

공간 통계를 이용한 원격탐사 화상 분류의 공간적 불확실성 분포 추정

박노욱* · 지광훈* · 권병두**

한국지질자원연구원 지질자원정보센터*, 서울대학교 지구과학교육과**

Assessing Spatial Uncertainty Distributions in Classification of Remote Sensing Imagery using Spatial Statistics

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

Geoscience Information Center, Korea Institute of Geoscience and Mineral Resources*,

Department of Earth Sciences, Seoul National University**

Abstract : The application of spatial statistics to obtain the spatial uncertainty distributions in classification of remote sensing images is investigated in this paper. Two quantitative methods are presented for describing two kinds of uncertainty; one related to class assignment and the other related to the connection of reference samples. Three quantitative indices are addressed for the first category of uncertainty. Geostatistical simulation is applied both to integrate the exhaustive classification results with the sparse reference samples and to obtain the spatial uncertainty or accuracy distributions connected to those reference samples. To illustrate the proposed methods and to discuss the operational issues, the experiment was done on a multi-sensor remote sensing data set for supervised land-cover classification. As an experimental result, the two quantitative methods presented in this paper could provide additional information for interpreting and evaluating the classification results and more experiments should be carried out for verifying the presented methods.

Key Words : Spatial uncertainty, Geostatistical simulation, Classification.

요약 : 이 논문은 원격탐사 화상 분류에서 공간적 불확실성 분포를 얻기 위해 공간 통계를 적용하였다. 분류 항목 할당과 참조 자료와 연계된 각각의 불확실성 표현을 위해 2가지 정량적 방법을 제안하였다. 우선 분류 항목 할당에 따른 불확실성 표현을 위해 3가지 정량적 지수를 제안하였다. 그리고 참조 자료와 분류 결과를 결합하고 이와 연계된 불확실성 혹은 정확성 분포를 얻기 위해 지구통계학적 시뮬레이션 기법을 적용하였다. 다중 센서 원격탐사 화상을 이용한 감독 토지 피복 분류 실험을 수행하여 제안 방법을 예시하고, 적용시 제안점을 논의하였다. 실험 결과, 이 논문에서 제시한 방법론을 통해 분류 결과의 해석과 평가를 위한 추가적인 정보 추출이 가능하였으며, 제시 방법론의 검증에 위한 다양한 자료에의 적용이 필요한 것으로 판단된다.

1. 서론

원격탐사 화상 분류를 통한 토지 피복도/이용도 작성은 원격탐사 자료의 여러 응용분야 중에서 많은 연구가 진행되어온 분야중 하나이다. 화상 분류를 통해 작성된 토지 피복도/이용도가 GIS 기본도 갱신 혹은 주제도로서 활용되기 위해서는 우선 보다 향상된 분류 정확도를 제공할 수 있는 분류 기법의 개발이 필요하다. 이와 관련하여 최근에는 분류 정확도 향상을 위해 다중 센서 화상을 이용하거나 확률/통계, 퍼지 이론, 인공지능 등 다양한 이론적 배경을 가지는 분류 기법의 개발이 활발하게 이루어지고 있다(Serpico *et al.*, 1996; Benediktsson and Kanellopoulos, 1999; Bruzzone *et al.*, 1999). 그러나 화상 분류 과정을 원격탐사 화상을 이용하여 직접 조사하지 않은 지역의 토지 피복을 예측하는 과정으로 간주한다면, 분류 기법의 개발과 병행하여 분류 결과의 해석을 위한 불확실성 추정 등과 같은 부가적인 정보 추출에 대한 연구가 필요한 상황이다.

현재까지 원격탐사 화상 분류 결과의 정확도는 훈련집단(training data)과 독립적으로 작성한 참조자료(reference data)를 화상 분류 결과와 비교하여 오차 행렬(error matrix)을 작성한 후에 이로부터 통계치(전체 정확도, 카파 계수 등)를 계산함으로써 얻어 왔다(Congalton, 1991). 그러나 오차 행렬은 화상 전체의 분류 오류를 나타내지만, 화상 분류 결과 정확도의 공간적인 분포를 제공하지는 못한다. 물론 연구 대상지역 전체에 대한 참조 자료가 없는 경우가 대부분이기 때문에 이러한 표본 통계치를 모집단의 통계치로 가정하고 해석을 할 수밖에 없지만, 오차 행렬로부터 얻어진 통계치로부터 화상 분류결과에서 어느 부분이 분류정확도가 낮은지, 혹은 추가적인 지표조사를 필요로 하는 지에 대한 정보를 얻을 수는 없다. 또한 화상 분류의 결과물은 개별 화소에서의 각 토지 피복별 확률값이나 소속함수 값으로부터 결정 규칙에 따라 최종 토지 피복을 나타내는 지도로만 제시되기

때문에, 해석에 유용하게 이용될 수 있는 분류 과정에서 생성되는 확률값이나 소속함수 값 등의 부산물 정보의 손실을 유발하게 된다. 결정 규칙에 의한 각 화소별 최종 토지 피복 피복 결과이외에 적용된 분류 방법에 따른 확률값이나 소속함수값의 특성을 고려하는 일종의 soft 분류로 퍼지 이론 등이 이용되고 있으나(Wang, 1990; Binaghi *et al.*, 1997), 대부분의 연구가 분류 정확도의 향상에 이러한 방법론을 이용하였고 해석 정보로의 활용은 미비한 상황이다. 따라서 최종 사용자의 입장에서 기존 주제도의 갱신이나, 토지 이용도를 입력 자료로 사용하는 여러 응용 분야에 직접적인 적용에는 한계가 있게 된다.

불확실성 분석과 관련하여 지질학, 토양학 분야에서 공간통계, 지구통계학과 연계된 연구가 1990년대부터 진행되고 있으나(Journel, 1996; Goovaerts, 2001), GIS와 원격탐사 분야와의 연계는 최근 중요성이 인식되어 연구가 진행되고 있는 상황이다(Steele *et al.*, 1998; De Bruin, 2000; Foody and Atkinson, 2002).

이 논문의 목적은 원격탐사 화상 분류시 불확실성 혹은 정확도의 공간적 분포를 추정하기 위한 방법론을 제시하고 적용하는데 있다. 이를 위해 불확실성 분포를 2가지 관점에서 분리하여 정의하고, 토지 피복간 구분 능력과 정확도에 대한 공간상의 정량화 정보를 제공하는 지수를 제시하고 이론적 배경을 논의하고자 한다. 우선 원격탐사 화상 분류에서 유발되는 불확실성을 1) 분류 항목 할당에 따른 불확실성과 2) 참조 자료와 연계된 불확실성으로 구분하여 정의하였다. 첫 번째 불확실성 추정을 위해 공간 통계에 기반한 지수를 제시하였고, 두 번째 불확실성 추정을 위해 소수의 참조 자료와 원격탐사 화상으로부터 얻어지는 연구지역 전체에 대한 확률값을 통합할 수 있도록 지구통계학의 시뮬레이션 기법을 적용하였다. 제안 기법의 예시와 적용을 위해 다중센서 원격탐사 화상을 이용한 분류를 목적으로 사례연구를 수행하였다.

2. 공간적 불확실성 분포 추정

1) 분류 항목 할당에 따른 불확실성 추정

이 범주에 속하는 불확실성은 분류 결과 자체에 수반되는 것으로 훈련 집단의 선정, 사용 원격탐사 화상, 분류 기법 등에 의해 좌우된다. 앞으로 제시할 지수들은 사용 자료와 분류 기법 등에 의해 어느 부분이 다양한 토지 피복의 구분 능력이 좋은지 혹은 어느 원격탐사 자료가 포함되었을 때 분류 분해능이 향상되는지를 확인하는데 이용될 수 있다. 따라서 분류 방법의 비교, 다중 센서 자료 이용시 효과를 살펴보는 데 이용할 수 있다. 그러나 Canters(1997)에 의해 논의된 것처럼, 참조 자료와 연관이 되지 않은 상황이기 때문에 아직 정확도(accuracy)라는 용어를 사용할 수는 없다.

연구지역의 특정 화소의 위치를 X 라 하고 모든 화소들을 K 개의 토지 피복 항목 $\{\omega_k, k=1, 2, \dots, K\}$ 으로 분류하는 상황을 가정하자. 특정 원격탐사 화상과 분류 기법을 이용하여 X 에서 모든 분류 항목 ω_k 에 대한 사후 확률 혹은 소속함수 $\{P(\omega_k|X), k=1, 2, \dots, K\}$ 를 얻었다고 할 때, 최종 토지 피복 $\omega(X)$ 의 결정은 식 (1)의 규칙에 따라 이루어지게 된다.

$$\omega(X) = \arg \max_k P(\omega_k|X), k=1, 2, \dots, K \quad (1)$$

여기서 만약 확률론적 분류 기법을 사용하였을 경우에는 $P(\omega_k|X)$ 는 사후 확률을, 퍼지 분류 기법 일 경우에는 소속함수를 각각 나타낸다. 논문의 이후 부분에서 $P(\omega_k|X)$ 는 사후 확률로 대표하여 논의하고자 한다.

보통의 hard 분류의 경우에는 $P(\omega_k|X)$ 에 대한 정보를 식 (1)을 통해 각 화소별로 토지 피복을 할당하는 과정에서만 이용하게 된다. 그러나 $P(\omega_k|X)$ 는 사용 자료와 분류 기법의 분해능을 살펴보는 데 유용한 정보를 제공할 수 있다. 분류 항목 할당의 애매함이 클 경우에는 분류 항목별 사후 확률의 차이가 크지 않을 것이고, 이러한 차이를 정량적으로 표현할 수 있다면 분류 항목 할당

에 따른 불확실성 분포를 얻을 수 있게 된다. 이러한 개념을 이용하여 다음의 3가지 지수를 분류 항목 할당에 따른 불확실성 지수로 정의하였다.

우선 첫 번째 지수로 특정 화소 X 에서의 불확실성을 1에서 최대 사후 확률값을 빼준 값 $\phi(X)$ 으로 정의하였다(식 (2)).

$$\phi(X) = 1 - P_{1st \max}(\omega_k|X) \in [0, 1] \quad (2)$$

여기서 $P_{1st \max}(\omega_k|X)$ 은 특정 화소 X 에서의 최대 사후 확률값을 나타낸다.

만약 특정 화소 X 에서의 최대 사후 확률값이 클 경우, $\phi(X)$ 은 0에 가깝게 되며 주어진 자료와 분류 기법에 의해 분류 항목 할당에 불확실성이 작다고 간주할 수 있다. $\phi(X)$ 는 개념적으로 식 (1)의 분류 항목 할당 결정 규칙을 적용하는 과정에서 완전한 소속도를 가정하는데 따르는 일종의 위험(risk) 정도를 나타내는 지수에 해당된다.

두 번째 불확실성 표현 지수로 정보이론에 근거하여 확률변수의 불확실성에 대한 척도로 사용될 수 있는 엔트로피 $H(X)$ 를 이용하였다(식 (3)).

$$H(X) = - \frac{\sum_{k=1}^K \ln[P(\omega_k|X)] \cdot P(\omega_k|X)}{\ln K} \in [0, 1] \quad (3)$$

식 (2)의 $\phi(X)$ 와 마찬가지로 불확실성이 작은 화소에서는 0에 가까운 값을 나타내게 된다. 만약 모든 분류 항목들의 사후 확률(소속함수)이 같을 때 최대값인 1을 가지며, 한 분류 항목만이 1이고 나머지는 0일 때 최소값 0을 가진다. $H(X)$ 는 $\phi(X)$ 와는 달리 다른 분류 항목의 사후 확률값을 고려한 분류 항목 할당에 따르는 위험 정도를 나타내는 지수에 해당된다. Zhu(1997)은 퍼지 논리의 이론적 틀 안에서 퍼지 유사도 개념을 이용하여 분류 항목 할당에 따른 불확실성을 과장(exaggeration)과 무시(ignorance)로 구분하여 두 가지 지수를 제안하였는데, 과장 지수는 식 (2)의 $\phi(X)$ 에, 무시 지수는 식 (3)의 $H(X)$ 에 각각 해당한다고 간주할 수 있다. 그러나 실제로 soft 분류의 개념을 사용한다면 식 (2)와 (3)의 지수는 확률이론의 분류 기법에도 적용이 가능하다.

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">1 $\omega(1) = 1$</td> <td style="width: 50%; padding: 5px;">2 $\omega(2) = 1$</td> </tr> <tr> <td style="width: 50%; padding: 5px;">3 $\omega(3) = 1$</td> <td style="width: 50%; padding: 5px;">4 $\omega(4) = 2$</td> </tr> </table>	1 $\omega(1) = 1$	2 $\omega(2) = 1$	3 $\omega(3) = 1$	4 $\omega(4) = 2$	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">1 $\omega_1 = 0.8$ $\omega_2 = 0.2$</td> <td style="width: 50%; padding: 5px;">2 $\omega_1 = 0.99$ $\omega_2 = 0.01$</td> </tr> <tr> <td style="width: 50%; padding: 5px;">3 $\omega_1 = 0.6$ $\omega_2 = 0.4$</td> <td style="width: 50%; padding: 5px;">4 $\omega_1 = 0.45$ $\omega_2 = 0.55$</td> </tr> </table>	1 $\omega_1 = 0.8$ $\omega_2 = 0.2$	2 $\omega_1 = 0.99$ $\omega_2 = 0.01$	3 $\omega_1 = 0.6$ $\omega_2 = 0.4$	4 $\omega_1 = 0.45$ $\omega_2 = 0.55$	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%; padding: 5px;">1 $\phi(1) = 0.2$ $H(1) = 0.72$ $\Delta(1) = 0.6$</td> <td style="width: 50%; padding: 5px;">2 $\phi(2) = 0.01$ $H(2) = 0.08$ $\Delta(2) = 0.98$</td> </tr> <tr> <td style="width: 50%; padding: 5px;">3 $\phi(3) = 0.4$ $H(3) = 0.97$ $\Delta(3) = 0.2$</td> <td style="width: 50%; padding: 5px;">4 $\phi(4) = 0.45$ $H(4) = 0.99$ $\Delta(4) = 0.1$</td> </tr> </table>	1 $\phi(1) = 0.2$ $H(1) = 0.72$ $\Delta(1) = 0.6$	2 $\phi(2) = 0.01$ $H(2) = 0.08$ $\Delta(2) = 0.98$	3 $\phi(3) = 0.4$ $H(3) = 0.97$ $\Delta(3) = 0.2$	4 $\phi(4) = 0.45$ $H(4) = 0.99$ $\Delta(4) = 0.1$
1 $\omega(1) = 1$	2 $\omega(2) = 1$													
3 $\omega(3) = 1$	4 $\omega(4) = 2$													
1 $\omega_1 = 0.8$ $\omega_2 = 0.2$	2 $\omega_1 = 0.99$ $\omega_2 = 0.01$													
3 $\omega_1 = 0.6$ $\omega_2 = 0.4$	4 $\omega_1 = 0.45$ $\omega_2 = 0.55$													
1 $\phi(1) = 0.2$ $H(1) = 0.72$ $\Delta(1) = 0.6$	2 $\phi(2) = 0.01$ $H(2) = 0.08$ $\Delta(2) = 0.98$													
3 $\phi(3) = 0.4$ $H(3) = 0.97$ $\Delta(3) = 0.2$	4 $\phi(4) = 0.45$ $H(4) = 0.99$ $\Delta(4) = 0.1$													
(a)	(b)	(c)												

Fig. 1. Example of computing the degree of spatial uncertainty by class assignment in the two-class problem(class 1 - grey color, class 2 - white color); (a) land-cover classes assigned as a result of classification, (b) posteriori probability values for two classes ω_1 and ω_2 , (c) three indices that describe the degree of spatial uncertainty.

마지막 지수로 퍼지 개념을 이용하여 최대 사후 확률을 보이는 분류 항목과 다음 최대 사후 확률을 보이는 분류 항목의 확률값의 차이 $\Delta(X)$ 를 정의하였다(식 (4)).

$$\Delta(X) = P_{1st\ max}(\omega_k|X) - P_{2nd\ max}(\omega_k|X) \in (0, 1] \quad (4)$$

만약 분류 항목 할당이 애매할 경우 $\Delta(X)$ 는 0에 근접한 값을 나타낼 것이고, 명확하게 분류가 될 경우에는 1에 근접한 값을 나타내게 된다.

Fig. 1은 2개의 토지 피복 분류 문제에서 불확실성 표현 지수를 예시한 것이다. Fig. 1(b)의 사후 확률값으로부터 Fig. 1(c)의 불확실성 지수의 계산함으로써 Fig. 1(a)의 분류 항목 할당의 질(quality)을 판단할 수 있다. $\phi(X)$ 와 $H(X)$ 작을수록, $\Delta(X)$ 는 클수록 분류 항목 할당에 따른 불확실성이 작다고 간주할 수 있는데, 이 예시에서 화소 3과 화소 4는 각각 클래스 1과 2로 분류가 되었지만 불확실성 표현 지수의 계산 결과 클래스 1과 2로의 구분이 애매함을 알 수 있다. 개별 화소에 대해서 이러한 지수들을 계산함으로써 연구 지역 전체의 분류 항목 할당에 따른 불확실성 분포를 얻게 된다. 그러나 여기서 유의할 점은 각각의 지수들이 계산과정에서 상대적인 값의 차이를 나타내는 정도가 다르기 때문에 해석에서는 모든 지수를 함께 고려해야할 필요가 있다. $H(X)$ 는 식 (3)

에 나타난 것처럼 분류 항목의 수에 의존하게 되는데, 분류 항목의 수가 증가할수록 값이 작게 나타나게 된다. 화소 1의 경우 $\phi(1)$ 과 $\Delta(1)$ 은 불확실성이 상대적으로 낮게 나오지만, $H(1)$ 은 2개의 분류 문제에서는 상대적으로 크게 나타나고 있다. 따라서 다른 지수와 함께 값을 고려하고, 절대값보다는 상대적 비교를 수행하는 것이 실제 적용시 필요하다.

2) 참조 자료와 연관된 불확실성 추정

첫 번째 범주의 불확실성이 분류 항목 할당에 따른 분류 결과의 질을 판단할 수 있는 정보를 제공한다면, 두 번째 범주의 불확실성은 참조 자료와 연계된 정보를 제공한다. 여기서 참조 자료와 연계된다는 의미는 분류 결과로 제시되는 사후 확률을 참조 자료와 결합한다는 것으로 실제 지표 상황과 연관 짓기 때문에 일종의 정확도의 의미를 내포하게 된다. 따라서 혼련집단에 대해 부트스트랩을 적용하여 얻어진 통계치를 지시자 크리깅(indicator kriging)을 수행하여 분류 자체의 공간적 불확실성 분포를 얻은 Steele et al.(1998)과는 접근이 다르다.

이 논문에서는 지구통계학적 시뮬레이션(simulation) 기법을 이용하여 사후 확률과 참조 자료를 연관시키고자 하였다. 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 번이를 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). 그러나 지구통계학이 기본적으로 sparse한 자료의 공간적 분포를 추정하기 때문에 연구지역에 대한 전체 자료를 획득할 수 있는 원격탐사 분야에는 많이 적용되지 못했다. 1990년대 들어오면서 그룹으로 인한 지표의 미관측지점의 값을 추정하기 위해 크리깅을 이용한 연구가 있었으며(Rossi et al.,

1994), 1990년대 중후반부터 sparse한 지표 조사자료와 원격탐사자료를 통합하는 사례연구가 진행되고 있는 상황이다(Dungan, 1998; De Bruin, 2000). 또한 최근에는 Goovaerts(2002)와 Park et al(2003)에 의해 토지 피복 분류 과정에 공간적 자기 상관성을 고려하기 위해 지시자 크리깅을 적용한 연구들이 수행되었다.

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 되

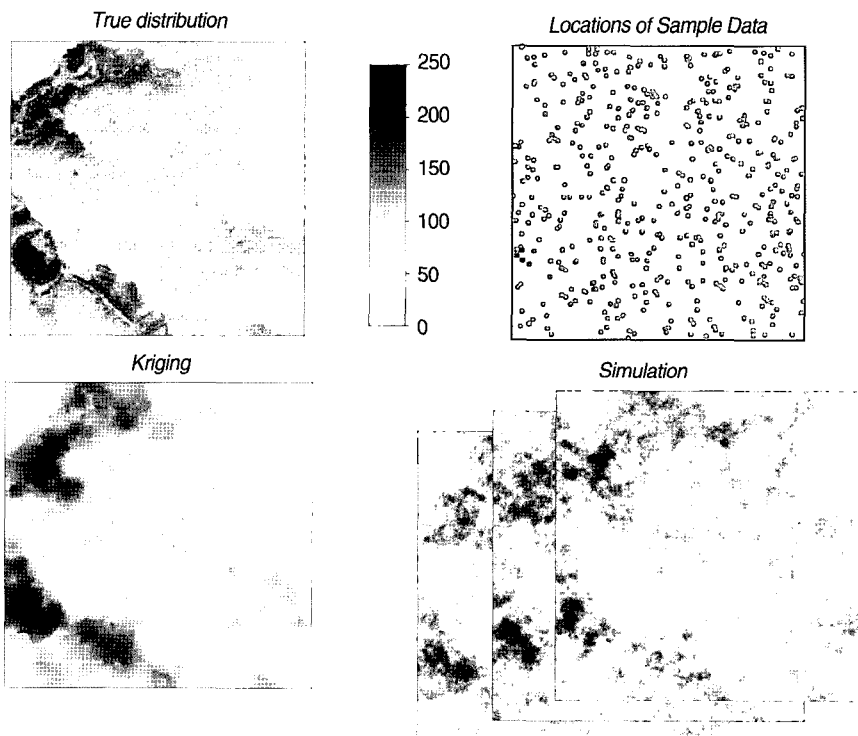


Fig. 2. Comparison of kriging and simulation algorithms(modified from Journel(1996)).

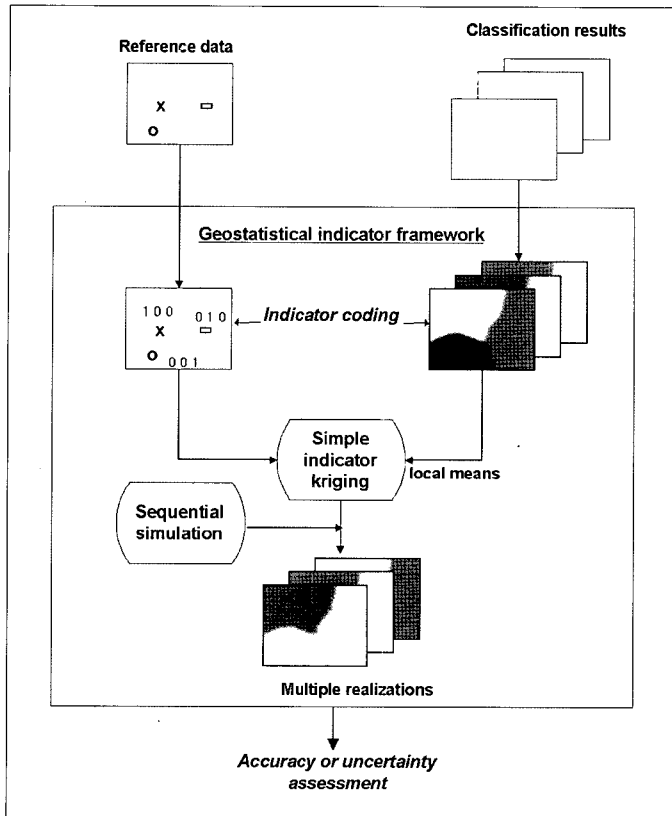


Fig. 3. Schematic view of SISLM.

며, 원래 자료의 배리오그램이나 히스토그램을 재 생산할 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. 2). 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 논문에서는 여러 시뮬레이션 기법 중에서 범주형 자료와 부가자료를 함께 고려할 수 있는 Sequential Indicator Simulation with Local Means (SISLM)를 적용하였다(Fig. 3). Sequential

Indicator Simulation(SIS)은 우선 지시자 변환(indicator transform)을 수행한 후에 주어진 공분산 함수를 만족하면서 매단계에서 시뮬레이션된 자료를 다음 계산 단계에서는 알려진 자료로 가정하고 시뮬레이션을 모든 화소위치에서 순차적으로 진행한다(Goovaerts, 1997).

SIS의 실제 수행단계는 다음과 같다.

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

한다.

연구지역에서 K 개의 토지 피복 항목 $\{\omega_k, k=1, 2, \dots, K\}$ 과 n 개의 참조 자료를 고려할 때, 참조 자료 위치 $\{X_\alpha, \alpha=1, \dots, n\}$ 에서의 지시자 변환을 다음과 같이 수행한다.

$$I(X_\alpha; \omega_k) = \begin{cases} 1 & X_\alpha \in \omega_k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

참조 자료가 없는 위치 X 에서 토지피복 항목 ω_k 의 발생 확률은 다음의 지시자 크리깅 시스템에 의해 결정된다.

$$I(X; \omega_k) = \sum_{\alpha=1}^n \lambda_\alpha(X_\alpha; \omega_k) [I(X_\alpha; k) - p_k] + p_k \quad (6)$$

여기서 p_k 는 k 번째의 토지 피복 항목의 초기 확률을 나타낸다.

SIS와 달리 SISLM은 지시자 크리깅을 수행할 때, 전체 평균으로 주어지는 즉 위치와 무관하게 주어지는 초기 확률 p_k 를 위치와 관계된 local means인 $p_k(X)$ 으로 대체함으로써 분류 결과와 참조 자료를 연관 지을 수 있다. Kyriakidis and Dungan(2001)의 연구에서는 오차 행렬로 계산되는 연구 지역 전체의 global statistics의 일종인 사용자 정확도를 $p_k(X)$ 로 이용하였는데, 이 경우 국소적으로 상이한 분류 확률과의 직접적인 연관성이 떨어지게 된다. 대신 이 연구에서는 일종의 soft 분류 결과 얻어지는 모든 화소 X 에서 모든 분류 항목 ω_k 에 대한 사후 확률 $\{P(\omega_k|X), k=1, 2, \dots, K\}$ 을 $p_k(X)$ 로 사용하였다(식 (7)). 만약 퍼지 이론을 적용하였을 경우에는 모든 화소에서의 소속함수값의 합이 1이 되도록 정규화 과정을 미리 수행해야 한다.

$$I(X; \omega_k) = \sum_{\alpha=1}^n \lambda_\alpha(X_\alpha; \omega_k) [I(X_\alpha; k) - p(\omega_\alpha|X)] + p(\omega_\alpha|X) \quad (7)$$

식 (7)에서 크리깅 가중치 $\lambda_\alpha(X_\alpha; \omega_k)$ 는 식 (8)을 계산하여 얻어진다.

$$\sum_{\beta=1}^n \lambda_\beta C_I(X_\beta - X_\alpha; \omega_k) = C_I(X - X_\alpha; \omega_k), \alpha = 1, \dots, n \quad (8)$$

여기서 $C_I(X - X_\alpha; \omega_k)$ 는 ω_k 의 잔차(residual)인 $I(X_\alpha; \omega_k) - p(\omega_k|X)$ 의 공분산을 나타낸다.

식 (7)과 (8)을 이용하여 L 번의 시뮬레이션을

수행하였으면 모든 화소에서는 L 개의 우세한 토지 피복 항목을 가지게 되는데, 이들로부터 최종적인 개별 토지피복 항목에 대한 분류 정확도의 분포값 $\text{Pr}^*(X; \omega_k)$ 는 식 (9)를 이용하여 얻을 수 있다.

$$\text{Pr}^*(X; \omega_k) = \frac{1}{L} \sum_{l=1}^L I^{(l)}[X; \omega_k] \quad (9)$$

여기서 $I^{(l)}[X; \omega_k]$ 은 식 (5)와 마찬가지로 X 에서의 토지 피복 항목의 지시자 변환을 나타내며 고려하는 토지 항목과 일치하면 1을, 일치하지 않으면 0의 값을 갖게 된다. 불확실성값은 1에서 식 (9)를 통해 얻어진 확률값을 빼줌으로써 얻을 수 있다.

만약 특정 위치에서 분류정확도가 높으면 여러 번의 시뮬레이션의 결과(multiple realizations)에서 최종 결정된 분류항목이 많이 나타나야 한다. 반면, 여러 항목이 시뮬레이션 결과에서 나타난다면 분류정확도가 상대적으로 낮다고 간주할 수 있다. 예를 들어 5개의 토지 피복 항목($\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$)을 고려할 때 100번의 시뮬레이션을 통해 특정 화소 X 에서 각 토지 피복 항목이 나타난 회수가 70, 15, 0, 5, 10일 경우, 개별 토지 피복에 대한 확률값은 각각 0.7, 0.15, 0.0, 0.05, 0.1로 얻게 된다. 최대 확률값이 1에 가까울수록 불확실성이 작음을 의미한다.

앞에서 언급한 것처럼 시뮬레이션 결과 얻게 되는 공간적 불확실성 혹은 정확성 분포는 분류 결과 확률값으로부터 계산할 수 있는 불확실성 분포와는 개념적으로 다르다. 후자는 분류 모델 자체에 대한 불확실성 분포를 제공하지만, 시뮬레이션 결과는 분류 모델 자체의 불확실성과 참조 자료의 공간적 상관성이 함께 결합된 정보를 제공하게 된다.

3. 사례 연구

1) 사용 자료

제안 기법의 예시와 적용을 위해 다중센서 원격탐사 화상을 이용한 분류를 목적으로 사례 연구를

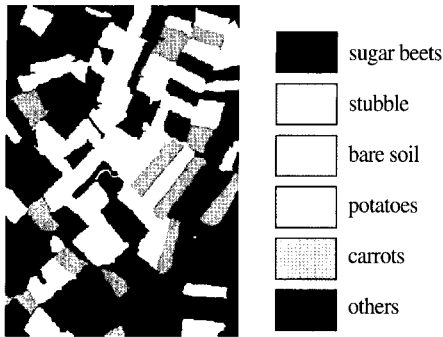


Fig. 4. Ground truth map in the study area. The areas shown in black (others class) are not considered in this study.

수행하였다. 사례 연구에는 박노옥 외(2004)에서 이용한 IEEE Geoscience and Remote Sensing Society의 Data Fusion Committee의 기하보정과 필터링 등의 기본 전처리가 수행된 ATM(Airborne Thematic Mapper) 광학 센서 자료 6개 밴드와 NASA JPL

polarimetric AIRSAR C, L, P 밴드의 다중 편광(HH, HV, VV)으로 구성된 9개의 밴드를 이용하였다. 연구지역의 토지 피복 항목으로는 sugar beets, stubble, bare soil, potatoes, carrots 등 5개의 토지 피복 항목을 고려하였다. 또한 연구지역 전체의 토지피복도를 분류결과와 분석을 위한 보조 자료로 이용하였다 (Fig. 4). 연구지역은 250×350 화소로 구성되어 있는데, 5개의 토지 피복 항목만으로 구성된 총 55,657 화소를 고려하였으며, 제외된 화소들은 Fig. 4에서 검은 색으로 표시되어 있다. 감독분류를 위한 훈련 집단과 참조 자료는 Fig. 4로부터 각 피복별로 임의 추출을 통해 서로 독립적으로 작성하였다.

2) 적용 결과

다중 센서 화상의 조합은 1) 광학 자료만을 이용한 경우, 2) SAR 자료만을 사용한 경우(C-, L- and P-band)와 3) 모든 자료를 고려한 경우 총 3

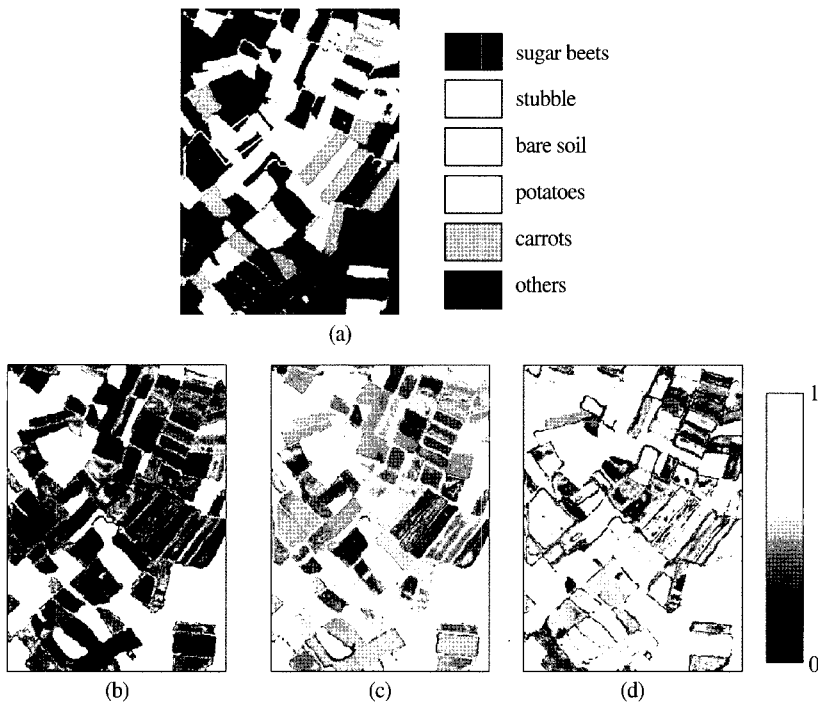


Fig. 5. Classification and uncertainty distribution results using ATM data; (a) classification map, (b) $\phi(X)$, (c) $H(X)$, (d) $\Delta(X)$. The areas with the black color (others class) in (a) are ones which are not considered in this study. In (b), (c) and (d), the areas which are not considered in this study are shown in white.

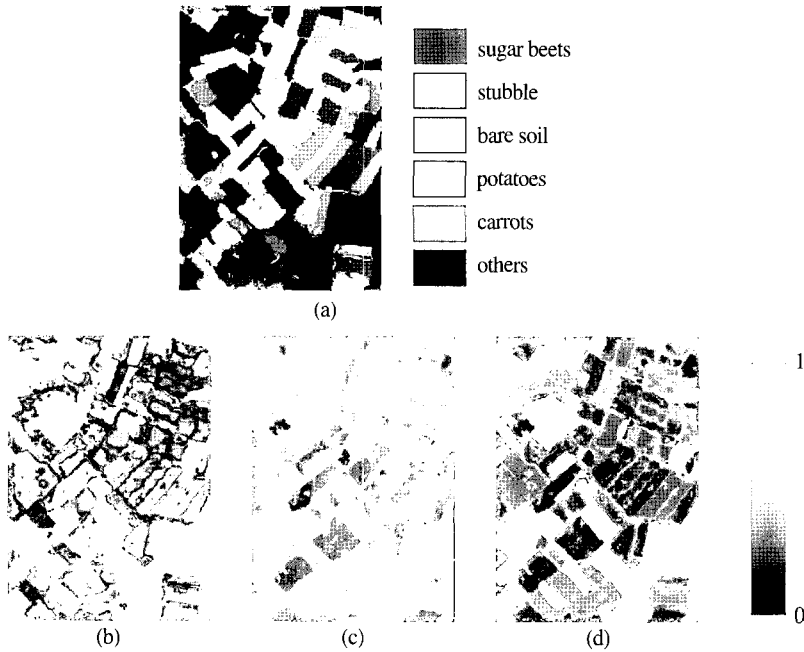


Fig. 6. Classification and uncertainty distribution results using multi-frequency SAR data; (a) classification map, (b) $\phi(X)$, (c) $H(X)$, (d) $\Delta(X)$. The areas with the black color (others class) in (a) are ones which are not considered in this study. In (b), (c) and (d), the areas which are not considered in this study are shown in white.

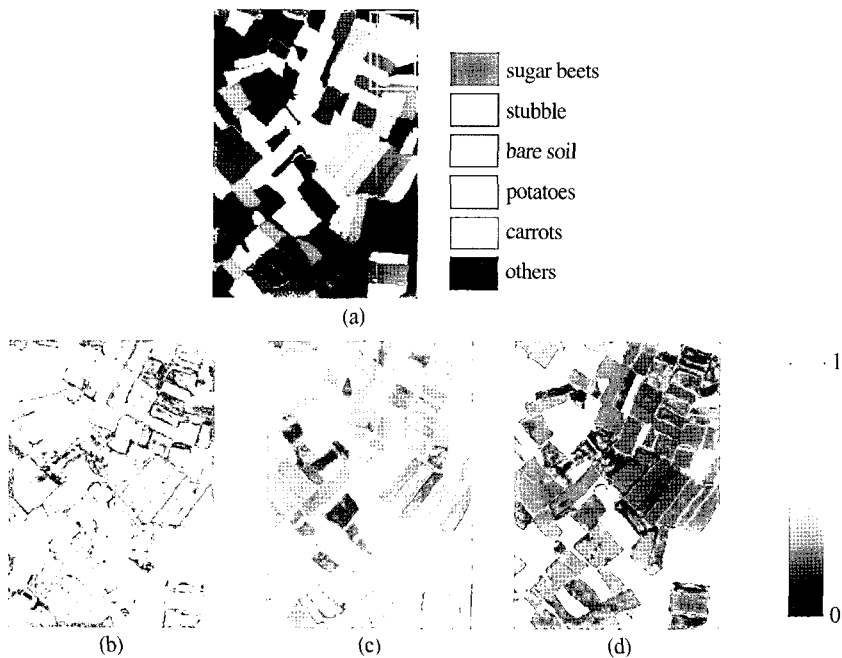


Fig. 7. Classification and uncertainty distribution results using all data; (a) classification map, (b) $\phi(X)$, (c) $H(X)$, (d) $\Delta(X)$. The areas with the black color (others class) in (a) are ones which are not considered in this study. In (b), (c) and (d), the areas which are not considered in this study are shown in white.

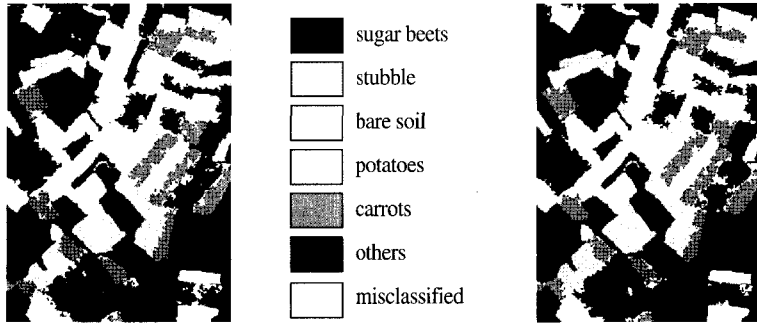


Fig. 8. Two realizations generated by SISLM. The areas with the black color (others class) are ones which are not considered in this study. The misclassified pixels shown in white were generated by comparing each realization with the ground truth map as shown in Fig. 4.

가지를 고려하였다. 분류 기법으로는 확률론적 인공 신경망(probabilistic neural network: Specht, 1990)을 적용하였다.

우선 광학자료인 ATM의 6개 밴드만을 이용하였을 경우(Fig. 5), $\phi(X)$ 과 $H(X)$ 는 토지 피복간의 경계부분에서 높은 값을 나타내어 불확실성이 큰 것으로 나타났다. 또한 오분류된 화소에서는 상대적으로 $\phi(X)$ 과 $H(X)$ 이 크게 나타나고 있으며, $\Delta(X)$ 는 작게 나타나서 분류가 애매한 것으로 나타났다. 그러나 Fig. 5의 박스로 표현된 연구지역 좌측 중앙의 stubble이 bare soil로 오분류된 지역은 $\phi(X)$ 과 $H(X)$ 는 작은 값을, $\Delta(X)$ 는 큰 값을 나타내어 오분류되었음에도 불구하고 오히려 불확실성이 작게 나타나고 있다. 이러한 오분류 결과는 연구지역 좌측 중앙 지역이 stubble보다는 bare soil의 ATM에서의 분광 특성을 대표하여 나타난 것으로 해석할 수 있다. 앞에서 언급한 것처럼 지표의 참조 자료와 아직 연계되지 못했기 때문에 실제 오분류 여부는 이러한 불확실성 지수만으로는 판단이 힘들며, 또한 불확실성이 작다고 반드시 분류 정확도가 높을 것으로 단정지어서는 안될 것으로 판단된다.

C, L, P 밴드 다중 편광 자료만을 이용한 결과를 살펴보면(Fig. 6), 대체로 오분류 지역에서 불확실성이 높은 것으로 나타나고 있다. 즉 $\phi(X)$ 과 $H(X)$ 는 상대적으로 크게, $\Delta(X)$ 는 작게 나타나고

있다. 또한 Fig. 5의 광학자료만을 이용하였을 때에 비해, 경계부분에서의 불확실성이 큰 것으로 나타났으며 각 불확실성 지수별로 값 차이가 크게 나타나고 있다. 광학자료만을 이용하였을 때, 오분류로 나타난 연구지역 좌측 중앙의 stubble은 다중 편광 자료에서는 맞게 분류된 것으로 나타났다.

모든 자료를 다 분류에 이용한 결과를 살펴보면(Fig. 7), 전반적인 양상은 Fig. 5와 Fig. 6의 양상이 혼재된 결과를 보이면서 값 차이가 커짐을 확인할 수 있다. 값 차이가 커졌다는 사실은 특정 분류 항목에서 불확실성이 자료 융합의 결과로 작아졌음을 의미한다. 여러 분류 항목 중에서 Fig. 7의 박스에 표시된 potatoes 항목에서 자료 융합의 효과가 가장 크게 나타나서 불확실성이 상대적으로 작게 나타나고 있다. 연구지역 좌측 중앙의 stubble은 자료 융합의 효과로 맞게 분류되었다. 광학자료에 의해 불확실성이 작지만 오분류된 이 지역이 SAR 자료와의 융합시 개별 자료 이용에 비해 불확실성이 작은 지역이 상대적으로 줄어들었지만 SAR 자료의 영향으로 맞게 분류된 것으로 판단된다.

세 가지 자료 결합의 경우 모두 공통적으로 세 가지 지수의 전반적인 양상은 유사하게 나타나서 각 지수별 상호보완적 해석 정보를 나타내지는 못하였다. 다만 $H(X)$ 와 $\Delta(X)$ 에 비해 $\phi(X)$ 이 상대적으로 값 차이가 작게 나타나고 있다. 값 차이가 작게 나타나는 경우에는 상대적인 비교가 힘든 경우

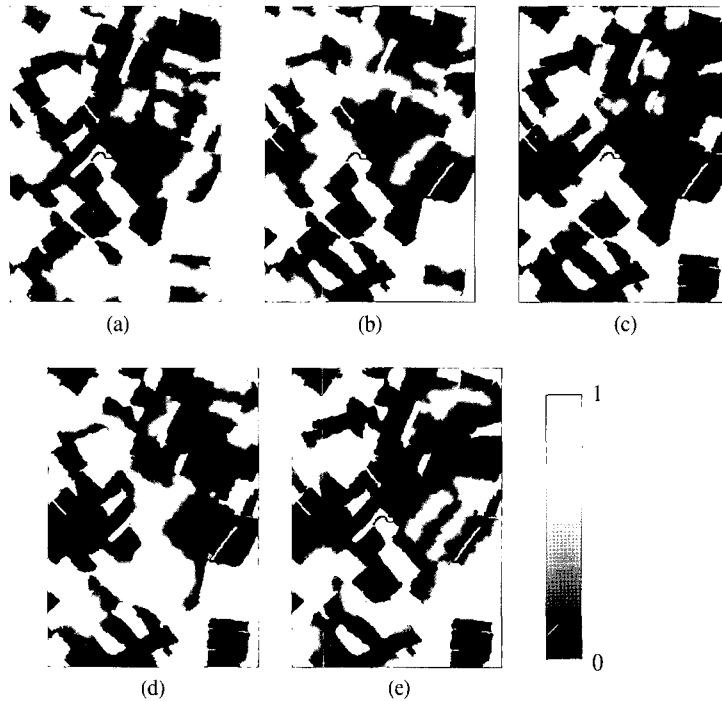


Fig. 9. Spatial accuracy probability distribution for each land-cover class from 100 simulated land-cover maps; (a) sugar beets, (b) stubble, (c) bare soil, (d) potatoes, (e) carrots. The areas which are not considered in this study are shown in white.

가 발생할 수 있기 때문에 상대적으로 큰 값차이를 나타내는 $\Delta(X)$ 가 사례연구지역의 자료 해석에는 보다 효율적인 것으로 판단된다.

참조 자료와 연계된 불확실성 추정을 위해 GSLIB(Deutsch and Journel, 1998)를 이용하여 SISLM을 수행하였다. SISLM 수행시 local means로 확률론적 인공 신경망 기법을 통해 얻어진 각 분류항목별 사후 확률을 이용하였다. 이 기법의 예시를 위해 박노옥 외(2004)에서 가장 높은 분류 정확도를 나타낸 모든 자료를 이용한 경우에 적용하였다. 우선 SISLM을 적용하기 전에, 크리깅 수행시 필요한 가정인 unbiasedness 조건을 만족시키기 위해 다중 로지스틱 회귀 분석을 통해 local means를 참조 자료에 대해 calibration을 수행하였다. 그리고 simple indicator kriging with local means를 수행하여 각 분류 항목별로 참조 자료와 연계된 조건부 확률을 계산하였다. 이 후에 서로 다른 임의의 경로를 가지는 100번의 시뮬레이션을 수행하였다.

Table 1. Class proportion statistics computed from 100 realizations. Reference indicates the ground truth map shown in Fig. 4 and the reference data set.

	Sugar beets	Stubble	Bare soil	Potatoes	Carrots
Reference	0.27	0.24	0.09	0.22	0.18
Min	0.26	0.23	0.09	0.21	0.16
Max	0.29	0.26	0.11	0.23	0.18
Average	0.27	0.25	0.10	0.22	0.17

Fig. 8은 100번의 시뮬레이션 중에서 1번째와 2번째 결과를 나타낸 것이다. Fig. 4의 토지 피복도와 비교하였을 때, 하얀색으로 표시된 오분류 화소들은 주로 토지 피복 항목간의 경계부분에 위치함을 확인할 수 있다. 경계부분에서 나타나는 패턴은 참조 자료의 공간적 배열과 경계부분에서의 토지 피복간 혼재나 피복 구분의 애매함 등으로 인해 시뮬레이션 결과 불규칙하게 나타난 것으로 판단된다. 그러나 Fig. 7의 분류 결과에서 부분부분 나타나는 고립된 화소들이 별로 나타나지 않고

있는데, 이는 각 화소와 근접한 참조 자료와의 공간적 상관성을 시뮬레이션 과정에서 배리오그램 모델링을 통해 고려하였기 때문이다.

Fig. 9는 5개의 토지 피복 항목별로 식 (9)를 이용하여 얻어진 참조 자료와 연관된 정확성 (accuracy) 분포를 나타낸다. 1에 근접한 값은 반대로 불확실성이 높음을 의미하며, 0에 근접한 값은 불확실성이 낮음을 의미한다. Fig. 8의 결과로부터 유추할 수 있는 경계부분에서의 낮은 정확성값 즉 높은 불확실성값을 나타내고 있다. 경계부분에서의 변화는 주로 근접한 토지 피복간의 혼재나 농작물이 아닌 다른 토지 피복(나무, 농로 등)의 영향으로 인해 나타난 결과로 해석될 수 있으며, 이는 결과적으로 참조 자료와 연관된 불확실성이 높음을 의미한다. Fig. 8과 Fig. 9의 결과는 참조 자료와 연계된 정확성 혹은 불확실성의 공간적 분포 이외에도 훈련집단을 이용하여 작성된 토지 피복도를 참조 자료와 연관지음으로써 새로이 갱신된 토지 피복 정보를 제공한다. 기존 연구에서는 오차 행렬을 계산할 때에만 참조 자료를 사용하였지만, 이 연구의 결과는 참조 자료의 속성과 공간적 분포를 분류 결과와 직접적으로 연관지어서 얻게 된다.

Table 1은 100번의 시뮬레이션 결과로부터 각 토지 피복 항목의 비율과 관련된 통계치를 정리한 것이다. 여기서 참조 자료의 각 토지 피복 항목의 비율은 실제 토지 피복도와 거의 동일하게 작성되었다. 결과를 비교하면, 각 시뮬레이션을 통해 얻어진 Fig. 8과 같은 토지 피복도는 실제 토지 피복도와 참조 자료에서의 각 토지 피복 항목의 비율을 거의 정확하게 나타내고 있음을 확인할 수 있다. 이러한 결과는 참조 자료의 통계치를 재생산하면서 참조 자료와의 공간적 상관성을 고려하여 미지의 값을 예측하는 시뮬레이션의 특성을 그대로 반영한 것이다.

4. 결론

이 연구에서는 원격탐사 자료의 분류 결과의

평가시 기존 오차 행렬로부터 얻어지는 통계치 이외에 공간적 불확실성 분포를 얻기 위해 공간통계에 기반을 둔 방법론을 제시 및 적용하였다. 분류시 불확실성을 분류 항목 할당에 따른 불확실성과 참조 자료와 연계된 불확실성으로 구분하여 정의하고 각각의 방법론을 제시하였다. 제안 방법론을 다중 센서 원격탐사 화상의 분류에 적용한 결과, 분류 결과를 해석할 수 있는 보조 정보, 참조 자료의 공간적 분포와 연관된 정확성 혹은 불확실성 분포 이외에 갱신된 분류 결과를 얻을 수 있었다. 이러한 결과들은 분류 결과의 질을 판단하는데 뿐만 아니라, 추가 지표 조사를 필요로 하는 지역의 추출에도 이용이 가능할 것으로 기대된다.

분류 항목 할당에 따른 불확실성 분포를 얻기 위해 제안된 세 가지 지수는 사례 연구에서는 큰 차이를 나타내지 않았지만, 개별 지수에서 얻기 힘든 상호 보완적인 정보를 제공할 수 있기 때문에 다른 자료에 적용할 때에는 모든 지수를 함께 해석하는 것이 나을 것으로 판단된다. 지구통계학적 시뮬레이션을 통해 얻어진 다중 결과로부터 정확성 혹은 불확실성 분포에 관한 확률값을 계산하였지만, 실제로는 Fig. 8처럼 개별 시뮬레이션 결과 자체에도 의미가 있다. 예를 들어 토지 피복도를 입력 자료로 사용하는 토양 유실, 경관 모델링 등에 시뮬레이션을 통해 얻어진 다수의 분류 결과를 입력자료로 활용함으로써 하나의 결과값이 아닌 결과값들의 분포를 얻을 수 있다. 이렇게 얻어진 분포는 각 응용에서의 가변성을 설명하거나 입력 자료의 모델에 대한 오차 전파를 추정하는 데에도 이용이 가능한데, 이러한 응용 목적으로 연구를 진행해 나갈 예정이다. 또한 다양한 자료를 대상으로한 사례 연구를 통해 제안 방법의 적용성 검증을 수행할 예정이다.

사 사

이 연구에 사용한 다중 센서 화상은 IEEE Geoscience and Remote Sensing Data Fusion

Committee(grss-dfc-0006)로부터 제공받았으며 이에 감사드립니다. 이 연구는 “위성자료 공공활용 연구”의 일환으로 한국지질자원연구원에서 운영 중인 공공원격탐사 지상센터의 기반 연구의 일부로 수행되었습니다.

참고문헌

- 박노욱, 지광훈, 권병두, 2004. 퍼지 논리 융합과 반복적 relaxation labeling을 이용한 다중 센서 원격탐사 화상 분류. 대한원격탐사학회지, 20(4): 275-288.
- Benediktsson, J. A. and I. Kanellopoulos, 1999. Classification of multisource and hyperspectral data based on decision fusion, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3): 1367-1377.
- Binaghi, E., I. Gallo, M. Pepe, P. A. Brivio, S. Musazzi, and A. Bassini, 1997. Fuzzy contextual classification of multisource remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 35(2): 326-340.
- Bruzzone, L., D. F. Prieto, and S. B. Serpico, 1999. A neural-statistical approach to multitemporal and multisource remote-sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3): 1350-1359.
- Canter, F., 1997. Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 63(4): 403-414.
- Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37(1): 35-46.
- De Bruin, S., 2000. Predicting the areal extent of land-cover types using classified imagery and geostatistics, *Remote Sensing of Environment*, 74(3): 387-396.
- Deutsch, C. V. and A. G. Journel, 1998. *GSLIB: Geostatistical software library and user's guide*, Oxford University Press, New York.
- Dungan, J. L., 1998. Spatial prediction of vegetation quantities using ground and image data, *International Journal of Remote Sensing*, 19(2): 267-285.
- Foody, G. M. and P. M. Atkinson, 2002. *Uncertainty in remote sensing and GIS*, John Wiley & Sons, Chichester, England.
- Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*, Oxford University Press, New York.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science, *Geoderma*, 103(1): 3-26.
- Goovaerts, P., 2002. Geostatistical incorporation of spatial coordinate into supervised classification of hyperspectral data, *Journal of Geographical Systems*, 4(1): 99-111.
- Journel, A. G., 1996. Modelling uncertainty and spatial dependence, *International Journal of Geographical Information Systems*, 10(5): 517-522.
- Kyriakidis, P. C. and J. L. Dungan, 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(4): 311-300.
- Park, N.-W., K.-H. Chi and B.-D. Kwon, 2003. Geostatistical integration of spectral and spatial information for land-cover mapping using remote sensing data, *Geosciences Journal*, 7(4): 335-341.
- 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(1): 32-40.
- Serpico, S. B., L. Bruzzone, and F. Roli, 1996. An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images, *Pattern Recognition Letters*, 17(13): 1331-1341.
- Specht, D. F., 1990. Probabilistic neural network, *Neural Networks*, 3(1): 109-118.
- Steele, B. M., J. C. Winne, and R. L. Redmond, 1998. Estimating and mapping of misclassification probabilities for thematic land cover maps, *Remote Sensing of Environment*, 66(2): 192-202.
- Wang, F., 1990. Fuzzy supervised classification of remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, 28(2): 194-201.
- Zhu, A. X., 1997. Measuring uncertainty in class assignment for natural resource maps under fuzzy logic, *Photogrammetric Engineering & Remote Sensing*, 63(10): 1195-1202.