

## POST-CLASSIFICATION APPROACH BASED ON GEOSTATISTICS TO REMOTE SENSING IMAGES : SPECTRAL AND SPATIAL INFORMATION FUSION

N. Yao<sup>a,\*</sup>, J. X. Zhang<sup>a</sup>, Z. J. Lin<sup>b</sup>, C. F. Ren<sup>a</sup>

<sup>a</sup> Remote Sensing information Engineering School, Wuhan University, 129 Luoyu Road, Wuhan, China  
nayao@foxmail.com; jxzhang@whu.edu.cn; ren\_cf@163.com

<sup>b</sup> Chinese Academy of Surveying and Mapping, 16 Beitaping Road, Beijing, China - lincasm@casm.ac.cn

**KEY WORDS:** Classification, Information, Fusion, Spectral, Spatial, Geostatistics, Kriging

### ABSTRACT:

Classification of remote sensing imagery provides an inexpensive yet efficient approach to land cover mapping. In supervised image classification, training samples are collected through certain sampling schemes, which are used to derive classification rules, aiming for adequate accuracy for the applications at hand. However, in conventional classification methods, the potential of training samples in terms of locational information is not tapped further, confounding the classification accuracy to the limited separability inherent to the given input feature vector. This paper explores two methods pertaining to geostatistics, i.e., simple kriging with local mean and cokriging, to predict class occurrences based on training samples' indicator transforms (location and classes) and spectrally derived class probabilities, thus calibrating the a posteriori class probability vectors derived from initial spectral classification. The results showed that classification accuracy is significantly increased by these two methods for utilizing spatial information contained in training samples and initial spectral classification, compared with those obtainable with spectral classification. Moreover, the proposed methods constitute a valuable strategy for making fuller use of information residing in training data for improving spectrally derived classification, which is independent of the specific classifiers initially adopted for image classification.

### 1. INTRODUCTION

Land cover, as a spatial factor impacting and linking human life and natural environment, refers to the observed (bio)physical cover on the earth's surface (FAO, 2000). Remote sensing is an attractive data source for land cover mapping. Although remote sensing has been used successfully in mapping a range of land covers at a variety of spatial and temporal scales, the land cover maps derived are often judged to be of insufficient quality for operational applications (Foody, 2002). Therefore, how to improve the quality of land cover maps has been a hot issue all the time.

Thematic mapping, exemplified by land cover mapping, from remotely sensed data is typically based on an image classification (Foody, 2002). Hence, the performance of a certain classifier would become a key factor which impacts on the classification accuracy and further impacts on the quality of the derived maps. An operating classifier can be considered as a system that reduces the initial uncertainty by consuming the information contained in the input vector (Battiti, 1995). Battiti further indicates that the final uncertainty will be zero in the ideal case (i.e., the class will be certain), while it can be higher in the actual applications for at least two different reasons, i.e., insufficient input information or suboptimal operation.

Hence, the performance of a classifier is one possibility that relates to suboptimal operation. That is, even if sufficient input information is given, classification accuracy quantified by the confusion matrix may be lower than its potential value due to the insufficient training of the classifier. On the other hand, the information loss during the training period manifests a conceivable promotion space of classification accuracy.

Therefore, if a strategy is capable of compensating or reducing the information loss, an accuracy promotion of classification would be foreseeable.

Traditional spectral classification of remotely sensed images applied on a pixel-by-pixel basis ignores the potentially useful spatial information between the values of proximate pixels (Atkinson, 2000; Zhang, 2009). Geostatistical approaches, adopted in this paper, indeed aim to employ the spatial information inherent in remotely sensed images to enhance the spectral classification. Spatial information mainly serves two purposes that derive texture "wavebands" for subsequent use in classification or smooth the imagery prior to or after classification (Atkinson, 2000).

This paper utilizes two kriging approaches, i.e., simple kriging with local and cokriging, to fusion the input (e.g., spectral ) and spatial information. In the following sections, first the principles of the two kriging paradigms are revisited, then an index which assesses the potential separability of a data set is introduced, and finally the experimental results are presented and analyzed.

### 2. METHOD

Spectral response in each waveband of a remotely sensed image may be treated as a continuous variable, which is also defined as a regional variable in geostatistics. The variogram (or covariance function) is a quantitative model which reflects the relationship and spatial structure of the regional variables.

---

\* Corresponding author.

Assume  $Z(\mathbf{x})$  to be a certain attribute, i.e., a regional variable, then the experimental variogram  $\gamma(\mathbf{h})$  (generally abbreviated to variogram) may be obtained from  $\alpha = 1, 2, \dots, N(\mathbf{h})$  pairs of observations  $\{Z_\nu(\mathbf{x}_\alpha), Z_\nu(\mathbf{x}_\alpha + \mathbf{h})\}$  defined on a support  $\nu$  at locations  $\{\mathbf{x}, \mathbf{x}+\mathbf{h}\}$  separated by a fixed lag  $\mathbf{h}$ :

$$\gamma_\nu(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [Z_\nu(\mathbf{x}_\alpha) - Z_\nu(\mathbf{x}_\alpha + \mathbf{h})]^2 \quad (1)$$

where  $\mathbf{h}$  = a vector of direction and distance.

If the study area is isotropic,  $\mathbf{h}$  simply degrades to distance  $h$ . In this paper, the support  $\nu$  represents a single pixel in a remotely sensed image.

The kriging system is essentially a generalized least square regression algorithm (Goovaerts, 2002), which is characterized by the following formula:

$$\hat{Z}(\mathbf{x}) - m(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) [z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha)] \quad (2)$$

where  $\hat{Z}(\mathbf{x})$  = estimate of the regional variable  $Z(\mathbf{x})$   
 $z(\mathbf{x}_\alpha)$  = realization of  $Z$  at  $\mathbf{x}_\alpha$   
 $\lambda_\alpha(\mathbf{x})$  = weight of  $z(\mathbf{x}_\alpha)$   
 $m(\mathbf{x}), m(\mathbf{x}_\alpha)$  = expected values of  $Z$  at  $\mathbf{x}$  and  $\mathbf{x}_\alpha$   
 $n(\mathbf{x})$  = total number of realizations in the neighborhood of  $\mathbf{x}$ .

The purpose of kriging is to minimize the estimated variance  $\sigma_E^2(\mathbf{x}) = \text{Var}[\hat{Z}(\mathbf{x}) - Z(\mathbf{x})]$  under the unbiased condition, i.e.,  $E[\hat{Z}(\mathbf{x}) - Z(\mathbf{x})] = 0$ .

The regional variable  $Z(\mathbf{x})$  is usually further decomposed to two parts, formulated as follows:

$$Z(\mathbf{x}) = R(\mathbf{x}) + m(\mathbf{x}) \quad (3)$$

where  $R(\mathbf{x})$  = the residual component modelled as a stationary random function with zero mean  
 $m(\mathbf{x})$  = the trend component

### 2.1 Simple Kriging with Varying Local Mean

Traditionally, simple kriging considers the mean  $m(\mathbf{x})$  to be known and constant, i.e.,  $m(\mathbf{x}) = m$ , through the study area (Goovaerts, 2002). In other words, the mean  $m$  does not depend on location  $\mathbf{x}$  but represents global information common to all unsampled locations under the assumption of stationarity. Once the trend component  $m(\mathbf{x})$  is known and varying with the location  $\mathbf{x}$ , Eq.2 turns to simple kriging with varying local mean.

Given  $K$  land cover types of a certain study area,  $Z(\mathbf{x})$  in Eq.2 may be replaced by a Boolean variable  $I_k(\mathbf{x})$  to indicate whether the unsampled location pertaining to the predefined  $k$ th class. The indicator transform of  $I_k(\mathbf{x})$  is

$$I(\mathbf{x}; k) = \begin{cases} 1, & \text{if class is } k \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Hence, Eq.2 may be rewritten as:

$$\hat{I}_k(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) [i_k(\mathbf{x}_\alpha) - p_k(\mathbf{x}_\alpha)] + p_k(\mathbf{x}), \quad (5)$$

where  $i_k(\mathbf{x}_\alpha)$  = indicator of a training sample at location  $\mathbf{x}_\alpha$   
 $p_k(\mathbf{x}), p_k(\mathbf{x}_\alpha)$  = the *a posteriori* probability (of hard classification) or membership (of fuzzy classification), obtained by a classifier, pertaining to unsampled location  $\mathbf{x}$  and training sample at location  $\mathbf{x}_\alpha$   
 $\hat{I}_k(\mathbf{x})$  = estimate of the probability that the unsampled location  $\mathbf{x}$  pertains to the  $k$ th class.

Given  $K$  classes under consideration, Eq.5 will be repeated  $K$  times. Then a normalization process will be applied to attain  $K$  estimated probabilities, which finally constitute a  $K$ -class indicator vector, i.e.,  $\hat{I}_K(\mathbf{x})$ .

Through information fusion of the indicator vectors of training samples and the predicted probability vectors, Eq.5 aims at revising the probability vectors. It incorporates both the known categorical information of training samples and the predicted *a posteriori* information of all the pixels throughout a remotely sensed image. From the perspective of the information theory, Eq.5 adequately excavates the information contained in the input vectors (i.e., spectral features) which is partially wasted by a classifier. It is  $I_k(\mathbf{x}_\alpha) - P_k(\mathbf{x}_\alpha)$  that is the wasted information, namely the aforementioned residual. Therefore, Eq.5 utilizes the linear combinations of residuals pertaining to the training samples to amend the posterior probabilities which are directly predicted by a classifier (Zhang, 2009).

### 2.2 Cokriging

In the kriging paradigm, another algorithm to combine the primary and secondary information is cokriging. Direct measurements of the primary attribute of interest are often supplemented by secondary information in order to improve the estimation (Goovaerts, 1997).

Similarly, given  $K$  land cover types in a study area, assume  $I_k(\mathbf{x})$  to be the primary variable and  $P_k(\mathbf{x})$  to be the secondary variable, the cokriging estimate for  $I_k(\mathbf{x})$  is

$$\hat{I}_k(\mathbf{x}) = m(\mathbf{x}) + \sum_{\alpha=1}^{n_{\alpha}(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) [i_k(\mathbf{x}_{\alpha}) - m(\mathbf{x}_{\alpha})] + \sum_{\beta=1}^{n_{\beta}(\mathbf{x})} \lambda_{\beta}(\mathbf{x}) [p_k(\mathbf{x}_{\beta}) - m(\mathbf{x}_{\beta})] \quad (6)$$

where  $i_k(\mathbf{x}_{\alpha})$  = indicator transform of the class pertaining to the training sample at  $\mathbf{x}_{\alpha}$

$p_k(\mathbf{x}_{\beta})$  = posterior probabilities or memberships pertaining to all the supports  $\mathbf{x}_{\beta}$ , i.e., all the pixels

$m(\mathbf{x}_{\alpha}), m(\mathbf{x}_{\beta})$  = mean values of  $I_k(\mathbf{x})$  and  $P_k(\mathbf{x})$  at  $\mathbf{x}_{\alpha}$  and  $\mathbf{x}_{\beta}$ , respectively

$\lambda_{\alpha}(\mathbf{x}), \lambda_{\beta}(\mathbf{x})$  = weights of primary variable and secondary variable, respectively

Obviously, it is the combination of the primary and secondary information that helps improve the directly predicted precision.

### 2.3 The Upper Bound of Accuracy: Arif Index

Eq.5 and Eq.6 are the mathematical bases for the fusion of input information (e.g. spectral information) and spatial information. The common ground of the two equations is that both the input information and spatial information depend only on the training samples. In other words, the differences between the indicator vectors and the posterior probability vectors of training samples are the premise on which the method proposed in this paper will function. The information loss due to classifier consumption will result in the residuals. It is easy to encounter this circumstance in remotely sensed image classification, so the premise on which to apply the methods proposed in this paper is prone to be satisfied.

Even if an ideal classifier exists, the predicted land cover types are probably different with the ground truth due to insufficient input vectors. In other words, a maximum achievable accuracy exists in pattern classification using a particular set of features. Hence, if an upper bound of the discrimination power of input vectors can be assessed, the difference between the upper bound and the predicted accuracy of a certain classifier may be regarded as a quantitative measurement of the information wasted by the classifier. Furthermore, the quantitative measurement manifests the existence of differences between true land cover types and the posterior probabilities which ascertain the premise of the applications of Eq.5 and Eq.6.

Arif index, adopted in this paper, can be used to directly assess the maximum achievable classification accuracy of a set of input features by any classifiers (Arif, 2009). This index varies from 0 to 1, with 0 representing completely separable classes while 1 representing completely overlapping classes. In other words, as overlapping among classes increases, the value of Arif index also increases and the classification accuracy decreases.

In Figure 1, the parameter  $N$  denotes the volume of training samples, and the parameter  $\theta$  ( $\theta > 1$ ) denotes a user defined threshold which controls the strength of clustering data points of the same class near a particular data point  $y$ . The Arif index is defined as

$$Arif\ Index = \left( N - \sum_{k=1}^N S(k) \right) / N \quad (7)$$

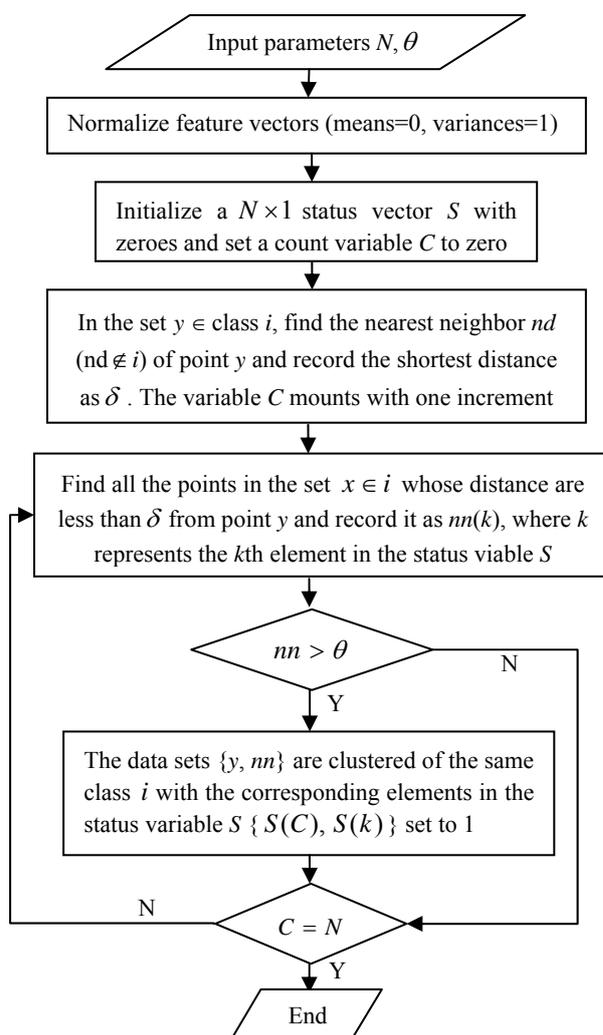


Figure 1. Flow Chart for Calculating Arif Index

Obviously, Arif index gives the ratio of data points which are not surrounded by data points of its class to the total number of data points (Arif, 2009). The relationship between Arif index and the maximum accuracy a feature set may achieve is computed as

$$Maximum\ Accuracy = 100 - (100 - Accuracy_{lower\_bound}) \times Arif\ Index \quad (8)$$

where  $Accuracy_{lower\_bound}$  = lower bound of accuracy which is the percentage of the majority class.

Hence, a linear trend can be interpolated between 100% classification accuracy and the lower bound of the classification accuracy.

### 3. EXPERIMENTS AND ANALYSIS

The land cover data used in this paper is acquired from the National Land Cover Dataset 2001 (abbreviated to NLCD2001) for conterminous United States. Sixty-five mapping zones and sixteen land cover types are involved in NLCD2001. All NLCD2001 products were generated from a standardized set of data layers mosaiced by mapping zone. Typical zonal layers included multi-season Landsat-5 TM and Landsat-7 Enhanced Thematic Mapper (ETM+) imagery centred on a nominal collection year of 2001 (Homer, 2007). All of the images are geo-registered to the Albers equal area projection grid, and resampled to 30m grid cells.

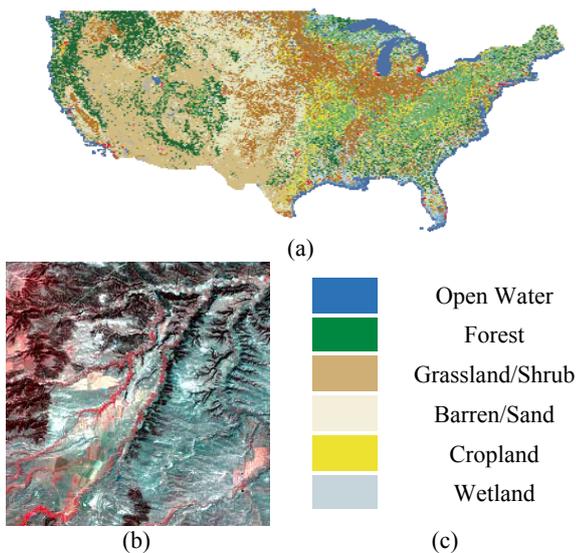


Figure 2. (a): Land cover types in NLCD2001 database; (b) and (c): an ETM+ image and six land covers of the test area

Because all the images utilized in NLCD2001 are provided by U.S. Geological Survey (USGS) Centre for Earth Resources Observation and Science (EROS), a corresponding Landsat 7 ETM+ image was downloaded from EROS. It was acquired on July 13, 1999. After the registration with the land cover data, a study area ranging from the latitude of  $47^{\circ}41' N$  to  $47^{\circ}54' N$ , and from the longitude  $109^{\circ}01' W$  to  $109^{\circ}21' W$  was clipped from the ETM+ image. The clipped image is made up of 500 by 500 pixels (Figure 2(a)). In our experiment, six classes including open water, forest, grassland/shrub, barren/sand, cropland and wetland are chosen from the NLCD2001 (Figure 2(b)). The tasseled cap transformation is applied to the ETM+ image after which the two components of soil brightness and greenness are selected.

Thanks to NLCD2011, the ground truth of each pixel in the image is known. Hence, 1500 training samples are random selected with others allocated as testing samples. The ratio of training samples to testing samples is rather small. The support vector machine (abbreviated to SVM) is utilized as the classifier in this paper. And the package **libsvm** interfacing with the statistical software **R** is adopted to implement  $K$ -class ( $K > 2$ ) land cover classification (Meyer, 2009).

For this Data set, the estimated Arif index is 0.411 which manifests a moderate separability and corresponding to the potential highest accuracy of 79.47%. Compared to the overall accuracy of 73.79% obtained by the SVM classifier, 79.47% not only reflects that part of the information is consumed by the classifier, but indicates an improvable accuracy of about 5%. It is easy to compute indicator transforms for training samples of known class labels. And, after prediction of the posterior probabilities pertaining to six classes by the SVM classifier, the residuals can be calculated as differences between binary indicators and predicted class probabilities. Table 1 lists the variogram models of simple kriging with local mean which reflect the spatial distribution of residuals of the training samples. And it also shows the variogram models of the primary variable and the secondary variable, and the covariogram of the cokriging method. It exhibits the spatial variation of the target variable at the locations of training samples and testing samples (i.e., all pixels except for training samples in this experiment), respectively. The trend of the cokriging method in this paper is obtained by applying spatial smoothing to the posterior probabilities.

In general, as is shown in Table 2, the cokriging method obtains a considerable improvement in overall accuracy and kappa coefficient, and simple kriging with local mean is no exception and even more effective. The former achieves 2 percents improvement in overall accuracy and an increase in kappa coefficient from 0.58 to 0.65, while the latter witnesses a 5% accuracy increase and an improved kappa coefficient of 0.68. The reason the SK method gains higher accuracies than cokriging may be that the trends of the primary and secondary variables of cokriging are obtained through smoothing in spatial domain, while the trend of the SK method is localized to each pixel. Therefore, the residuals are more accurate. Moreover, the latter demands less variogram models and thus costs less time for modelling. Therefore, the method of simple kriging with local mean is more worthy of recommendation for the fusion of input information and spatial information.

Furthermore, the SK method is made as an example to account for the effects of kriging paradigm. Four groups of testing samples, each of which contains fifteen samples with ground truth as farmland but classified as other land cover types are randomly chosen and exhibited, shown in Figure 3. The posterior probabilities predicted by the SVM classifier and those revised by the SK method are compared in this figure. Generally, the posterior probabilities obtained by the classifier would first be corrected by residuals and then be normalized. However, in order to more clearly reveal the probability fluctuations before and after residual corrections, the normalization procedure was skipped over. In Figure 3, the abscissa represents the number of testing samples, while the ordinate denotes the probabilities. For the selected testing samples, the red circle ● denotes probabilities pertaining to the land cover type of farmland after SVM prediction, while the black triangle ▲ represents the highest posterior probabilities pertaining to the prevailing class type other than farmland after the prediction of SVM. Figures 3(a)-(b) illustrate that the classifier failed to make accurate predictions. Correspondingly, the reversed purple triangle ▼ denotes the probabilities pertaining to farmland after residual corrections by the SK method; while the blue squares ■ represent the revised ones corresponding to the original black triangles ▲.

Class	SK	Cokriging		
	Variogram	Variogram of primary variable	Variogram of secondary variable	Covariogram
Open water	0.0007 Sph(30)	0.001 Sph(30)	0.0003 Sph(30)	0.0003 Sph(30)
Forest	0.046 Exp(21)	0.044 Exp(58)	0.011 Exp(60)	0.005 Gau(250)
Grassland/Shrub	0.012 Gau(37)	0.122 Exp(44)	0.018 Sph(50)	0.008 Exp(100)
Barren/Sand	0.130 Gau(53)	0.130 Exp(80)	0.011 Gau(65)	0.005 Gau(180)
Cropland	0.056 Gau(90)	0.056 Exp(430)	0.0009 Exp(75)	0.0002 Exp(100)
Wetland	0.019 Exp(92)	0.019 Gau(100)	0.001 Gau(200)	0.0007 Gau(100)

Table 1. Variograms and covariograms of SK and cokriging

Method	SVM Classifie		Cokriging		Simple Kriging with Local Mean	
	Producer's Accuracy	User's Accuracy	Producer's Accuracy	User's Accuracy	Producer's Accuracy	User's Accuracy
Open water	37.29	53.14	24.41	48.65	31.86	50.81
Forest	78.95	84.17	76.52	81.59	77.95	83.12
Grassland/Shrub	57.36	66.33	58.86	64.91	61.29	64.47
Barren/Sand	91.96	74.51	86.25	82.76	87.53	81.82
Cropland	15.92	59.75	73.28	76.11	72.80	73.93
Wetland	10.32	41.80	35.98	20.51	27.16	54.83
Overall Accuracy	73.79		75.77		77.02	
Kappa Coefficient	0.58		0.63		0.65	

Table 2. Accuracy Assessment Indexes of SVM Classification and Two Kriging Methods (Accuracy Unit: %)

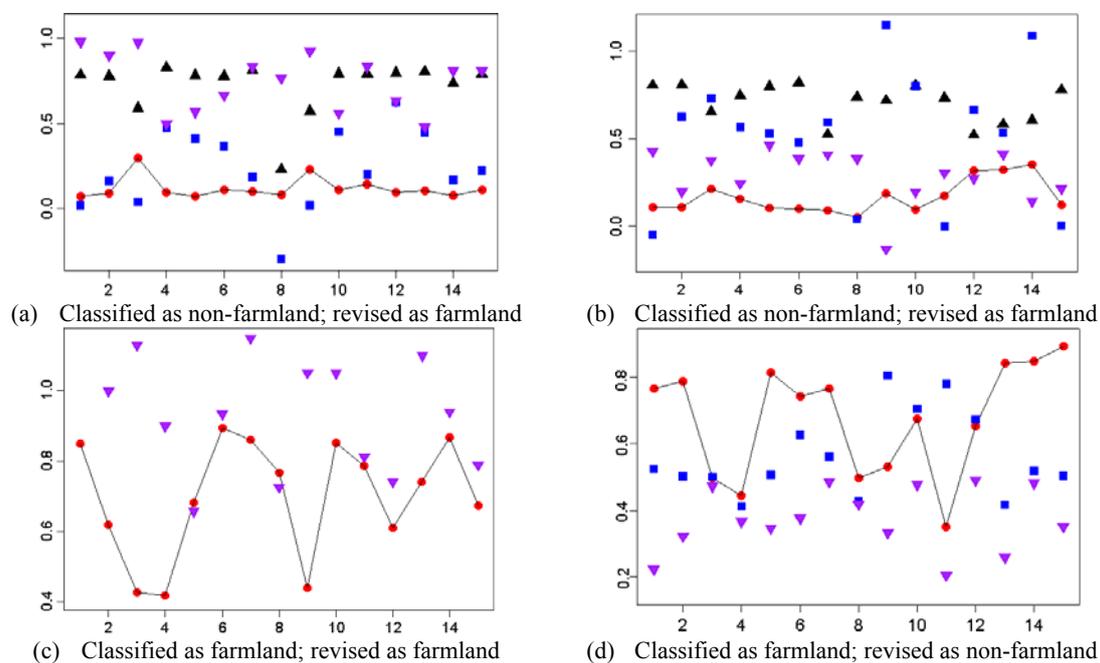


Figure 3. The Probabilities of the SVM Classifier and the SK Method (e.g., Farmland)

On the whole, Figures 3(a)-(c) reflect the effect of the SK method, i.e., it utilizes the information of spatial distribution provided by training samples to improve the posterior probabilities pertaining to the target land cover type and accordingly reduce those pertaining to the confusing types. Take Figure 3(a) for example. Given the testing samples with

ground truth as farmland, the predicted posterior probabilities of farmland are less than those of other confusing types, which results in omission. However, after the residual corrections, the probabilities of farmland are improved with a relative probability decrease of other types, which may also be reflected by the producer's accuracy in Table 2. In addition, as is shown

in Figure 3(b), the increased posterior probability pertaining to the right type is still not predominant to exceed the decreased probability pertaining to the confusing types. However, the effect that the SK method helps to improve the predicted probability of the right land cover type is still traceable.

In Figure 3(d), the SK method fails to further improve the originally prevailing posterior probabilities pertaining to the right type, but instead meets the exact reverse. It may come down to two reasons: (1) the input vectors are prone to confusion which results in the slight advantage in the posterior probability of the right type; (2) the training samples in a local domain are sparse and unevenly distributed which results in an unfaithful modelling of residual variation and a naive spatial interpolation. That is, the less accurate residual variogram models would easily reverse the slight advantage. However, samples of this kind in Figure 3(b) only occupy 1.1 percent of the total samples. On the contrary, samples of the kind in Figure 3(a), which are poorly classified but correctly revised, occupy 58 percent.

Moreover, compared to the producer's and user's accuracy acquired by the SVM classification, as are listed in Table 2, the corresponding accuracy fluctuations after the application of the SK method may not equivalent to mean that the kriging method is particularly suitable to some certain land cover types.

In order to further testify the efficiency of the kriging paradigm, another TM image including 17 land cover types is adopted. A total of ten variables were available: Landsat TM channels 1-5, 7, modified normalized difference vegetation index (MNDVI), scaled elevation, slope in degrees, and a combined slope-aspect variable. Further detail of this data set is documented in Zhang and Goodchild (2007). The estimated Arif index manifests a corresponding highest classification accuracy of 72.22%. The overall accuracy and kappa coefficient achieved by generalized linear model (abbreviated to GLM) are 65.55% and 0.62, respectively. After residual corrections, the former is improved to 75.45% and the latter is increased to 0.73. It is interesting to notice that the revised accuracy of 75.45% is larger than the estimated potential highest accuracy of 72.22%. The is that the potential highest is just estimated by the input vectors without considering the introduced spatial information during the post-classification corrections.

#### 4. CONCLUSION

The proposed two kriging methods are independent of the specific classifiers initially adopted for image classification. Hence, these methods may be treated as post-classification approaches. They aim at compensating part of the information consumed by classifiers due to the insufficient learning process. Moreover, these methods are independent of the land cover types, although the improvements vary in the producer's and user's accuracies of different land covers. The factors, including the sampling methods, sample sizes, and sample distributions, need to be further investigated, for they are close related to the spatial information of the training samples.

#### REFERENCE

Atkinson P. M., Lewis P., 2000. Geostatistical classification for remote sensing: an introduction. *Computers and Geosciences*, 26, pp. 361-371.

Arif M., Afsar F. A., Akram M. U., et al., 2009. Arif index for predicting the classification accuracy of features and its application in heart beat classification problem. In: *Advances in Knowledge Discovery and Data Mining: proceedings of the 13th Pacific-Asia Conference*, Thailand.

Battiti R., 1995. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4), pp. 537-551.

Food and Agriculture Organization of the United Nations, 2000. Land Cover Classification System (LCCS): Classification Concepts and User Manual. <http://www.fao.org/docrep/003/x0596e/x0596e00.htm>(2011)

Foody G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80, pp.185-201.

Goovaerts P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, pp. 125-233.

Goovaerts P., 2002. Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. *Geographical systems*, 4, pp. 9-111.

Homer C., Dewitz J., Fry J., et al., 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4), pp. 337-341.

Homer C., Huang C. Q., Yang L. M., et al., 2004. Development of a 2001 National Land-Cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70(7), pp. 829-840.

Meyer D., 2009. Support Vector Machines-The Interface to Libsvm in Package e1071. <http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf> (2011)

Zhang J. X., 2009. *Scale, Uncertainty and Fusion of Spatial Information*. Wuhan University Press, Wuhan, pp. 221-234.

Zhang J. X., Goodchild, M. F., Steele, B. M. and et al., 2007. A discriminant space-based framework for scalable area-class mapping. In: *Proceedings of SPIE-The International Society for Optical Engineering*, Nanjing, China, Vol. 6751.

#### ACKNOWLEDGEMENT

The research is partially supported by "973 Program" grants (No. 41071286) and (No. 41171346).