A Feature-Space Indicator Kriging Approach for Remote Sensing Image Classification

Jie-Lun Chiang, Jun-Jih Liou, Chiang Wei, and Ke-Sheng Cheng

***Abstract*—An indicator kriging (IK) approach for remote sensing image classification is proposed. By introducing indicator variables for categorical data, the work of image classification is transformed into estimation of class-dependent probabilities in feature space using ordinary kriging. Individual pixels are then assigned to the class with maximum class probability. The approach is distribution free and yields perfect classification accuracies for training data provided that collocated data in feature space do not exist. Technical considerations regarding implementation of IK such as indicator semivariogram modeling and handling of collocated data in feature space are also described. The IK, Gaussian-based maximum likelihood, nearest neighbor, and support vector machine (SVM) classifiers were applied to study areas within the Shimen reservoir watershed (case A: FORMOSAT-2) and Taipei city (case B: SPOT 4). The results show that the overall accuracies of the proposed IK classifier and SVM can achieve higher than 97% for training data and 81% for testing data. (The overall accuracies of IK are a little higher than those of SVM.) IK and SVM are found to be superior to the other two classifiers in terms of overall accuracies for both training and testing data. The proposed IK classifier has the following advantages: 1) It can deal with anisotropic problem in feature space; 2) it is a nonparametric method and needs not to know the type of probability distribution; and 3) it yields 100% classification accuracy for the training data provided that collocated data in feature space do not exist.**

***Index Terms*—Image classification, image recognition, statistics.**

I. Introduction

LAND use/land cover classification is frequently applied using remote sensing images. Many classification methods exist in the literature, e.g., maximum likelihood (ML) classification, Bayes classification, and nearest neighbor classification, and new methods using artificial neural network (ANN), support vector machine (SVM) [1]–[3], geostatistics, and other algorithms are also evolving in recent years. Many of these classification methods root in specific statistical criteria since both the target objects (land surface objects) and the remote sensing data exhibit certain random characteristics in space and in time.

Therefore, terms like statistical pattern recognition or statistical discriminant analysis are also used to describe such work. Supervised classification requires the user's involvement in choosing training data. Implementing statistical criteria in classification, particularly for supervised classification, may require the data to be fitted to certain distribution types, and distribution parameters are estimated using training data. Such methods are termed the parametric (distribution specified) approach. For example, ML classification in many commercialized remote sensing image processing software packages assumes Gaussian distributions for classification features, although the concept of ML method only requires modeling each class with a probability density function, but the density function used may be of any form.

A drawback of the parametric (distribution specified) approach of remote sensing image classification is that probability distributions of classification features are often complicated and assumptions on these probability distributions are generally invalid. For example, land cover classification involves several land cover classes, and each class is characterized by several spectral or textural features. It is unlikely that all features will have the same type of probability distribution. In particular, the Gaussian-based ML classifier assumes that all classification features form a joint Gaussian distribution. Unfortunately, we often find classification features being non-Gaussian and having multimodal probability density functions. In such cases, splitting multimodal classes into separate unimodal subclasses for classification is a common practice [4], [5]. However, forcing a unimodal Gaussian assumption is not always justified. There are many reasons for features to be non-Gaussian, for example, areas of the same vegetation class in an image may have different slopes, aspects, and densities. Even though many applications have been conducted using Gaussian-based ML and Bayes classifications without a careful check of normality, quest for releasing the constraint on distribution type, i.e., the nonparametric (distribution free) classification, remains.

Geostatistics, which has been successfully applied to many spatial processes, also finds its way into remote sensing image analysis. Geostatistics is a set of techniques, often referred to as kriging methods, which utilize the spatial covariance function or the semivariogram for spatial data analysis. Since remote sensing images are spatial data, it is therefore attractive for

Manuscript received November 5, 2012; revised April 20, 2013; accepted July 11, 2013. Date of publication September 9, 2013; date of current version February 27, 2014. This work was supported in part by the National Science Council of Taiwan under Grants NSC101-2625-M-020-001 and NSC102-2625-M020-001.

J.-L. Chiang is with the Department of Soil and Water Conservation, National Pingtung University of Science and Technology, Neipu 912, Taiwan (e-mail: jlchiang@mail.npust.edu.tw).

J.-J. Liou is with the Information Division, National Science and Technology Center for Disaster Reduction, Taipei 23143, Taiwan (e-mail: leojeanz@gmail.com).

C. Wei is with the Experimental Forest, National Taiwan University, Nantou 55750, Taiwan (e-mail: d87622005@ntu.edu.tw).

K.-S. Cheng is with the Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: rslab@ntu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2279118

people who are familiar with geostatistics to apply such methods for remote sensing image processing. Atkinson and Curran studied the effect of regularization due to change of support (or pixel resolution) on local estimation and estimation of regional mean using remote sensing images [6]. They demonstrated that the semivariogram and kriging variance provide a powerful means of assessing the support in potential remote sensing investigations. Inggs and Lord applied ordinary kriging in interpolation of satellite-derived wind field [7]. Cheng *et al.* applied an anisotropic kriging approach for remote sensing image rectification [8]. Carr and Miranda compared the variogram measure computed within a local window with traditional co-occurrence-matrix-based textural features for image classification [9]. They found that the semivariogram yields better accuracy than co-occurrence-matrix-based features for classifying microwave imagery. Other studies also used the semivariogram as a textural feature for remote sensing image classifications [10], [11]. Atkinson and Lewis point out that the main problem with using semivariogram as a measure of texture is that the homogeneous regions of different texture within the image must be sufficiently large to allow computation of the variogram up to a reasonable number of lags [12].

Unlike the aforementioned methods which use the semivariogram as a measure of texture for classification, Van der Meer [5] applied an indicator kriging (IK) approach to classification of remotely sensed images for calcite–dolomite mineral mapping. The approach involved defining indicator variables for each feature (spectral band) of specified classes, based on predetermined lower and upper threshold values of individual features in spatial space (not in feature space). The spectral ranges used for class assignment and subsequent calculation of indicator variables are based on laboratory studies of spectral characteristics of certain minerals rather than on ground truth data [13]. Then, class-specific average values of individual indicator variables within block areas were estimated by IK. For each class, an average value over all indicator variables (or spectral bands) was obtained to represent the probability that the central pixel of the block belongs to that class. Finally, IK-classified images were derived by thresholding at 0.5 probability for both of the two classes used in their study. This study marked an important and the first application of geostatistics to probability-based image classification. However, the block average of indicator variables is specific to both spectral bands and land cover classes; therefore, when different spectral bands are considered, the approach yields different probabilities that the central pixel belongs to specific classes. For a certain class, the average value of indicator variables over all spectral bands represents the probability of the central pixel belonging to that class. Ambiguous class assignment may result since it is possible to have two or more classes having the average indicator variables higher than 0.5.

Since probability density and correlation structure between classification features are insightful, probability-based classification methods are appealing to many researchers and practitioners. The work of probability-based classification can be conceived as a spatial estimation problem for which the estimates are probabilities that a pixel with certain feature vector belongs to different classes. Van der Meer's [13] study performed the probability estimation in 2-D geographical space for individual spectral features. To circumvent the previously mentioned ambiguity in Van der Meer's study, this study proposes a feature-space-based IK approach of image classification which is nonparametric (distribution free) and integrates all classification features in probability estimation.

II. IK in Feature Space

Different from Van der Meer's study, we estimated class probabilities in feature space by using IK instead of probability estimation in geographic space because the spatial structure will be more obviously in feature space than in geographical space. For example, the pixels with similar features (spectrums) will be more likely the same class than those pixels with close location in geographic space. In this section, we describe the detailed procedures of IK in feature space for image classification. We begin with the definition of indicator variable for continuous random variables and the method of IK, followed by the indicator variable for categorical data. An algorithm for estimating class probabilities (the probabilities that a pixel belongs to certain classes) in feature space is finally described.

A. Indicator Variable and IK

For a continuous random variable Z with sample space Ω and cumulative distribution function $F_Z(z)$, an indicator variable (or indicator function) $I_D(z)$ can be defined as

$$I_D(z) = \begin{cases} 1, & \text{if } z \in D \\ 0, & \text{if } z \notin D \end{cases} \tag{1}$$

where D represents a subset of the sample space Ω. Given a random sample $\{z_1, z_2, \ldots, z_n\}$ of Z, the empirical cumulative distribution function can be defined as

$$\hat{F}_Z(z) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,z]}(z_i) = \sum_{i=1}^{n} w_i I_{(-\infty,z]}(z_i) \tag{2}$$

with $w_i = 1/n$. Equation (2) indicates that the average of the indicator variables corresponding to a random sample is an estimate of the cumulative distribution function. All weights in (2) are equal since values of a random sample are observed from a set of independent variables with a common probability density function.

To extend usage of indicator variable to spatially distributed data, let us consider a continuous random field $Z(x)$ with n measurements $\{z(x_1), z(x_2), \ldots, z(x_n)\}$, where x_i, $i = 1, 2, \ldots, n$, represents the spatial location and z represents the corresponding measurements. We want to estimate the cumulative distribution function of $Z(x_0)$ using the nearby measurements $\{z(x_1), z(x_2), \ldots, z(x_n)\}$. Unlike random samples of a random variable, measurements of a random field are generally correlated, and therefore, unequal weights must be used. Weighted linear combination of the nearby sample indicators is used to estimate the local cumulative distribution

function [14]. Given a threshold value v_c for $Z(x_0)$, the nearby sample indicator variables are determined by

$$I_{v_c}(x_i) = \begin{cases} 1, & \text{if } z(x_i) \le v_c \\ 0, & \text{if } z(x_i) > v_c \end{cases} \tag{3}$$

and the cumulative probability of $Z(x_0)$ corresponding to v_c is estimated as

$$\hat{F}_{Z_0}(v_c) = \sum_{i=1}^{n} \lambda_{i0} I_{v_c}(x_i). \tag{4}$$

Weights λ_{i0}, $i = 1, 2, \ldots, n$, are determined by solving the following ordinary kriging system equation of indicator variables or the IK system:

$$\begin{cases} \sum\limits_{i=1}^{n} \lambda_{i0}\gamma_{ij} + \ell = \gamma_{j0}, \quad j = 1, 2, \ldots, n \\ \sum\limits_{i=1}^{n} \lambda_{i0} = 1 \end{cases} \tag{5}$$

where ℓ is a Lagrange multiplier and $\gamma_{ij} = \gamma(|x_i - x_j|)$ or $\gamma_{j0} = \gamma(|x_j - x_0|)$ is the semivariogram, or simply the variogram, of the indicator random field $I_{v_c}(x)$ and is defined as

$$\gamma(|x_i - x_j|) = \frac{1}{2} E\left\{ [I_{v_c}(x_i) - I_{v_c}(x_j)]^2 \right\}. \tag{6}$$

Semivariograms calculated from observed data, also termed as the experimental semivariograms, must be fitted to theoretical models in order to be feasible [20]. The $(\gamma(h) = \omega[1 - \exp(-(h/a))])\gamma(h)$ semivariogram will reach its maximum value ω as the distance h approaches infinity; however, at distance $h_c = 3a$, the semivariogram will be approximately 0.95ω, and $3a$ is termed the practical influence range of the exponential semivariogram. Random variables corresponding to two spatial points that are at least $3a$ distance apart are considered to be mutually independent. Readers are reminded that, for IK, the semivariogram of indicator variables (indicator semivariogram) must be utilized. $\hat{F}_{Z_0}(v_c)$ is obtained by solving the IK system, i.e., (5), for kriging weights λ_{i0} and then substituting them into (4). Estimate of the local cumulative distribution $F_{Z_0}(z)$ can be obtained by using different values for the threshold v_c. For detailed description of ordinary kriging and IK, readers are referred to [14] and [15].

Since land cover types are categorical data, the definition of indicator variable must be modified to accommodate the image classification problem. Similar to the case of a continuous random field, the indicator variable can be used to estimate the probability that a pixel belongs to a certain class for a categorical random field. Let the indicator variable be defined as

$$I_j(x) = \begin{cases} 1, & \text{if } s(x) \in C_j \\ 0, & \text{if } s(x) \notin C_j \end{cases} \qquad j = 1, 2, \ldots, k \tag{7}$$

where C_j represents the jth class and $s(x)$ represents the pixel at location x. $I_j(x)$ is the value of the indicator variable related to the jth class. The weighted average of values of n neighboring indicator variables is an estimate of the probability that a pixel at location x_0 belongs to the jth class, i.e.,

$$P\left(s(x_0) \in C_j\right) = P_j(x_0) = \sum_{i=1}^{n} \lambda_{i0} I_j(x_i) \tag{8}$$

where $P_j(x_0)$ is the probability that $s(x_0)$ belongs to class C_j.

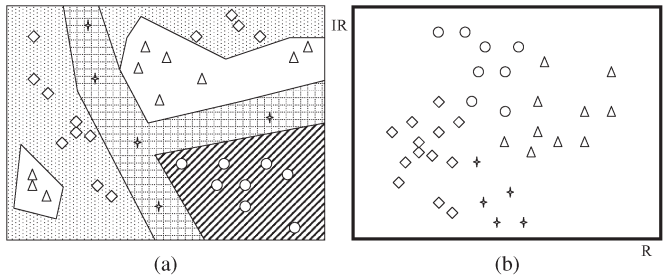


Fig. 1. Illustrative example showing (a) discontinuity in geographical space and (b) stronger continuity in feature space. (a) Image in geographical space with four types of land cover (shown in different patterns). Marked symbols represent training data points. (b) Distribution of training data in a 2-D feature space (e.g., IR and R bands).

B. Transforming Classification Into Estimation

The work of remote sensing image classification assigns one class identity to each pixel in the image. Class identities are categorical data and are noncontinuous in geographical and feature spaces. Thus, spatial estimation of a class identity in space is logically incorrect. In order to transform the work of image classification into an estimation problem, two factors need to be considered. First, we need to have a noncategorical measure that associates with class identities. Second, we need to find a space in which the chosen measure is continuous and spatial estimation of the chosen measure will be made. In this study, we adopt the class probabilities, i.e., the probability that a pixel belongs to certain classes, as our spatial estimation parameters. Since spatial discontinuity of land cover often occurs in geographical space due to human activities, our estimation of class probabilities takes place in a feature space rather than a geographical space. Fig. 1 illustrates an example of spatial discontinuity in geographical space, whereas it exhibits stronger spatial continuity in feature space.

Consider an image classification problem of k classes using m classification features. A random field $\{Z(x) \in \{1, 2, \ldots, k\}; x \in \Omega\}$ represents the distribution of class identity in an m-dimensional spatial domain Ω. For convenience of illustration, let us assume that $k = 3$ and $m = 2$. From a set of training pixels, we first establish the three-class scatter plot in feature space [Fig. 2(a)]. In order to estimate the class probabilities in feature space, we then transform the three-class scatter plot in 2-D feature space into three two-class scatter plots corresponding to each of the three individual classes [see Fig. 2(b)–(d)]. Hereafter, we shall refer to these two-class scatter plots as binary scatter plots of individual classes. For each binary scatter plot, we consider the spatial distribution of indicator variables as a random field associated with that particular class. By conducting ordinary kriging of class-specific

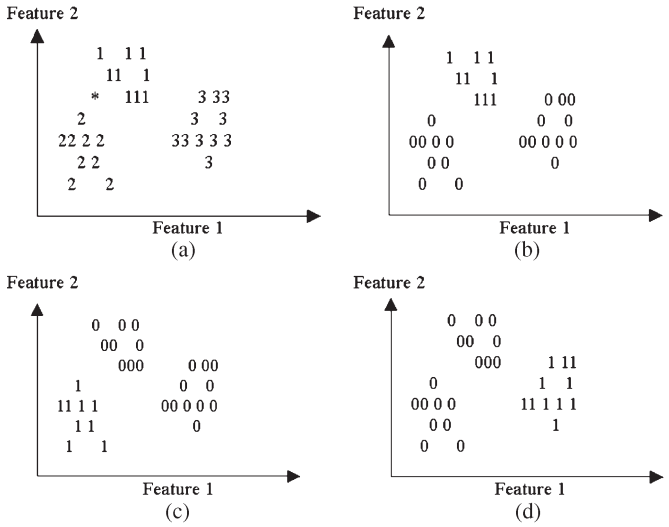


Fig. 2. Binary scatter plots in 2-D feature space. (a) Three-class scatter plot in 2-D feature space ($*$ denotes collocated training data). (b) Class-1/non-class-1 scatter plot. (c) Class-2/non-class-2 scatter plot. (d) Class-3/non-class-3 scatter plot.

indicator variables, as defined by (7), in feature space, we obtain the class probability for each individual class. Finally, the pixel of interest is assigned to the class with maximum class probability.

C. Semivariogram Modeling and SVM Classification

Experimental indicator semivariograms of different land cover classes are fitted to exponential (9)/Gaussian (10) models. Since indicator semivariograms $\gamma(h)$ are calculated in a feature space formed by four spectral bands, the distance h is in unit of digital number of the satellite images. It is also important to compare and consider the scale of classification features in calculating the separation distance h. In our study, digital numbers of the three bands are of the same scale, and scale adjustment is not necessary. However, if image classification involves texture features which are not all of the same order of scale, it may be necessary to adjust the scale of classification features in the calculation of the separation distance.

The experimental indicator semivariograms of land cover classes 1–5 in case A are fitted to an exponential model, and the others (land cover classes 6–8 and the classes in case B) are fitted to a Gaussian model

$$\gamma(h) = \omega\left[1 - \exp\left(-\left(\frac{h}{a}\right)\right)\right] \tag{9}$$

$$\gamma(h) = \omega\left[1 - \exp\left(-\left(\frac{h}{a}\right)^2\right)\right]. \tag{10}$$

Another important consideration in semivariogram modeling is the anisotropy. Digital numbers of different spectral bands are often correlated, and class-specific training data may scatter along certain directions in a spectral feature space. Under such situation, an anisotropic semivariogram which considers directional distribution of training data can be adopted. Anisotropic semivariogram modeling in a 2-D feature space involves identification of the principal (maximum and minimum) variation

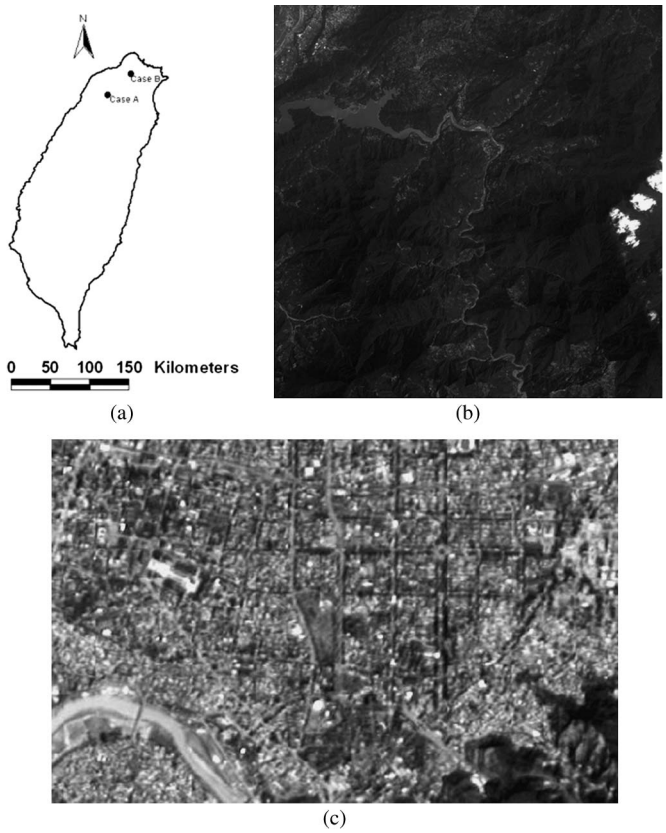


Fig. 3. Locations and images of the study area. (a) Locations of cases A and B. (b) Case A image (FORMOSAT-2). (c) Case B image (SPOT 4).

directions and an anisotropic ratio. Semivariogram modeling is needed in one of the principal directions only, and a semivariogram in any other direction can be derived using the principal-direction semivariogram and an equivalent distance calculated using the anisotropic ratio. For feature space with number of dimensions higher than or equal to three, anisotropic semivariogram modeling becomes more complex and difficult to apply. Since remote sensing image classification often involves more than three classification features, anisotropic semivariogram modeling is practically difficult and therefore is not discussed in this paper.

Recently, in addition to ANN, SVM is a commonly used classification technique as well and had been widely applied in remote sensing image classification [1]–[3], [16]–[18].

In recent years, SVM has been widely used in various fields, particularly in image classification. SVM is a powerful nonlinear pattern recognition technique. A user-friendly SVM tool, namely, Library for Support Vector Machine (LIBSVM) [19], was used to implement the multiclassification for comparative studies. At the data preprocessing stage, raw digital numbers of different spectral bands were linearly rescaled into $[-1, 1]$ using the ranges of their minimums and maximums. In LIBSVM, two unknown parameters (cost and gamma) were determined by trial and error, while a kernel function was given as a radial basis function. At the model establishing stage, a K-fold cross-validation strategy (five disjoint subsets of training data) was adopted to avoid the overfitting problem. The optimal parametric pair (cost and gamma) was decided

TABLE I
EVALUATION OF CLASSIFICATION ACCURACIES—GAUSSIAN-BASED ML METHOD. (a) TRAINING DATA (CASE A). (b) TESTING DATA (CASE A). (c) TRAINING DATA (CASE B). (d) TESTING DATA (CASE B)

(a)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	148	38	80	0	0	0	0	0	266	55.64
	Bamboo	22	158	58	0	12	0	0	0	250	63.20
	Hardwood	32	11	202	0	0	0	0	0	245	82.45
	Paddy field	0	1	0	60	23	3	6	0	93	64.52
	Dry farming field	0	5	8	13	164	1	17	0	208	78.85
	Build-up area	0	0	0	0	0	211	1	4	216	97.69
	Bared soil	0	1	0	15	12	5	65	0	98	66.33
	Water	0	0	0	0	0	2	0	877	879	99.77
	SUM	202	214	348	88	211	222	89	881	2255	Overall Accuracy (%)
	Producer's Accuracy (%)	73.27	73.83	58.05	68.18	77.73	95.05	73.03	99.55		83.59

(b)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	64	0	22	0	0	0	0	0	86	74.42
	Bamboo	10	54	63	0	0	0	0	0	127	42.52
	Hardwood	2	19	68	0	0	0	0	0	89	76.40
	Paddy field	0	5	0	17	28	4	9	0	63	26.98
	Dry farming field	2	15	4	15	55	1	2	0	94	58.51
	Build-up area	0	0	0	0	0	106	0	10	116	91.38
	Bared soil	0	2	2	6	5	1	39	1	56	69.64
	Water	0	0	0	0	0	0	0	436	436	100.00
	SUM	78	95	159	38	88	112	50	447	1067	Overall Accuracy(%)
	Producer's Accuracy(%)	82.05	56.84	42.77	44.74	62.50	94.64	78.00	97.54		78.63

(c)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	91	0	0	0	91	100
	Woods	0	101	14	0	115	87.83
	Grass Lands	0	29	81	5	115	70.43
	Built-up	3	0	0	178	181	98.34
	SUM	94	130	95	183	502	Overall Accuracy(%)
	Producer's Accuracy (%)	96.81	77.69	85.26	97.27		89.84

(d)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	60	0	0	0	60	100
	Woods	0	74	4	0	78	94.87
	Grass Lands	0	12	57	0	69	82.61
	Built-up	6	34	0	167	207	80.68
	SUM	66	120	61	167	414	Overall Accuracy (%)
	Producer's Accuracy (%)	90.91	61.67	93.44	100		86.47

by the best averaged multiclassified result among different parametric pairs. The parametric space of the former parameter (cost) was varied between e^{-5} and e^{15} with a step of e^2, while the latter (gamma) was between e^{-15} and e^5 with the same step as the former.

SVM is a famous and powerful nonlinear classifier, but IK is linear. Both SVM and IK divide samples belonging to different categories in the feature space.

Nonlinear classifiers should yield better results than linear ones in most cases. However, the more effort was done in parametric optimization, the better classified results might be yielded. This should also be true in both SVM and IK multiclassifications. The point of this study is not SVM but the first application of replacing space distance by spectral distance in the semivariogram model. Therefore, this study had made a lot of work in parametric calibration of a variogram model but few at parameters of SVM's kernel function. This should be mentioned before making a conclusion of classified results.

III. STUDY AREA AND IMAGE DATA

The following are two case study areas which contain a mountain area (case A) and an urban area (case B) and are the most commonly seen land covers in Taiwan (see Fig. 3).

Case A: A study area of approximately 415 km^2 mostly within the Shimen reservoir watershed [Fig. 3(b)] was selected for test of the proposed IK classification approach. Taoyuan county occupies the northwestern corner, and vegetation covers

TABLE II
EVALUATION OF CLASSIFICATION ACCURACIES—NEAREST NEIGHBOR METHOD. (a) TRAINING DATA (CASE A). (b) TESTING DATA (CASE A). (c) TRAINING DATA (CASE B). (d) TESTING DATA (CASE B)

(a)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	202	1	1	0	0	0	0	0	204	99.02
	Bamboo	0	213	0	0	0	0	0	0	213	100.00
	Hardwood	0	0	347	0	0	0	0	0	347	100.00
	Paddy field	0	0	0	88	0	0	0	0	88	100.00
	Dry farming field	0	0	0	0	211	0	0	0	211	100.00
	Build-up area	0	0	0	0	0	222	0	0	222	100.00
	Bared soil	0	0	0	0	0	0	89	0	89	100.00
	Water	0	0	0	0	0	0	0	881	881	100.00
	SUM	202	214	348	88	211	222	89	881	2255	Overall Accuracy (%)
Producer's Accuracy (%)		100.00	99.53	99.71	100.00	100.00	100.00	100.00	100.00		99.91

(b)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	40	3	10	0	0	0	0	0	53	75.47
	Bamboo	16	51	27	0	0	0	0	0	94	54.26
	Hardwood	22	19	121	0	1	0	0	0	163	74.23
	Paddy field	0	0	0	11	22	3	2	0	38	28.95
	Dry farming field	0	22	1	20	65	2	16	0	126	51.59
	Build-up area	0	0	0	3	0	104	1	0	108	96.30
	Bared soil	0	0	0	4	0	2	31	0	37	83.78
	Water	0	0	0	0	0	1	0	447	448	99.78
	SUM	78	95	159	38	88	112	50	447	1067	Overall Accuracy (%)
Producer's Accuracy (%)		51.28	53.68	76.10	28.95	73.86	92.86	62.00	100.00		81.54

(c)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	94	0	0	0	94	100
	Woods	0	123	3	0	126	97.62
	Grass Lands	0	7	92	0	99	92.93
	Built-up	0	0	0	183	183	100
	SUM	94	130	95	183	502	Overall Accuracy (%)
Producer's Accuracy (%)		100	94.62	96.84	100		98.01

(d)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	64	0	0	0	64	100
	Woods	0	77	2	0	79	97.47
	Grass Lands	0	18	59	0	77	76.62
	Built-up	2	25	0	167	194	86.08
	SUM	66	120	61	167	414	Overall Accuracy (%)
Producer's Accuracy (%)		96.97	64.16	96.72	100		88.65

of forests, farms, parks, campuses, river banks, etc., widely disperse over the area. Eight major land cover types—softwood, bamboo, hardwood, paddy field, dry farming field, buildup area, bared soil, and water—were identified using 1 : 5000-scale orthophotographs.

Multispectral FORMOSAT-2 satellite images (8-m resolution) acquired on March 5, 2008, were used for land cover classification. We use the blue (0.45–0.52 μm), green (0.50–0.60 μm), red (0.63–0.69 mm), and near-infrared (0.76–0.92 μm) spectral bands of the FORMOSAT-2 images as classification features.

Case B: A study area of approximately 70 km^2 mostly within the Taipei metropolitan [Fig. 3(c)] was selected as case B. Four major land cover types—water, woods, grasslands, and built-up areas—were identified due to the relatively simple land use condition within the study area.

Multispectral SPOT 4 satellite images (20-m resolution) acquired on December 24, 2001, were used for land cover classification. In case B, we use the blue (0.45–0.52 μm), green (0.53–0.59 μm), red (0.62–0.69 mm), and infrared (0.76–0.89 μm) spectral bands of the SPOT 4 images as classification features.

IV. RESULTS AND DISCUSSIONS

A. Classification Results

The Gaussian-based ML, nearest neighbor, SVM, and IK classification methods were applied to the study area. The classification accuracies of all the four methods using training and testing data are listed in Tables I–IV. With complex land use condition in the study area, the ML classifier achieves

TABLE III
EVALUATION OF CLASSIFICATION ACCURACIES—SVM METHOD. (a) TRAINING DATA (CASE A). (b) TESTING DATA (CASE A). (c) TRAINING DATA (CASE B). (d) TESTING DATA (CASE B)

(a)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	180	6	11	0	0	0	0	0	197	91.37
	Bamboo	10	203	7	0	1	0	0	0	221	91.86
	Hardwood	12	5	330	0	0	0	0	0	347	95.10
	Paddy field	0	0	0	87	0	0	0	0	87	100.00
	Dry farming field	0	0	0	1	210	0	0	0	211	99.53
	Build-up area	0	0	0	0	0	222	0	0	222	100.00
	Bared soil	0	0	0	0	0	0	89	0	89	100.00
	Water	0	0	0	0	0	0	0	881	881	100.00
	SUM	202	214	348	88	211	222	89	881	2255	Overall Accuracy (%)
	Producer's Accuracy (%)	89.11	94.86	94.83	98.86	99.53	100.00	100.00	100.00		97.65

(b)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	39	3	7	0	0	0	0	0	49	79.59
	Bamboo	15	54	23	0	0	0	0	0	92	58.70
	Hardwood	24	17	118	0	0	0	0	0	159	74.21
	Paddy field	0	0	0	13	23	2	4	0	42	30.95
	Dry farming field	0	18	1	16	59	1	9	0	104	56.73
	Build-up area	0	3	10	3	5	109	2	4	136	80.15
	Bared soil	0	0	0	6	1	0	35	0	42	83.33
	Water	0	0	0	0	0	0	0	443	443	100.00
	SUM	78	95	159	38	88	112	50	447	1067	Overall Accuracy (%)
	Producer's Accuracy (%)	50.00	56.84	74.21	34.21	67.05	97.32	70.00	99.11		81.54

(c)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	94	0	0	0	94	100
	Woods	0	130	0	0	130	100
	Grass Lands	0	0	95	0	95	100
	Built-up	0	0	0	183	183	100
	SUM	94	130	95	183	502	Overall Accuracy (%)
	Producer's Accuracy (%)	100	100	100	100		100

(d)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	66	0	0	0	66	100
	Woods	0	118	3	0	121	97.52
	Grass Lands	0	2	58	0	60	96.67
	Built-up	0	0	0	167	167	100
	SUM	66	120	61	167	414	Overall Accuracy (%)
	Producer's Accuracy (%)	100	98.33	95.08	100		98.79

overall accuracies of 83.59% for training data and 78.63% for testing data. The results of the nearest neighbor and SVM classifiers are 99.91% (training) and 81.54% (testing), and 97.65% (training) and 81.54% (testing), respectively. The IK classifier yields overall accuracies of 99.91% and 82.01% for training and testing data, respectively. Figs. 4–6 show the classification results of the IK method.

In case B, although all classification methods result in high overall accuracies, it is noteworthy that the IK classifier has superiority over other classifiers with regard to the producer's and user's accuracies using testing data. For example, the producer's accuracy of woods is 61.67% for the Gaussian-based ML method, as compared to 100% for the IK method. Detailed inspection using high-resolution aerial photographs and site visit also confirm higher accuracies achieved by the IK classifier. The true color details of three exemplar sites (the circled sites in Fig. 5) shown in Fig. 6 demonstrate that the land cover map derived from IK can show the details (e.g., in sites A, B, and C).

B. Exact Estimation of IK

As shown in Table IV(a), the IK classification achieves 99.91% of overall accuracy, and its error matrix of classification

TABLE IV
EVALUATION OF CLASSIFICATION ACCURACIES—IK METHOD. (a) TRAINING DATA (CASE A). (b) TESTING DATA (CASE A). (c) TRAINING DATA (CASE B). (d) TESTING DATA (CASE B)

(a)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	202	1	1	0	0	0	0	0	204	99.02
	Bamboo	0	213	0	0	0	0	0	0	213	100.00
	Hardwood	0	0	347	0	0	0	0	0	347	100.00
	Paddy field	0	0	0	88	0	0	0	0	88	100.00
	Dry farming field	0	0	0	0	211	0	0	0	211	100.00
	Build-up area	0	0	0	0	0	222	0	0	222	100.00
	Bared soil	0	0	0	0	0	0	89	0	89	100.00
	Water	0	0	0	0	0	0	0	881	881	100.00
	SUM	202	214	348	88	211	222	89	881	2255	Overall Accuracy (%)
Producer's Accuracy (%)		100.00	99.53	99.71	100.00	100.00	100.00	100.00	100.00		99.91

(b)

		Referenced Classes									
		Softwood	Bamboo	Hardwood	Paddy field	Dry farming field	Build-up area	Bared soil	Water	SUM	User's Accuracy (%)
Assigned Classes	Softwood	38	1	4	0	0	1	0	0	44	86.36
	Bamboo	18	53	22	0	0	0	0	0	93	56.99
	Hardwood	22	18	132	0	1	0	0	0	173	76.30
	Paddy field	0	0	0	10	20	4	2	0	36	27.78
	Dry farming field	0	23	1	18	66	1	15	0	124	53.23
	Build-up area	0	0	0	4	0	100	2	0	106	94.34
	Bared soil	0	0	0	6	1	5	29	0	41	70.73
	Water	0	0	0	0	0	1	2	447	450	99.33
	SUM	78	95	159	38	88	112	50	447	1067	Overall Accuracy (%)
Producer's Accuracy (%)		48.72	55.79	83.02	26.32	75.00	89.29	58.00	100.00		82.01

(c)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	94	0	0	0	94	100
	Woods	0	130	0	0	130	100
	Grass Lands	0	0	95	0	95	100
	Built-up	0	0	0	183	183	100
	SUM	94	130	95	183	502	Overall Accuracy (%)
Producer's Accuracy (%)		100	100	100	100		100

(d)

		Referenced Classes					
		Water	Woods	Grass Lands	Built-up	SUM	User's Accuracy (%)
Assigned Classes	Water	66	0	0	0	66	100
	Woods	0	120	2	0	122	98.36
	Grass Lands	0	0	59	0	59	100
	Built-up	0	0	0	167	167	100
	SUM	66	120	61	167	414	Overall Accuracy (%)
Producer's Accuracy (%)		100	100	96.72	100		99.52

is the same as the result of the nearest neighbor classifier using training data. Such results are not unexpected since IK is an exact estimator which yields exactly correct estimates at the measurement points. This is a significant contrast with a least squares fitting of a polynomial, which will never give the true value at the measurement points [20]. As an exact estimator, the IK can expect to yield very high classification accuracies if enough numbers of training data are available.

C. Collocated Training Data in Feature Space

Theoretically, IK yields exact estimates, i.e., completely correct classification results; however, ambiguity in class assignment may occur when collocated but different-class training data in feature space exist. Assume that m pixels of training data have the same digital numbers of classification features and among which k_1 and k_2 pixels belong to classes 1 and 2, respectively (see $*$ in Fig. 2). Under such situation, all these collocated data are assigned to the class with more collocated data. As a result, it may also be possible for the IK classifier to yield less than 100% accuracies. The amount of different-class and collocated data generally accounts only a very small portion of the total number of pixels.

V. CONCLUSION

In this study, we have proposed a feature-space-based IK approach for remote sensing image classification. The approach is

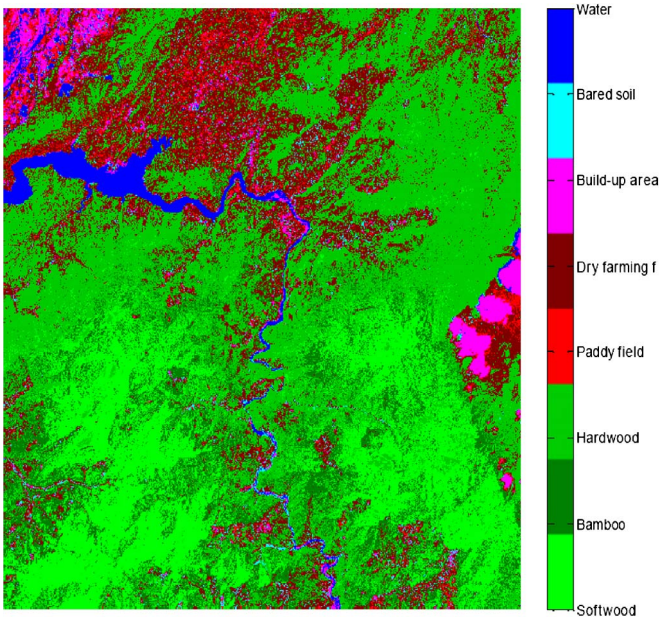


Fig. 4. Land cover classification results of the IK classifier (case A).

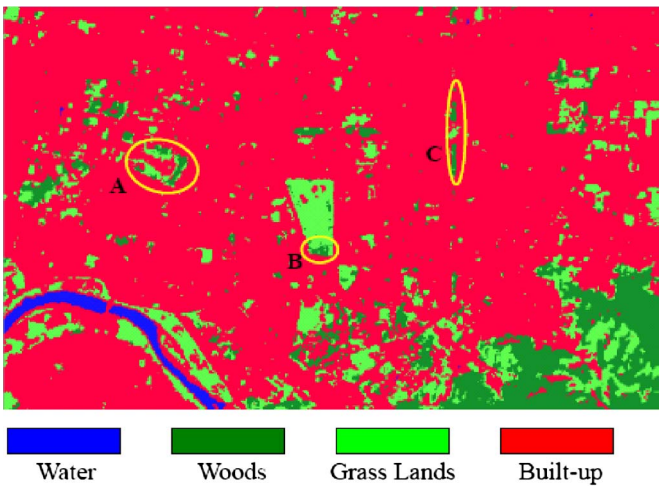


Fig. 5. Land cover classification results of the IK classifier (case B) (details of the circled sites are shown in Fig. 6).

nonparametric (distribution free) and does not require a specific type of distribution for spectral features and yields *a posteriori* probabilities for land cover class assignment. The IK classifier is capable of achieving perfect classification accuracies for training data if no different-class collocated training data occur in feature space owing to its exact estimation property. We have demonstrated that the IK classifier is superior to the Gaussian-based ML, nearest neighbor, and SVM classifiers in terms of overall accuracies using training data and testing data as well. Although no rigorous investigation about the number of points to be considered in kriging estimation was conducted, our experiences from this study indicate that inclusion of more points may have adverse effect on classification accuracy, possibly due to reduction of the kriging weights assigned to points within close vicinity of the point of interest. The results show that the proposed IK classifier provides an alternative to recently commonly used SVM.

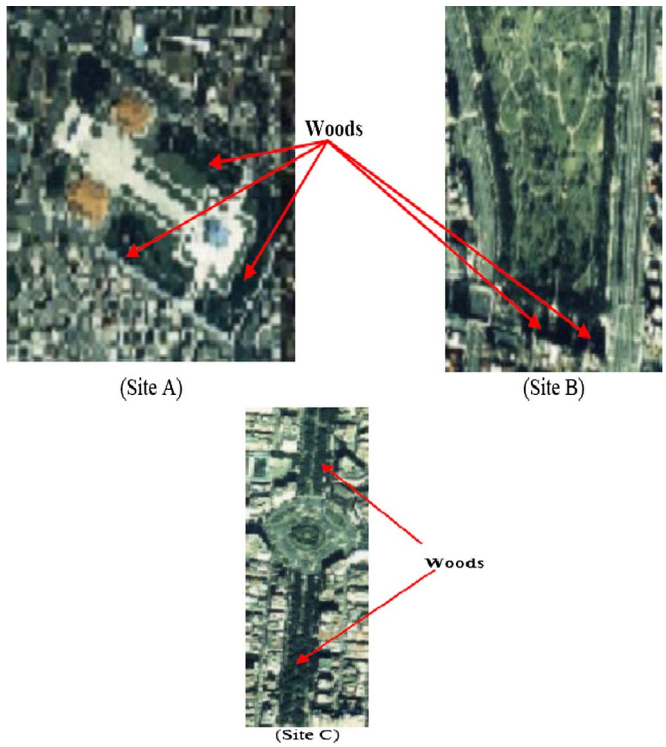
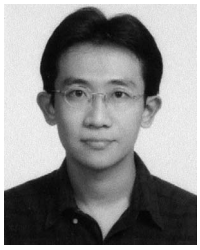


Fig. 6. True color details of three exemplar sites (site locations are shown in Fig. 5).

REFERENCES

[1] J. Ren, "ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging," *Knowledge-Based Syst.*, vol. 26, pp. 144–153, Feb. 2012.
[2] G. M. Foody, "RVM-based multi-class classification of remotely sensed data," *Int. J. Remote Sens.*, vol. 29, no. 6, pp. 1817–1823, 2008.
[3] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.
[4] J. R. Schott, *Remote Sensing: The Image Chain Approach*. New York, NY, USA: Oxford Univ. Press, 1997.
[5] F. Van der Meer, "Classification of remotely-sensed imagery using an indicator kriging approach—Application to the problem of calcite–dolomite mineral mapping," *Int. J. Remote Sens.*, vol. 17, no. 6, pp. 1233–1249, Apr. 1996.
[6] P. M. Atkinson and P. J. Curran, "Defining an optimal size of support for remote sensing investigations," *IEEE Trans. Geoscience and Remote Sensing*, vol. 33, no. 3, pp. 768–776, May 1995.
[7] M. R. Inggs and R. T. Lord, "Interpolating satellite derived wind field data using ordinary kriging with application to the nadir gap," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 250–256, Jan. 1996.
[8] K. S. Cheng, H. C. Yeh, and C. H. Tsai, "An anisotropic spatial modeling approach for remote sensing image rectification," *Remote Sens. Environ.*, vol. 73, no. 1, pp. 46–54, Jul. 2000.
[9] J. R. Carr and F. P. Miranda, "The semivariogram in comparison to the co-occurrence matrix for classification of image texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 6, pp. 1945–1952, Nov. 1998.
[10] R. M. Lark, "Geostatistical description of textures on an aerial photograph for discriminating classes of land-cover," *Int. J. Remote Sens.*, vol. 17, no. 11, pp. 2115–2133, Jul. 1996.
[11] M. Chica-Olmo and F. Arbarca-Hérnandez, "Computing geostatistical image texture for remotely sensed data classification," *Comput. Geosci.*, vol. 26, no. 4, pp. 373–383, May 2000.
[12] P. M. Atkinson and P. Lewis, "Geostatistical classification for remote sensing: An introduction," *Comput. Geosci.*, vol. 26, no. 4, pp. 361–373, May 2000.
[13] F. Van der Meer, "Extraction of mineral absorption features from high-spectral resolution data using non-parametric geostatistical techniques," *Int. J. Remote Sens.*, vol. 15, no. 11, pp. 2193–2214, 1994.
[14] E. H. Isaaks and R. M. Srivastava, *Applied Geostatistics*. New York, NY, USA: Oxford Univ. Press, 1989.

[15] J. P. Chilès and P. Delfiner, *Geostatistics—Modeling Spatial Uncertainty*. New York, NY, USA: Wiley, 1999.
[16] F. Samadzadegan, H. Hasani, and T. Schenk, "Simultaneous feature selection and SVM parameter determination in classification of hyperspectral imagery using ant colony optimization," *Can. J. Remote Sens.*, vol. 38, no. 2, pp. 139–156, Apr. 2012.
[17] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.
[18] H. W. Chen, N. B. Chang, R. F. Yu, and Y. W. Huang, "Urban land use and land cover classification using the neural-fuzzy inference approach with FORMOSAT-2 data," *J. Appl. Remote Sens.*, vol. 3, no. 1, pp. 033558-1–033558-18, Oct. 2009.
[19] C. C. Chang and C. J. Lin, LIBSVM: A Library for Support Vector Machines, 2001, [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[20] C. De Marsily, *Quantitative Hydrogeology*. Orlando, FL, USA: Academic, 1986.



Jie-Lun Chiang received the B.S. degree in hydraulic engineering from Feng Chia University, Taichung, Taiwan, in 1994 and the M.S. degree in agricultural engineering and the Ph.D. degree in bioenvironmental system engineering from National Taiwan University, Taipei, Taiwan, in 1996 and 2006, respectively.

He is currently an Associate Professor of geographic information system (GIS) and remote sensing with the Department of Soil and Water Conservation, National Pingtung University of Science and Technology, Neipu, Taiwan. His research interests include GIS and remote sensing application to watershed management, uncertainty of temporal–spatial data, and climate change impact on soil and water resources.

Jun-Jih Liou received the B.S. and M.S. degrees in agricultural engineering and the Ph.D. degree in bioenvironmental system engineering from National Taiwan University, Taipei, Taiwan, in 1998, 2000, and 2008, respectively.

Since 2010, he has been an Assistant Researcher with the Information Division, National Science and Technology Center for Disaster Reduction, Taipei. His research interests include random-field simulation, remote sensing, hydrology, and climate change.

Chiang Wei received the B.S. and M.S. degrees in agricultural engineering and the Ph.D. degree in bioenvironmental system engineering from National Taiwan University, Taipei, Taiwan, in 1993, 1995, and 2003, respectively.

Since 2003, he has been a Research Scientist with the Experimental Forest, National Taiwan University, Nantou, Taiwan, where he is also currently an Associate Research Fellow. He continued his research in the fields of remote sensing, hydrology, and forest environment monitoring and assessment.

Ke-Sheng Cheng received the B.Sc. and M.Sc. degrees from National Taiwan University, Taipei, Taiwan, in 1979 and 1983, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1989.

He is currently a Professor with the Department of Bioenvironmental Systems Engineering, National Taiwan University. His research interests focus on three interrelated fields: hydrologic modeling and forecasting, environmental remote sensing, and spatial and geostatistical modeling. He is particularly interested in applying stochastic modeling and simulation to environmental processes and climate change.