Application of Geostatistical Indicator Simulation to Assessment of Spatial Uncertainty Distributions in Remote Sensing Data Classification

Kwang-Hoon Chi
Korea Institute of Geoscience and Mineral Resources
30 Gajeong, Yuseong-gu, Daejeon 305-350, Korea
khchi@kigam.re.kr

No-Wook Park
Korea Institute of Geoscience and Mineral Resources
30 Gajeong, Yuseong-gu, Daejeon 305-350, Korea
nwpark@kigam.re.kr

Abstract: This paper applied geostatistical indicator simulation to obtain the spatial accuracy probability in remote sensing data classification. Among various geostatistical simulation algorithms, sequential indicator simulation with local means was adopted. By applying this algorithm, the spatial accuracy distribution can be obtained from many realizations of land-cover class labels at each pixel. This scheme was applied with the case study of supervised land-cover classification with multi-sensor remote sensing data.
Keywords: Geostatistical Simulation, Classification.

1. Introduction

In remote sensing data classification, the quality of classification results is traditionally given in terms of accuracy statistics (e.g. overall accuracy, kappa statistic) derived from a confusion matrix. This accuracy measure only provides a global statistics, but fails to give any spatial information on classification accuracy or uncertainty. If we regard the remote sensing data classification as the predictions of ground properties or attributes at unsampled locations, spatial distributions of accuracy or uncertainty should be considered.

In relation to the processing of spatial data, geostatistics provides quantitative descriptions of variables distributed in space and/or time [2]. These days, geostatistics is increasingly used to infer the local and spatial uncertainty and integrate various types of data. Especially, stochastic simulation, which can generate multiple alternative realizations of the spatial distribution of attributes prevailing over the study area, can be applied to assess spatial accuracy distributions. Recently, sequential indicator simulation was applied to map thematic classification accuracy and to evaluate the impact of inaccurate spatial data on ecological model predictions [4]. They integrated sparse ground samples with user's accuracy and finally spatial accuracy distributions were generated.

In this paper, we applied geostatistical indicator simulation to assessment of spatial uncertainty distributions in remote sensing data classification. Unlike the previous research [4], the sparse ground samples were integrated with exhaustive class-wise probability by applying sequential indicator simulation with local means. Supervised land-cover classification using multi-sensor remote sensing data was carried out to illustrate the application of the method.

2. Methodology

Unlike kriging, stochastic simulation allows generation of maps that reproduce the spatial pattern of the data without smoothing effect [2]. It can generate multiple realizations, each representing alternative representations of the unknown truth. In remote sensing data classification, the simulated alternative land-cover class representations can be used for assessing spatial accuracy or uncertainty. This approach can not only generate spatial accuracy distributions but also connect the classification result to the reference data.

For categorical attributes like land-cover classes, geostatistical modeling can be done using an indicator algorithm. The indicator approach provides a non-parametric distribution estimated directly at two possible outcomes i.e. 0 and 1 [3]. Sparse ground sample data, which are precise measurement of the attribute of interest, are coded into a set of binary hard indicator probabilities. The soft probabilities as a by-product of the classification procedure provide prior probabilities of occurrence for land-cover classes. After the prior indicator coding, the next step is to generate posterior probabilities by combining the hard and soft prior probabilities. For updating, simple indicator kriging with local means was applied. This algorithm estimates the unknown residual from the residual data using simple indicator kriging and the resulting estimate is added to the prior local mean (in our case, the soft prior probabilities derived from the classification). Once the local uncertainty model has been set up, sequential indicator simulation proceeds. From several simulation images, the final probability for each class that can be regarded as the spatial distribution probability of accuracy is computed as the normalized class-frequency (Fig. 1). For the details of sequential indicator simulation with local means, interested readers should refer to [2].

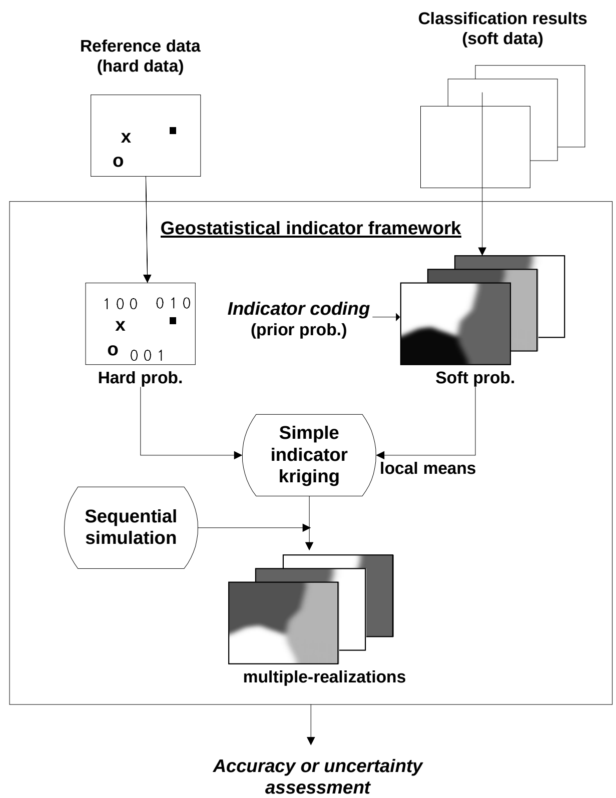

Fig. 1. Schematic flow of sequential indicator simulation with local means applied in this paper.

3. Experiment

The multi-sensor remote sensing data set provided by the IEEE GRSS Data Fusion Committee [5] is used in this experiment. The data set includes 6 optical ATM images and 9 SAR images of the P, L and C bands with full polarizations, a total of 15 images. The land-cover classes considered here are five agricultural classes (i.e. sugar beets, stubble, bare soil, potatoes and carrots). Originally, the data set includes the training and reference data set prepared for the comparison of several algorithms. Since the training and reference data sets are spatially clustered, spatial accuracy distributions may be overestimated. Thus, in this experiment, we newly generated the training and reference data sets which occupied the 2.5% proportions in the study area.

For modeling the spatial distributions of classification accuracy, two preprocessing steps were employed. The first step is to generate soft probabilities for 5 land-cover classes. Probabilistic neural network was applied for the classification. Secondly, the soft probabilities were calibrated against the hard indicator data by applying multinomial logistic regression. This was done to approximate equality of the stationary means of hard and soft indicator data and thus validity of unbiasedness condition [1]. After getting calibrated soft probabilities, sequential indicator simulation with local means was repeated 50 times with different random paths. For each pixel in a random path of pixels except reference hard pixels in a study area, the algorithm first computes a conditional probability in a given land-cover class, given the soft and hard data for the ten closest pixels for which hard data in the form of a reference land-cover class. A stochastic land-cover for the pixel is then simulated by choosing a land-cover class randomly with the probability proportional to the conditional probability derived from simple indicator kriging with local means.

Two such realizations (out of 50 generated) of alternative class labels are shown in Fig. 2. The simulated land-cover maps are locally different and some variations located on boundary pixels exist. This means that the high spatial uncertainties of accuracy distributions exist.

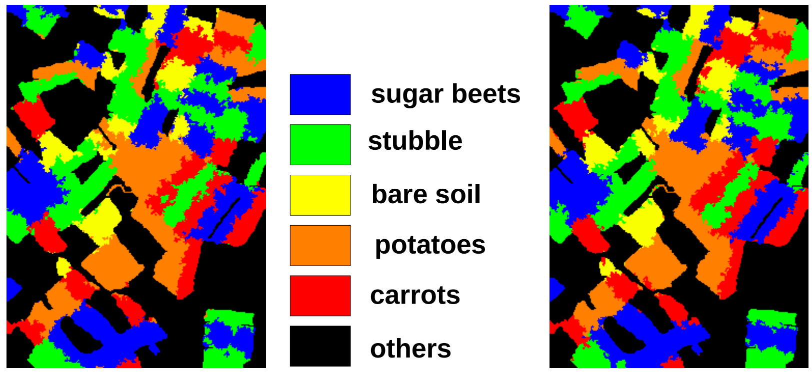

Fig. 2. Two realizations of land-cover class generated by .sequential indicator simulation with local means.

The spatial accuracy probability for each class to prevail at any pixels is obtained from the 50 realizations of class labels at each pixel. From the 50 realizations, for example, if the number of the simulation results belonging to sugar beets, stubble, bare soil, potatoes and carrots is 20, 15, 0, 10 and 5, the spatial accuracy probability for each class is 0.4, 0.3, 0.0, 0.2 and 0.1, respectively.

Above probability for each class is mapped over the study area in Fig. 3. The uncertainties (i.e. low accuracy values) are dominant on boundary pixels, due to mixed pixel information along the boundaries between neighboring fields. The results can also be regarded as land-cover maps updated by combination of ground data and classification results obtained by remote sensing data.

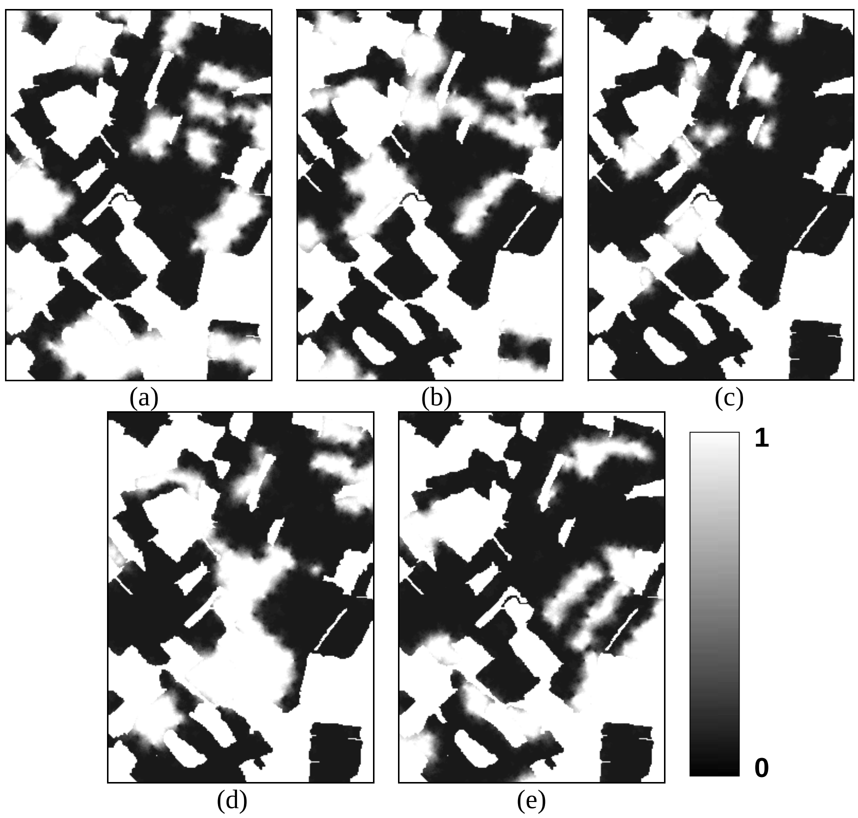

Fig. 3. Spatial accuracy probabilities for each land-cover class generated from 50 simulated land-cover maps. (a) sugar beets, (b) stubble, (c) bare soil, (d) potatoes, (e) carrots.

4. Conclusions

To assess the spatial accuracy or uncertainty distributions in remote sensing data classifications, geostatistical indicator simulation (i.e. sequential indicator simulation with local means) was applied. Multiple realizations can be used to estimate those spatial distributions. The results can provide useful information for evaluating the classification result or further ground survey. As well, it is expected that multiple realizations generated from simulation can also be used as inputs to any modeling process in which the land-cover map is used as an input parameter and thus the resulting distributions can be used for investigating the effects or uncertainties of the input parameter.

Acknowledgement

This authors thank the IEEE GRSSS Data Fusion Committee for providing the data set (grss-dfc-0006) used in this paper.

References

[1] 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.
[2] Goovaerts, P., 1997. Geostatistics for natural resources evaluation, Oxford University Press.
[3] Journel, A., 1983. Non-parametric estimation of spatial distributions, *Mathematical Geology*, 15(3): 4445-468.
[4] 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): 311-330.
[5] **URL:** IEEE GRSSS Data Fusion Committee. Available at: http://www.dfc-grsss.org.