

지구통계학적 시뮬레이션을 이용한 원격탐사 화상 분류 결과의 공간적 불확실성 분포의 추정

Assessing spatial uncertainty distributions in remote
sensing data classification using geostatistical simulation

박노옥*¹⁾, 지광훈¹⁾, 권병두²⁾

No-Wook Park*, Kwang-Hoon Chi, Byung-Doo Kwon

1) 한국지질자원연구원 지질자원정보센터

E-mail: nwpark@kigam.re.kr

2) 서울대학교 지구과학교육과

요약

이 연구에서는 원격탐사 자료를 이용하여 얻어진 분류 결과로부터 분류 정확도의 공간적 불확실성을 추정하고자 하였다. 기존 분류결과로부터 얻어지는 토지 피복별 확률값을 지구통계학적 시뮬레이션 기법을 이용하여 참조자료의 공간적 분포와 통합하였다. 다중센서 화상 통합에 적용한 결과, 각 토지피복 항목별로 공간적인 정확도 분포를 얻을 수 있을 수 있었으며 이러한 자료는 분류결과를 해석하는데 유용하게 사용될 수 있을 것으로 기대된다.

1. 서론

원격탐사 화상을 이용한 토지 피복/이용도 작성은 여러 활용분야 중에서 많은 연구가 진행되어온 분야중 하나이다. 원격탐사 화상을 이용하여 얻어진 토지 피복/이용도가 GIS 기본도 갱신 혹은 주제도로서 활용되기 위해서는 지속적인 갱신뿐만 아니라 분류 결과의 정확도에 대한 부가적인 정보를 필요로 한다. 현재까지 원격탐사 화상 분류 결과의 정확도는 훈련집단과 독립적으로 준비한 참조자료(reference data)를 화상 분류 결과와 직접적으로 비

교하는 오차 행렬(error matrix)을 통해 얻어 왔다([1]). 그러나 오차 행렬로부터 얻어진 통계치는 화상 전체에 대한 일종의 global statistics를 제공하지만, 화상 분류 결과 정확도의 공간적인 분포를 제공하지는 못한다. 따라서 오차행렬로부터 얻어진 자료로부터 화상 분류결과에서 어느 부분이 분류정확도가 낮은지, 혹은 추가적인 지표조사를 필요로 하는 지에 대한 정보를 얻을 수 없다.

이 연구에서는 원격탐사 화상 분류 결과와 참조자료를 이용하여 분류정확도의 공간적 분포를 추정하는 방법을 제안하였다.

이를 위해 소수의 참조자료와 원격탐사 화상으로부터 얻어지는 연구지역 전체에 대한 확률값을 통합할 수 있도록 지구통계학의 시뮬레이션 기법을 적용하였다. 적용 시뮬레이션 기법은 범주형 자료에 적용이 가능하며, 연구지역에 대해서 exhaustive하게 획득한 확률값과 sparse한 참조자료를 통합할 수 있는 Sequential indicator simulation with local means (SISLM)이다. 기존 참조자료와 원격탐사 화상 처리결과를 결합하기 때문에 훈련집단에 대해 부트스트랩을 이용하여 얻어진 통계치를 지시자 크리깅을 수행하여 분류 자체의 공간적 불확실성 분포를 얻은 Steele et al. ([2])과는 접근이 다르다. 제안된 기법의 적용을 위해 다중센서 화상 통합을 목적으로 실험을 수행하였다.

2. 제안 기법

(1) 지구통계학

1960년대 이후에 광상의 위치를 추정하기 위해 개발된 지구통계학은 기존에 내삽 방법으로 알려진 크리깅으로 대표되어 왔다. 지구통계학과 통계학과의 가장 큰 차이는 기본적으로 통계값을 추출하는데 있어서 공간적인 분포를 고려한다는 것이다. 자료의 변이를 나타내는 통계치로 분산 혹은 상관계수 등이 주로 사용되지만, 지구통계학에서는 거리와 방향까지 고려한 배리오그램(variogram)을 기본적으로 이용한다. 1980년대 말까지는 내삽방법으로 많이 활용되어 왔으나, 1990년대 들어 시뮬레이션 기법이 개발됨으로써 국소 혹은 공간적 불확실성 분석에 많이 활용되고 있다([3]). 지구통계학이 기본적으로 sparse한 자료의 공간적 분포를 추정

은 시뮬레이션하기 때문에 연구지역에 대한 전체 자료를 획득할 수 있는 원격탐사 분야에는 많이 적용되어 오지 못했다. 1990년대 초부터 구름으로 인한 미관측지점의 값을 추정하기 위해 크리깅을 사용하여 왔는데([4]), 1990년대 중후반부터 sparse한 지표조사자료와 원격탐사자료를 통합할 수 있는 기법들이 개발되어서 이에 대한 관심이 증가되고 있는 상황이다([5], [6]).

(2) 크리깅 vs. 시뮬레이션

크리깅과 시뮬레이션 모두 배리오그램으로부터 얻어지는 공간적 상관관계를 기반으로 하지만, 계산과정과 최종 결과물에 대해서는 차이가 크다.

크리깅은 일종의 최소오차 회귀분석 방법으로 자료 근처에서는 분산을 작게 하고, 자료가 드문 곳에서는 분산을 크게함으로써 전체 분산을 최소로 하면서 미관측지점의 값을 예측하게 된다. 따라서 전반적으로 국소적인 특성이 사라진 평활화된 결과를 얻게 되며, 원래 자료의 배리오그램이나 히스토그램을 재생산할 수 없는 단점이 있다. 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 원래 자료의 값과 분포, 배리오그램 등의 통계치를 재생산할 수 있다. 또한 몬테카를로 기법을 이용함으로써 동일한 시뮬레이션 조건하에서도 확률적으로 등가인 다른 결과를 나타내므로 이들의 비교를 통해 공간적인 불확실성 분포를 추정할 수 있다([3]).

(3) SISLM

이 연구에서는 여러 시뮬레이션 기법중에서 범주형 자료처리와 부가자료를 함께 고려할 수 있는 Sequential indicator simulation with local means (SISLM)를 적용하였다.

Sequential Indicator Simulation(SIS)은 지시자 변환을 우선 수행한 후에 주어진 공분산 함수를 만족하면서 매단계에서 시뮬레이션된 자료를 다음 계산 단계에서는 알려진 자료로 가정하고 시뮬레이션을 모든 화소위치에서 순차적으로 진행한다.

SIS의 실제 수행과정은 다음과 같다 ([7]).

- ① 주어진 자료에 대해 지시자 변환을 수행하고, 지시자 배리오그램을 계산한다.
- ② 각 지시자 값에 따른 지시자 크리깅을 수행한다.
- ③ 지시자 크리깅으로부터 구한 누적활동분포로부터 난수를 이용하여 하나의 값을 생성한다.
- ④ 생성된 값을 다음 단계의 계산값에서는 주어진 자료값으로 가정하여 임의의 경로를 따라 ② - ③의 과정을 반복하여 모든 화소에서의 값을 계산한다.

연구지역에서 K개의 토지피복 항목 { $\omega_k, i=1, \dots, K$ }과 n개의 참조자료를 고려할 때, 참조자료 위치 { $u_\alpha, \alpha=1, \dots, n$ }에서의 지시자 변환을 다음과 같이 수행한다.

$$I(u_\alpha; \omega_k) = \begin{cases} 1 & u_\alpha \in \omega_k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

참조자료가 없는 위치 u에서 토지피복

항목 ω_k 의 발생 확률은 다음의 지시자 크리깅 시스템에 의해 결정된다.

$$I(\mathbf{u}; k) - p_k = \sum_{\alpha=1}^n \lambda_\alpha(\mathbf{u}_\alpha; k) \cdot [I(\mathbf{u}_\alpha; k) - p_k] \quad (2)$$

SIS와 달리 SISLM은 지시자 크리깅을 수행할 때, 전체 평균으로 주어지는 즉 위치와 무관하게 주어지는 초기 확률값 p_k 을 위치와 관계된 초기확률값 $p_k(\mathbf{u})$ 으로 대체함으로써 연구지역 전체에 걸쳐 이미 얻어진 값을 고려하게 된다.

$$I(\mathbf{u}; k) - p_k(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha(\mathbf{u}_\alpha; k) \cdot [I(\mathbf{u}_\alpha; k) - p_k(\mathbf{u})] \quad (3)$$

이 연구에서는 local means로 일종의 soft 분류결과 얻어지는 각 토지 피복별 확률값을 이용하였다.

식 (3)을 이용하여 N번의 시뮬레이션을 수행하였다고 하면, 모든 화소에서는 N개의 우세한 토지피복 항목을 가지게 되며 이들의 평균값을 취함으로써 최종적인 개별 토지피복항목에 대한 분류 정확도의 공간 분포를 얻게 된다.

시뮬레이션 결과 얻게 되는 공간적 불확실성 혹은 정확성 분포는 분류결과 확률값으로부터 계산할 수 있는 불확실성 분포와는 다르다. 후자는 분류 모델 자체에 대한 불확실성 분포를 제공하지만, 시뮬레이션 결과는 분류 모델 자체의 불확실성과 참조자료와 함께 통합하여 연관지어 해석할 수 있는 정보를 제공하게 된다.

3. 적용 결과

(1) 대상지역과 사용 자료

제안된 분류 기법의 적용 가능성을 실험하기 위하여 IEEE GRSS Data Fusion Committee에서 제공하는 grss-dfc-0006 자료([9])에 적용하였다. 사용 자료는 AMS 광학자료 6채널, 다중 주파수(P, L, C 밴드), 다중 편광(HH, HV, VV) SAR 자료 9채널 등 총 15개 채널의 화상을 포함한다. 고려한 토지피복 항목은 sugar beets, stubble, bare soil, potatoes, carrots 등 5개이다. 훈련집단과 오차 행렬 작성을 위한 참조자료는 원래 제공하는 자료를 이용하지 않고, stratified random sampling을 통해 새로이 작성하였다.

(2) 적용 결과

15개 다중 센서 자료로부터 분류를 수행하기 위해 Probabilistic Neural Network을 적용하였다. 이렇게 얻어진 분류결과와 참조자료를 함께 SISLM에 적용하였고, 50번의 시뮬레이션을 수행하였다. 만약 특정 위치에서 분류정확도가 높다면 50번의 시뮬레이션의 결과에서 최종 결정된 분류항목이 많이 나타나야 한다. 반면, 여러 항목이 시뮬레이션 결과에서 나타난다면 분류정확도가 상대적으로 낮다고 간주할 수 있다. Fig. 1은 처음 2번의 시뮬레이션 결과를 나타내고 있다. 원래 분류결과에서도 주로 토지 피복의 경계부분에서 오차가 나타났는데, 이러한 양상은 시뮬레이션 결과에서도 나타났다. 즉 2개의 시뮬레이션 결과에서 차이가 나는 부분은 주로 토지 피복 사이의 경계에 해당된다.

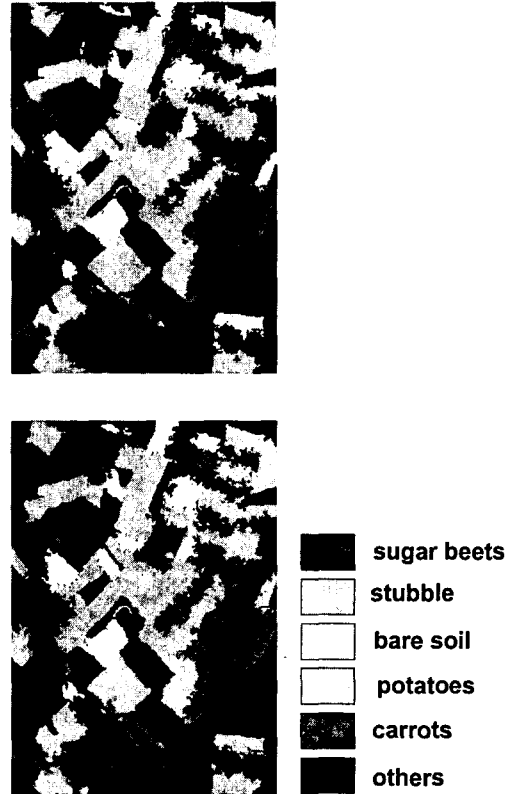


Fig. 1. Two realizations generated by sequential indicator simulation with local means.

50번의 시뮬레이션으로부터 각 화소에서 우세하게 나타나는 토지 피복 항목의 공간적인 정확도에 관한 확률값은 아래 식을 이용하여 얻게된다.

$$P(\mathbf{u}; \omega_k) = \frac{1}{50} \sum_{i=1}^{50} I^{(i)}(\mathbf{u}; \omega_k) \quad (4)$$

만약 특정 화소에서 sugar beets, stubble, bare soil, potatoes, carrots이 50번의 시뮬레이션을 통해 나타난 회수가 20, 15, 0, 10, 5일 경우, 개별 토지 피복에 대한 확률값은 각각 0.4, 0.3, 0.0, 0.2, 0.1로 얻게 된다. 최대 확률값이 1에 가까울수록 불확실성이 작음을 의미한다.

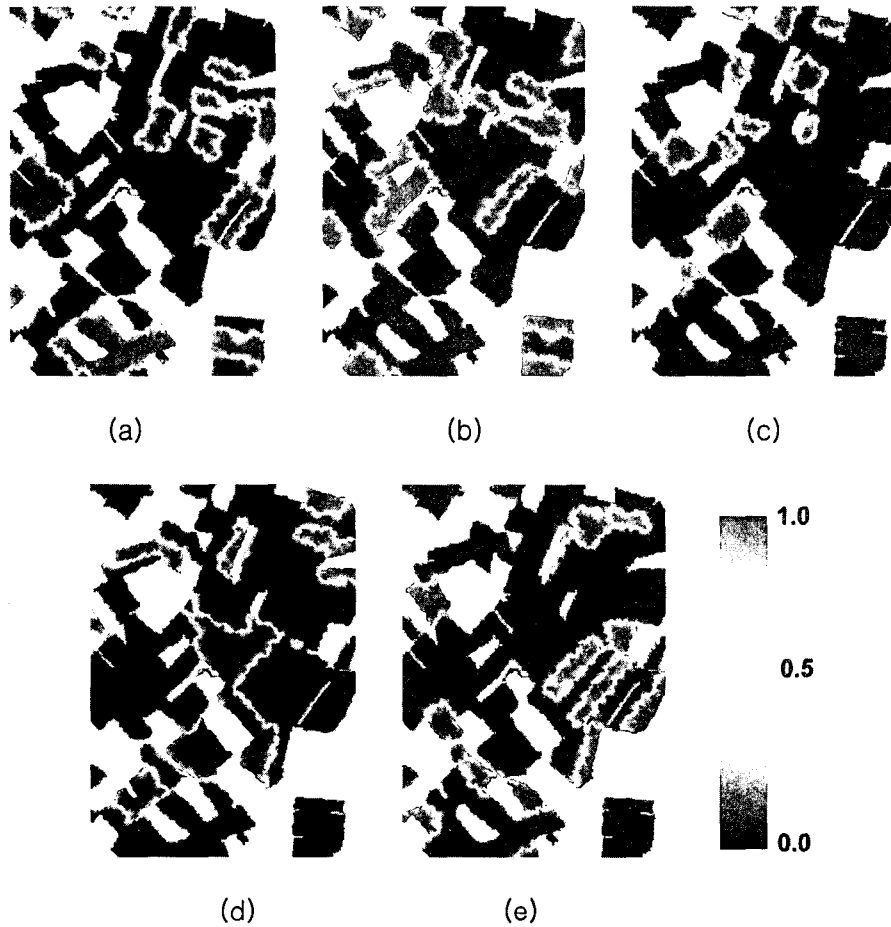


Fig. 2. Probabilities of spatial accuracy distributions for each land-cover class from 50 simulated land-cover maps; (a) sugar beets, (b) stubble, (c) bare soil, (d) potatoes, (e) carrots.

즉 $1 - P(\mathbf{u}; \omega_k)$ 이 불확실성의 척도로 정의할 수 있다. Fig. 2는 5개의 토지피복 항목에 대한 공간적 정확성의 분포를 나타낸다. Fig. 1의 결과의 시각적 판독을 통해 확인한 경계 부근에서의 불확실성이 개별 토지 피복별로 나타남을 확인할 수 있었다. 사례연구 지역은 격자 형태를 가지는 농업지역에 해당되는데 경계부에서 토지 피복의 혼재 때문에 이러한 결과

가 나타난 것으로 판단된다.

4. 결론

이 연구에서는 원격탐사 화상분류결과와 불확실성을 모사하기 위하여 지구통계학적 시뮬레이션을 적용하였다. 제안 방법은 기존 오차행렬로 제공되는 정보 이외의 불확실성의 공간분포를 제공할 수 있기

때문에 분류 결과의 해석에 보조자료로 활용할 수 있으며, 또한 추가 지표조사를 수행할 지역에 대한 정보를 제공할 수 있다. 또한 시뮬레이션을 통해 얻어진 확률적으로 등가인 여러 토지 피복도는 분류 결과의 질을 판단하는데 이용할 수 있을 뿐만 아니라 토지 피복도를 입력자료로 사용하는 모델의 불확실성 추정에도 사용할 수 있을 것으로 기대된다. 예를 들어 토양 유실 혹은 경관 모델링 등에 시뮬레이션을 통해 얻어진 다수의 토지 피복도를 입력 자료로 사용함으로써 여러 가지 결과값을 얻을 수 있다. 이렇게 얻어진 결과값은 입력자료의 모델에 대한 오차 전파 혹은 결과값의 신뢰도 추정에 사용할 수 있는데, 앞으로 이러한 적용 연구를 수행할 예정이다.

사사

이 연구는 과학기술부 원격탐사 기술개발 사업과 위성자료 공공활용연구의 지원을 받았습니다. 실험에 사용된 다중센서 원격탐사 화상은 IEEE GRSS DFC로부터 제공받았으며 이에 감사드립니다.

참고 문헌

[1] Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Environment*, 37: 35-46.

[2] Steele, B.M., J.C. Winne, R.L. Redmond, 1998. Estimation and mapping of local misclassification probabilities for thematic land cover maps, *Remote Sensing of Environment*, 66: 192-202.

[3] Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*, Oxford University Press.

[4] Rossi, R.E., J.L. Dungan and L.R. Beck, 1994. Kriging in the shadows: geostatistical interpolation for remote sensing, *Remote Sensing of Environment*, 49: 32-40.

[5] Dungan, J.L., 1998. Spatial prediction of vegetation quantities using ground and image data, *International Journal of Remote Sensing*, 19: 267-285.

[6] Goovaerts, P., 2002. Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data, *Journal of Geographical Systems*, 4:99-111.

[7] Deutsch, C.V. and A.G. Journel, 1998. *GSLIB: Geostatistical software library and user's guide*, Oxford University Press.

[8] Lillesand, T.M. and R.W. Kiefer, 1994. *Remote sensing and image interpretation*, John Wiley & Sons, Inc.

[9] <http://www.dfc-grss.org>. IEEE GRSS Data Fusion Committee, data set grss-dfc-0006.