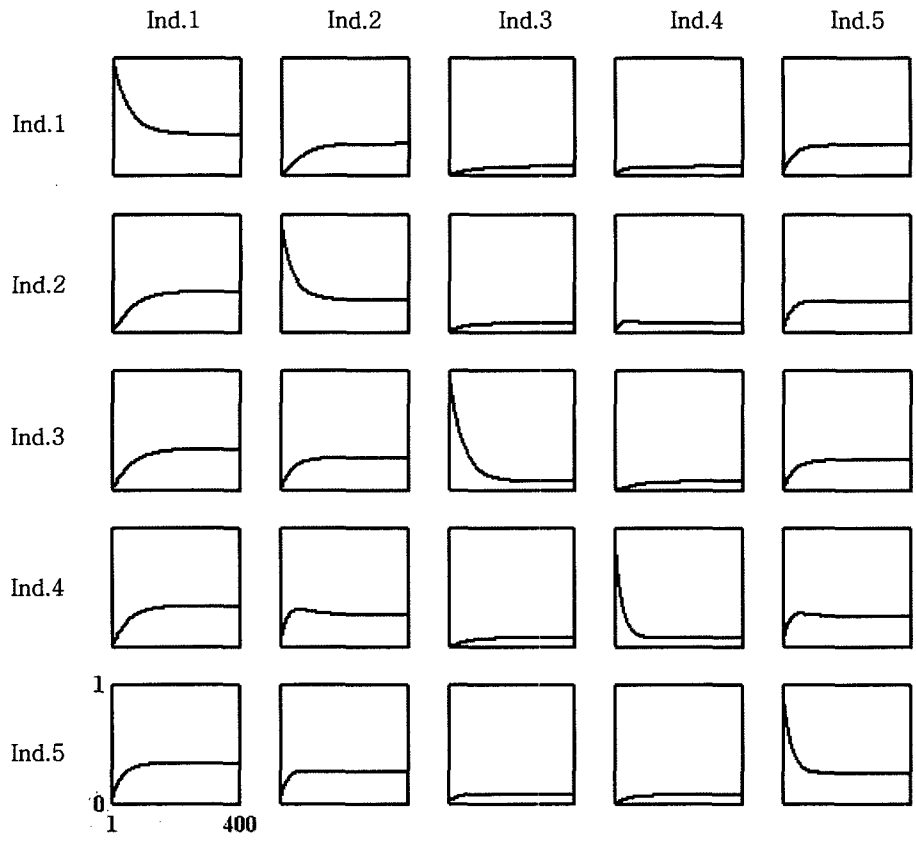

Fig. 6. Correlation decaying(with distance) matrix of x-positive-direction transition probability(the horizontal axis is lag distance from 1 to 400 m and the vertical axis is correlation from 0 to 1 where 1 means perfect correlation and 0 is no correlation).

Fig. 7은 Monte Carlo 모사를 통해 제작한 각 지시자의 확률 분포이다. 이를 위해 1개/1000 m^2의 빈도로 샘플링한 데이터를 이용하였으며, 총 30개의 재현을 생성하였다. Fig. 7을 통해 일반화된 연쇄 Markov Chain은 대부분의 지시자에 대해 안정적으로 예측함을 볼 수 있다. 그러나 Fig. 7c에서 보는 것과 같이 저빈도로 샘플링(under-sampling)된 지시자에 대해서 상대적으로 취약함을 볼 수 있다. 이러한 단점을 개선하기 위한 노력이 현재 진행 중이며, 복합적인 데이터를 이용한 데이터 융화(data fusion or assimilation)가 그 대안으로 검토 중에 있다.

4. 요 약

본 연구는 2차원 GCMC 확률식의 개발과 이를 활용하는 연산 알고리듬의 개발, 그리고 최종적으로 2차원 GCMC 모델을 가상의 토양도에 적용함으로써 기존 지구통계 기법과의 비교 및 대안적 지구통계 기법으로의 가능성 제시를 목적으로 하였다. 본 연구를 통해 새롭게 개발된 2차원 GCMC 확률식은 기존 CMC 확률식에 비해 보다 유연한 참조 정보 활용 가능성을 가지며 특수한 경우로 기존 CMC 확률식이 유도되었다. 또한 순차적 연산의 인위적 오류 발생 가능성 및 실제 야외 데이터의 낮은 빈도를 고려하여 무작위 위치에서 각 범위를 이용한 연산 알고리듬이 개발되었다. 개발된 모델은 가상의 2차원 토양도에 적용 되었으며 기존 지구통계 기법인 SIS에 비하여 손색이 없는 새로운 지구통계 기법으로 다양한 예측에 이용될 수 있는 가능성을 제시하였다.

낮은 빈도로 샘플링 된 지시자에 대해서는 기존 지구통계 기법과 마찬가지로 저평가(underestimate) 되는 현상을 나타났으며 이를 보완하기 위하여 소프트 데이터를 포함한 다양한 소스의 데이터 융합 등을 바탕으로 한 계속적인 연구가 요구된다. 또한 본 모델의 지하매체 분포에의 적용을 위해 3차원 공간상의 확장이 요구되며 이에 대한 연구가 현재 진행 중이다.

본 연구의 결과는 오염물질에 의한 위해성 파악, 자원광물의 경제성 파악, 지하수 산출 가능성 파악 등 다양한 추계론적 예측에 응용이 가능하다.

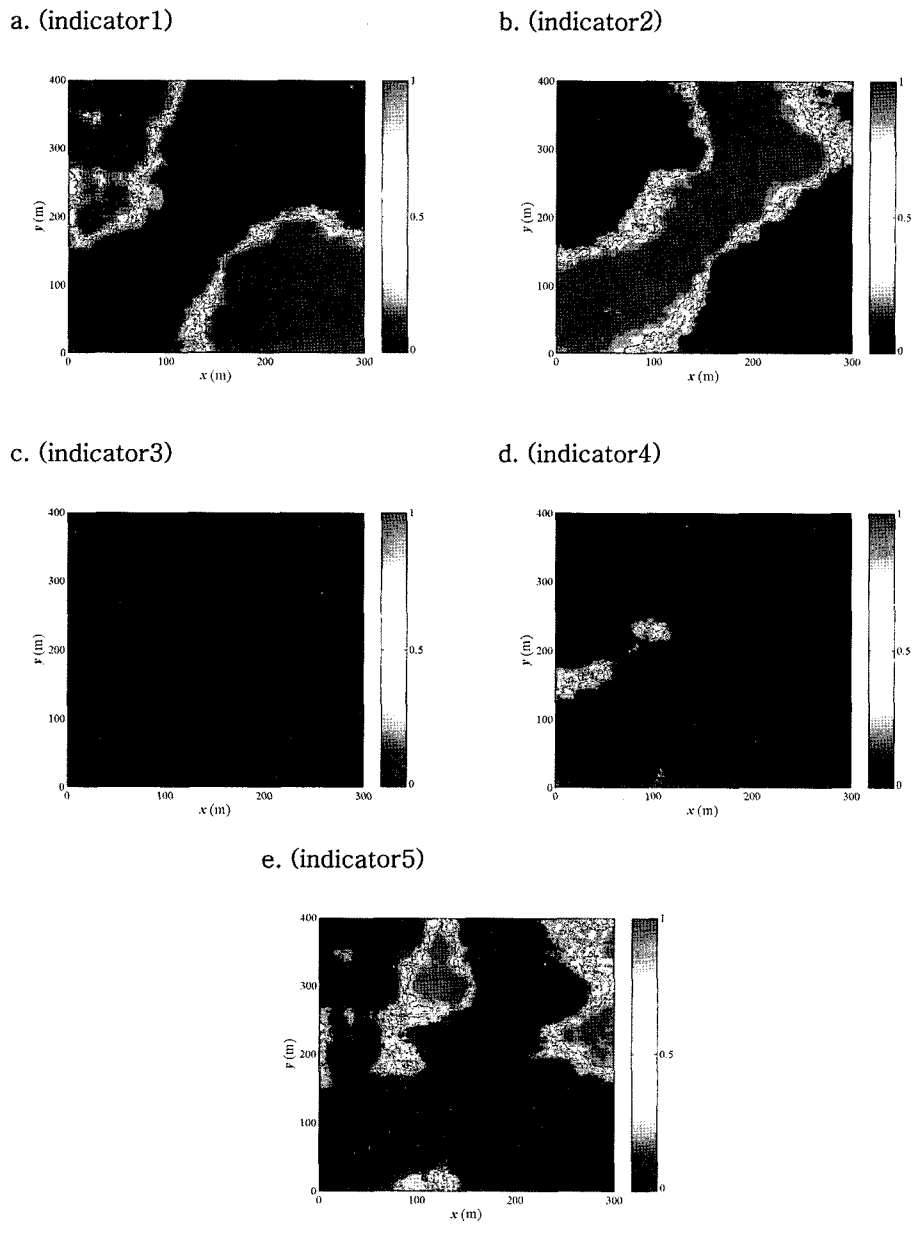

Fig. 7. Probability maps of each indicator analyzed from multiple realizations through Monte Carlo analysis (red color indicates high probability and blue color is low probability): (a) probability map of indicator1; (b) probability map of indicator2; (c) probability map of indicator3; (d) probability map of indicator4; (e) probability map of indicator5.

사 사

본 연구는 2005년도 경북대학교학술진흥연구비(20050107-00)에 의하여 연구되었음

참 고 문 헌

Carle, S.F. and Fogg, G.E., 1996, Transition probability-based indicator geostatistics, *Mathematical Geology*, **28**(4), 453-476.

Carle, S.F. and Fogg, G.E., 1997, Modeling spatial variability with one- and multidimensional continuous Markov chains, *Mathematical Geology*, **29**(7), 891-917.

Carle, S.F., LaBolle, E.M., Weissmann G.S., VanBrocklin D., and Fogg, G.E., 1998, Conditional simulation of hydrofacies architecture: A transition probability/Markov approach, in Hydrogeologic Models of Sedimentary Aquifers, *SEPM Concepts in Hydrol. Environ. Geol.*, Tulsa, OK, p. 147-170.

Deutsch, C.V. and Journel, A.G., 1992, GSLIB: Geostatistical Software Library and User's Guide, New York, Oxford Univer-

sity Press, p. 340.

Elfeki, A.M.M., 1996, Stochastic Characterization of Geological Heterogeneity and Its Impact on Groundwater Contaminant Transport. Ph.D.Thesis, TU Delft, Delft, The Netherlands.

Elfeki, A.M.M. and Dekking, M., 2001, A Markov Chain Model for Subsurface Characterization: Theory and Applications, *Mathematical Geology*, **33**(5), 569-589.

Isaak, E., 1990, The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data, PhD dissertation, Stanford University, Stanford, California.

Krumbein, W.C., 1967, FORTRAN IV Computer Program for Markov Chain Experiments in Geology: Computer Contribution 13, Kansas geological survey, Lawrence, Kansas, p. 38.

Li, S.-G., McLaughlin, D., and Liao, H.-S., 2003, A computationally practical method for stochastic groundwater modeling *Advances in Water Resources*, **26**, 1137-1148.

Li, W., Zhang, C., Burt, J.E., Zhu, A-.X., and Feyen, J., 2004 Two-dimensional Markov Chain Simulation of Soil Type Spatial Distribution, *Soil Society of America Journal*, **68**, 1479 1490.

Ross, S., 2000, Introduction to Probability Models, 7th ed., Academic Press, San Diego, California, p. 693.