Spatial discrimination analysis of land-cover categories

Ryuei NISHII* and Shojiro TANAKA**

*Faculty of Integrated Arts and Sciences, Hiroshima University
**Faculty of Engineering, Shimane University

Abstract — We examine classifiers of land-cover categories based on multispectral data. First, a parameter-estimation method by training data consisting of pure and mixed cells is discussed. By actual data, it is shown that the use of mixels in the training data improves unmixing results significantly. Next, classifiers taking spatial continuity of categories into account are examined through actual Landsat data. They show an excellent performance in discrimination, especially in unmixing.

Key words – linear discriminant function, penalized likelihood, spatial continuity.

I. Introduction

Statistical discriminant procedures based on multispectral images are widely used for land-cover classification. If data follow multivariate normal distributions with a common variance-covariance matrix, then mean vectors and the variance-covariance matrix are estimated by training data, and a linear discriminant function (LDF) will be used for discrimination.

In this case, ordinary methods for parameter estimation use pure cells only in the training data. Hence we must avoid mixels even though their mixture proportions of the categories are known. Consequently, only half of the whole training data or less may be used for parameter estimation, especially in low-resolution case.

Obviously, such a strategy losses much information on respective categories, and it is inefficient to estimate the parameters based on the sparse pure-cell data. Nishii [6] proposed an additive model for mean vectors and derived a parameter-estimation procedure based on multivariate regression models. We note that the linear unmixing for estimating category-proportions of a given pixel is a different issue, see Section III for detail.

In this article, we examine a classifier by [6] through actual data. His parameter-estimation method uses all information on the training data including mixels. Further, the proposed classifier takes spatial continuity of categories into account. See McLachlan [5] for statistical image analysis including classification, and Cressie [1] for general spatial statistics.

In Section II, the parameter-estimation method based on training data consisting of pure and mixed cells due to [6] is illustrated. Then, Section III gives classification procedures based on a penalized sum of squared errors (PSSE). Test data are classified by taking the spatial continuity of categories into account. Section IV introduces a measure assessing accuracy of unmixing procedures, which is an extension of the over-all accuracy in discriminant analysis. Finally, in Section V, the effect of using mixel information is examined through actual Landsat data, and the efficiency is shown. Concluding remarks are given in Section VI.

II. Parameter estimation based on pure and mixed cells

We take statistical procedures for discrimination or unmixing of land-cover categories. Consider K land-cover categories $C_1, \ldots, C_K$. We assume that a p-dimensional spectral vector z observed at a pure pixel from C_k follows a p-variate normal distribution with mean vector $\mu_k : p \times 1$ and common variance-covariance matrix $\Sigma : p \times p$, denoting $z \sim N(\mu_k, \Sigma)$ for $k = 1, \ldots, K$. Next, we assume distributions on mixles as follows.

Let s be a mixel whose proportions covered by the categories C_k are known as $a^{(k)} \geq 0$ for $k = 1, \ldots, K$ with $\sum_{k=1}^{K} a^{(k)} = 1$. Then, we suppose that a spectral vector z observed at the mixel s is normally distributed with mean vector $\sum_{k=1}^{K} a^{(k)} \mu_k$ and Σ in common. Further, all spectral data z's are assumed to be independent.

We call a $K \times 1$ vector $a \equiv (a^{(1)}, \ldots, a^{(K)})'$ a category-proportion vector of the pixel s, where A' denotes the transposition of a vector/matrix A. By the assumption, an expected vector of z is expressed by

$$E(z) = \sum_{k=1} a^{(k)}\mu_k = Ma \quad \text{with} \quad M \equiv [\mu_1, \dots, \mu_\kappa] \qquad (1)$$

where M is a $p \times \kappa$ mean matrix. Thus, $z \sim N_p(Ma, \Sigma)$.

The relation (1) for training data is considered as a multivariate regression model with a response vector z and an explanatory vector a. Hence, the unknown mean vectors $\mu_1, \dots, \mu_\kappa$ and the unknown variance-covariance matrix Σ are estimated through training data.

Consider training data $z_1, \dots, z_n$ of size n observed at pure or mixed cells $s_1, \dots, s_n$, where z_α are p-dimensional vectors for $\alpha = 1, \dots, n$. Let $a_\alpha^{(k)}$ $(k = 1, \dots, \kappa)$ denote a known proportion such that the category C_k covers the pixel s_α. A vector $a_\alpha \equiv (a_\alpha^{(1)}, \dots, a_\alpha^{(\kappa)})' : \kappa \times 1$ is called a category-proportion vector of s_α. Then, by the distributional assumption, z_α's are independently distributed as normal distributions:

$$z_\alpha \sim N_p(Ma_\alpha, \Sigma) \quad \text{for} \quad \alpha = 1, \dots, n. \qquad (2)$$

For the sake of expressing the joint distribution, we define a spectral matrix and a category-proportion matrix as:

$$Z \equiv \left[z_1, \dots, z_n\right] : p \times n; \quad A \equiv \left[a_1, \dots, a_n\right] : \kappa \times n. \quad (3)$$

By the distribution (2), an expected matrix of Z is expressed by MA. Hence, column vectors of a $p \times n$ matrix $U \equiv Z - MA$ are independently and identically distributed with $N_p(0, \Sigma)$.

Thus, the likelihood function based on the training data is expressed by

$$|2\pi\Sigma|^{-n/2} \exp\left[-\mathrm{tr}\left\{\Sigma^{-1}(Z - MA)(Z - MA)'\right\}/2\right] \qquad (4)$$

because it holds $\sum_{\alpha=1}^{n}(z_\alpha - Ma_\alpha)'\Sigma^{-1}(z_\alpha - Ma_\alpha) = \mathrm{tr}\left\{\Sigma^{-1}(Z - MA)(Z - MA)'\right\}$, where "tr(B)" denotes the trace of a squared matrix B.

By maximizing (4) with respect to M and Σ, the exact maximum likelihood estimates (MLEs) are obtained by [6] as:

$$\hat{M} = ZA'(AA')^{-1} \equiv \left[\hat{\mu}_1, \dots, \hat{\mu}_\kappa\right] \qquad (5)$$

$$\hat{\Sigma} = Z\left\{I_n - A'(AA')^{-1}A\right\}Z'/n. \qquad (6)$$

Distributions of the MLE's are also found there. See Eaton [2] for multivariate linear models.

The MLEs in (5) and (6) are always derived as long as the rank of the category-proportion matrix A in (3) is $\kappa (\le n)$. If the training data consist of pure pixels only, then MLE (5) of the mean vector μ_k is nothing but the sample mean of the observed vectors from C_k. By numerical studies in Section V, we will confirm that the mixel training data are useful for statistical inference of the categories.

III. Classifiers derived by penalties

Suppose that we are required to estimate a category-proportion matrix A^* of m pixels $s_1^*, \dots, s_m^*$ by using spectral data $z_\beta^* : p \times 1$ observed at the pixels s_β^* for $\beta = 1, \dots, m$. Put $Z^* \equiv [z_1^*, \dots, z_m^*] : p \times m$. Then, the MLE of $A^* : \kappa \times m$ is derived by minimizing the sum of squared errors (SSE):

$$Q(A^*) \equiv \mathrm{tr}\left\{\Sigma^{-1}\left(Z^* - MA^*\right)\left(Z^* - MA^*\right)'\right\} \qquad (7)$$

under one of the following conditions:

(D) Discrimination: all entries of $\hat{A}^*$ consist of zero or one, and $1_\kappa' \hat{A}^* = 1_m'$.

(U) Unmixing: all entries of $\hat{A}^*$ are non-negative, and $1_\kappa' \hat{A}^* = 1_m'$.

Consider the case $m = 1$. Put $Z^* = z^*$ and $a^{(k)} = (0, \dots, \overset{k}{1}, \dots, 0)'$. Then the quadratic form (7) for $a^{(k)}$ is reduced to $Q(a^{(k)}) = (z^* - \mu_k)'\Sigma^{-1}(z^* - \mu_k)$. Hence, the minimization of $Q(a^{(k)})$ with respect to k is equivalent to the classification result based on the ordinary linear discriminant function (LDF). In general, MLE of the category-proportion matrix A^* with 0-1 entries coincides with the classification based on LDF.

When we try to estimate category-proportions through an observed vector z^*, MLE of the category-proportion vector $a = (a^{(1)}, \dots, a^{(\kappa)})'$ is derived by minimizing the quadratic form weighted by the inverse matrix of the variance-covariance matrix Σ as

$$Q(a) = \left(z^* - \sum_{k=1} a^{(k)}\mu_k\right)' \Sigma^{-1} \left(z^* - \sum_{k=1} a^{(k)}\mu_k\right). \qquad (8)$$

A usual unmixing procedure takes the weight matrix in (8) by the unit matrix, see e.g. Hu, Lee and Scarpace [3]. Our procedure would be, however, more suitable because dispersions of spectral variables are normalized. Note that the formula (8) is the squared Mahalanobis distance.

However, it is reported by [6] that the estimation of A^* by LDF gives poor results because LDF ignores the spatial-continuity of the categories and the estimated matrix overfits to test data. Using the geometrical information, a penalized approach to SSE was taken there. The penalized SSE (PSSE) is as follows:

$$PS(A^*; \lambda) = Q(A^*) + \lambda\, \mathrm{tr}\left(M'\Sigma^{-1}MA^*PA^{*\prime}\right) \qquad (9)$$

where $Q(A_*)$ is SSE defined in (7), P is an $m \times m$ positive semi-definite matrix assessing the spatial variability of the categories, and $\lambda \geq 0$ is a trade-off parameter between the fitness and the spatial smoothness. A penalized method for segmentation of intensity SAR images is treated by Smits and Dellepiane [8].

A. An example of the penalty matrix P

When a test area is a rectangular region of size $u \times v$, the penalty matrix $P : uv \times uv$ due to [6] is given by

$$P \equiv D_u \otimes I_v + I_u \otimes D_v \qquad (10)$$

where

$$D_t \equiv \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} : t \times t \qquad (11)$$

for $t = u, v$. The penalty based on (10) implies that no penalty is added to SSE if two adjacent pixels belong to the same category. If they belong to different categories, for example C_1 and C_2, then penalty $\lambda(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$ is added. Variability of mixels are similarly evaluated. Thus, P assesses the spatial variability of the categories of all pairs of two adjacent pixels (first-order neighborhoods).

In general, the matrix P should be chosen so that the penalty term of (9) evaluates a sort of quantity orthogonal to SSE.

B. Estimation of category-proportions

The matrix A_* minimizing $PS(A_*; \lambda)$ under the constraint $1_K' A_* = 1_m'$ is found by Lagrange's multiplier method, see [3] for the case without penalty ($\lambda = 0$). Unfortunately, this solution of A_* has negative entries and shows poor performance by the simulation study due to [6]. Second approach is to find the optimal matrix by the solution of the differential equation: $\partial PS(A_*; \lambda) \big/ \partial A_* = 0$. Consequently, we have the matrix equation:

$$M'\Sigma^{-1}MA_* = M'\Sigma^{-1}Z^*(I_m + \lambda P)^{-1}. \qquad (12)$$

The estimate $\hat{A}_*$ should satisfy one of the two restrictions (D) and (U) following the formula (7). The solution of the matrix equation (12) is obtained by m linear equations derived by respective columns of A_*.

In the case $\lambda = 0$, $\hat{A}_0$ is the MLE of A_* derived by LDF, and an estimation procedure in terms of the linear unmixing is discussed by [3].

C. Other discrimination procedures

Other typical discrimination method is based on a quadratic discriminant function (QDF). This is derived when the categories C_k are supposed to follow multivariate normal distributions with respective variance-covariance matrices, say $N_p(\mu_k; \Sigma_k)$.

Another method is due to Switzer [9] called the smoothed method. He introduced the method when the observed vectors are spatially dependent. But, we review it under the spatially-independent assumption for use in Section V. See Mardia [4] for its statistical refinement.

s_6	s_2	s_5
s_3	s_0	s_1
s_7	s_4	s_8

Fig. 1. A neighborhood of s_0 with eight adjacent pixels

Suppose that we try to discriminate the center pixel s_0. We assume that all pixels in the window of Fig. 1 came from one of the normal populations $N_p(\mu_k; \Sigma); k = 1, \ldots, K$ at the same time (the local-continuity assumption of the categories). Then, s_0 is allocated to the category C_k which maximizes the joint density function of nine spectral vectors observed at the window. This method is abbreviated as Switzer8.

Thus, Switzer's method is based on decision by majority, whereas PSSE in (9) can be regarded a penalized likelihood. Prior distributions of the category-proportion matrix A_* are found in [6].

D. Other unmixing procedures

The discrimination methods LDF and QDF are also used for unmixing by using posterior probabilities. By our normality assumption in Section II, the posterior probabilities $\hat{p}_k$ such that a vector z^* came from the category C_k are given by

$$\hat{p}_k = \phi(z^*; \mu_k; \Sigma) \Big/ \sum_{\ell=1}^{K} \phi(z^*; \mu_\ell; \Sigma) \qquad (13)$$

for $k = 1, \ldots, K$, where

$$\phi(z; \mu; \Sigma) = |2\pi\Sigma|^{-1/2} \exp\{-(z - \mu)'\Sigma^{-1}(z - \mu)/2\}$$

is a density function of $N_p(\mu; \Sigma)$.

If the categories C_k have individual variance-covariance matrices Σ_k, exactly $N_p(\mu_k;\Sigma_k)$ for $k = 1,\ldots,\kappa$, the posterior probabilities $\hat a_k$ are given by

$$\hat a_k = \phi(z^*;\mu_k;\Sigma_k) \Big/ \sum_{\ell=1}^{\kappa} \phi(z^*;\mu_\ell;\Sigma_\ell). \tag{14}$$

The unmixing procedures (13) and (14) are respectively abbreviated by P_{LDF} and P_{QDF}.

IV. Accuracy assessment of classifiers

In discriminant analysis, various measures for accuracy assessment in terms of error matrices are proposed, see e.g. Nishii and Tanaka [7]. Here, we introduce a measure for assessment of unmixing. The measure is an extension of the over-all accuracy and a function of category-proportion matrices.

Let $A_* = \left(a_\beta^{(k)} : \kappa \times m\right)$ be an actual category-proportion matrix and $\hat A_* = \left(\hat a_\beta^{(k)} : \kappa \times m\right)$ be its estimate, where m denotes a sample size of the test data. Then, the absolute distance between two matrices meets the inequality:

$$\sum_{k=1}^{\kappa}\sum_{\beta=1}^{m} \left|\hat a_\beta^{(k)} - a_\beta^{(k)}\right| \le \sum_{k=1}^{\kappa}\sum_{\beta=1}^{m} \left(\hat a_\beta^{(k)} + a_\beta^{(k)}\right) = 2m \tag{15}$$

because $\sum_{k=1}^{\kappa} a_\beta^{(k)} = \sum_{k=1}^{\kappa} \hat a_\beta^{(k)} = 1$, see the restrictions (D) and (U). Standardizing (15) and changing the sign, we define the goodness of the estimated matrix $\hat A_*$ by

$$g(\hat A_*; A_*) \equiv 1 - \sum_{k=1}^{\kappa}\sum_{\beta=1}^{m} \left|\hat a_\beta^{(k)} - a_\beta^{(k)}\right| \Big/ (2m). \tag{16}$$

It holds that $0 \le g(\hat A_*; A_*) \le 1$, and $g(\hat A_*; A_*) = 1$ means the complete discrimination or unmixing.

In the discrimination case, or equivalently $a_\beta^{(k)}$ and $\hat a_\beta^{(k)}$ take only 0 or 1, we have

$$1 - \sum_{k=1}^{\kappa} \left|\hat a_\beta^{(k)} - a_\beta^{(k)}\right| \Big/ 2 = \begin{cases} 1 & \text{correct discrimination} \\ 0 & \text{wrong discrimination} \end{cases} \tag{17}$$

for $\beta = 1,\ldots,m$. Hence, $g(\hat A_*; A_*)$ is an averaged value of (17), and coincides with the over-all accuracy. So, we call the measure an extended over-all accuracy.

V. Applications to Landsat data

A geocoded Landsat image on December 20th, 1989 in Saitama, Japan used for our experiment. Using TM data of 30m resolution for classification, we aggregate land-use data of 10m resolution and yield data consisting of mixels with nine-level gradient of 30m resolution, see Fig. 2.

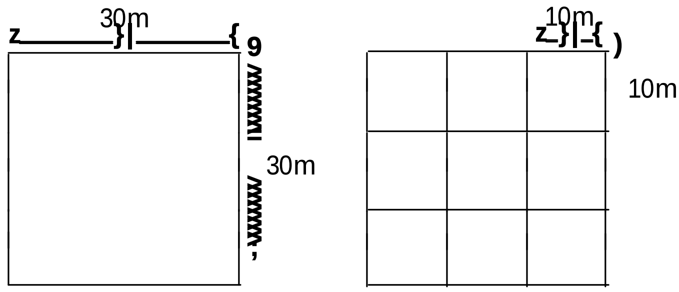


Fig. 2. Resolutions of TM data (left) and of the detailed digital land-use data (right)

Fifteen land-use categories are transformed and merged into five land-cover categories ($\kappa = 5$): C_1 : vegetation, C_2 : barren grounds (paddy field), C_3 : developed areas, C_4 : residential area, and C_5 : water area. Fig. 3 gives the image based on the detailed digital land-use data of size 9km$\times$9km with 10m resolution. In Fig. 3, pixels with 10m square are colored by corresponding categories based on the detailed digital land-use data. There are 90000 pixels with 30m square in the whole area, and the number of pure cells is only 38901.

We partition the whole area of size 300×300 as Fig. 3, and set G1a, G1d of size 50×50, G2a of size 100×100, and G3 of size 200×200 to training areas. A test area is set to T1 of size 100×100 in this experiment.

For assuring the independence between training data and test data, we choose separate regions. We classify the test area T1 by using one of the training areas G1a, G1d, G2a and G3.

We examine the effect of mixel information and compare the classifiers. Relations between training data used for parameter estimation and test data classified by the classifiers are summarized in Table 1.

Table 1. Training data and test data used in Tables 2-4.

Training data	Test data: pure pixels	Test data: all pixels
pure pixels	Tables 2, 3	Table 3
all pixels	Table 3	Tables 3, 4

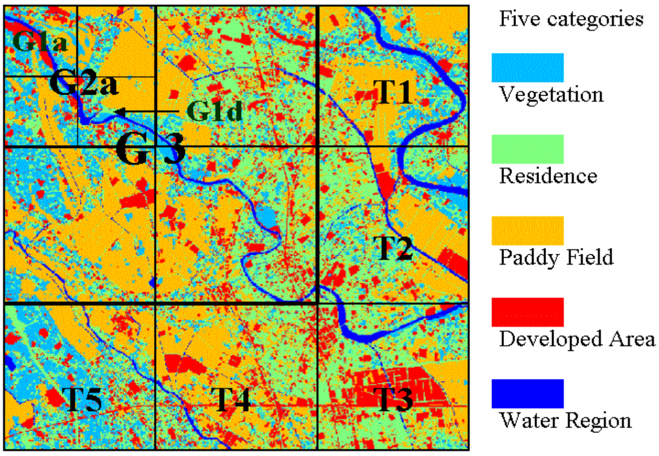


Fig. 3. Detailed digital land-use data and partition of the region of size 9km×9km with 10m resolution, Saitama, Japan, Dec. 1989

A. Discrimination by pure pixels of training data

Using the data, we study how mixel information of the training data is useful for discrimination. We examine four classifiers: LDF whose result coincides with MLE in this case, QDF, Switzer8 (see C of Section III for both), and the proposed PSSE method (A_λ with 0-1 entries).

In QDF, the variance-covariance matrices of the category C_k is estimated by the sample variance-covariance matrix of data observed at pure pixels from C_k. Similar to Switzer8, Switzer4 was also proposed, and we examined the performance of Switzer4. However, numerical examples suggest the superiority of Switzer8. Hence, results on Switzer4 are omitted in this section. In PSSE, the trade-off parameter is set by $\lambda = K \times \mathrm{rank}(P)/\bar{p}$, where $\bar{p}$ is a mean of the penalty terms due to randomly-generated category-proportion matrices.

We discriminate 4433 pure cells in the test area T1 by 1164 pure cells in the training area G1a. The sample sizes of the training data are given by $(n_1, \ldots, n_5) = (459, 2262, 465, 777, 470)$. Error matrices X_L, X_Q, X_S and X_P due to LDF, QDF, Switzer8 and PSSE are respectively obtained by

$$X_L = \begin{pmatrix} 179 & 55 & 55 & 126 & 44 \\ 58 & 1794 & 252 & 152 & 6 \\ 34 & 103 & 140 & 159 & 29 \\ 82 & 33 & 107 & 513 & 42 \\ 63 & 24 & 93 & 32 & 258 \end{pmatrix}; \quad X_Q = \begin{pmatrix} 172 & 93 & 41 & 99 & 54 \\ 62 & 1979 & 93 & 121 & 7 \\ 59 & 51 & 215 & 119 & 21 \\ 116 & 78 & 136 & 405 & 42 \\ 73 & 33 & 79 & 46 & 239 \end{pmatrix} \quad (18)$$

$$X_S = \begin{pmatrix} 234 & 47 & 51 & 116 & 11 \\ 53 & 1936 & 174 & 97 & 2 \\ 33 & 111 & 147 & 171 & 3 \\ 90 & 22 & 104 & 553 & 8 \\ 58 & 8 & 120 & 21 & 263 \end{pmatrix}; \quad X_P = \begin{pmatrix} 88 & 74 & 30 & 227 & 40 \\ 14 & 2079 & 68 & 95 & 6 \\ 10 & 119 & 114 & 193 & 29 \\ 14 & 50 & 67 & 621 & 25 \\ 19 & 10 & 107 & 48 & 286 \end{pmatrix}. \quad (19)$$

Rows and columns of the error matrices are corresponding to the ground-truth and the classified categories respectively. For example, the sum of first rows of all error matrices is equal to $459 = n_1$.

The accuracy of the error matrices is assessed through the over-all accuracy, the class-averaged accuracy, and the measure J_{pro} proposed by [7]. It is defined by entropy as

$$J_{pro} \equiv \prod_{k=1} \left(\frac{x_{kk} + 1/2}{n_k + 1/2} \right)^{n_k/n}$$

where x_{kk} denote k^{th} diagonal elements of the error matrix and $n = n_1 + \cdots + n_K$. Accuracy measures of (18), and (19) are tabulated in Table 2.

Table 2. Accuracy assessment of error matrices
Training data: 1164 pure cells in G1a:50 × 50
Test data: 4433 pure cells in T1:100 × 100

Assessments	Error matrices			
	X_L	X_Q	X_S	X_P
Over-all	.6506	.6790	.7067	.7192*
Class-averaged	.5387	.5483	.5906*	.5527
Jpro	.6190	.6451	.6751*	.6346

* denotes the best value in each accuracy assessment.

The error matrices (18), (19) and Table 2 imply that the proposed PSSE gives better results for the categories C_2 and C_4, but poor for C_1. Thus, PSSE attaches importance to large categories.

B. Discrimination by pure/mixed cells

Next, we discriminate pure pixels only or all pixels in the test area T1 based on pure pixels only or all pixels in four training areas, see Table 3. Category-proportion matrices are estimated under the restriction (D) in Section III, and discrimination results are assessed by the extended over-all accuracy (16). In QDF, the mean vectors are estimated by (5), and respective variance-covariance matrices are estimated by pure-pixel data from corresponding categories.

The use of mixel information improves all discrimination results due to LDF and QDF. The procedure PSSE shows the best performance with small training data, and QDF comes top with large training data.

Fig. 4 illustrates the discriminated results of the test area T1 by LDF, Switzer8 and PSSE based on pure pixels only (left) and all pixels (right) of the training areas G1a, G1d, G2a and G3. The water region is misclassified to developed area. This comes from that a part of the riverbed is dried up. The method LDF detects spatially-

Table 3. Extended over-all accuracies of discrimination methods
Training data: pure cells only or all pixels in G1a, G1d, G2a, G3
Test data: pure cells only (sample size = 4433), or all pixels (sample size = 10000) in T1

Training areas			Discrimination methods							
			LDF		QDF		Switzer8		PSSE	
code	cells	no. of cells	T1 pure	T1 all	T1 pure	T1 all	T1 pure	T1 all	T1 pure	T1 all
G1a	pure	1164	.6506	.4960	.6790	.5053	.7067	.5402	.7192	.5459
G1a	all	2500	.6619	.5021	.6803	.5102	.7011	.5373	.7232¤	.5511y
G1d	pure	1127	.5615	.4396	.6397	.4963	.6000	.4726	.6519¤	.5007
G1d	all	2500	.5964	.4562	.6510	.5039y	.6291	.4898	.6375	.4884
G2a	pure	4919	.5682	.4486	.7029	.5368	.6172	.4848	.6736	.5236
G2a	all	10000	.5766	.4501	.7031¤	.5377y	.6165	.4822	.6429	.4969
G3	pure	17623	.6253	.4739	.6966	.5328	.6578	.5041	.6614	.5049
G3	all	40000	.6325	.4772	.6979¤	.5330y	.6704	.5093	.6650	.5056

¤ denote the best values for discriminating pure cells in the test area T1 based on respective training data.
y denote the best values for discriminating all cells in the test area T1 based on respective training data.

local changes (over¯tting to the test data), but the Switzer and PSSE show spatially-stable performances because they take spatial-continuity of the categories into account.

C. Unmixing of all pixels in T1

Finally, all pixels in the test area T1 are unmixed based on pure pixels only, or all pixels in the four training areas. Table 4 gives accuracy measures (16) of unmixing procedures: P_{LDF}, P_{QDF} and PSSE, see B and D of Section III. The posterior probabilities of the training data $z^{¤}$ are calculated by (13) and (14).

Table 4. Extended over-all accuracies of unmixing methods. Training data: pure cells only or all pixels in G1a, G1d, G2a, G3; Test data: all data in T1

Training areas			Unmixing methods		
code	cells	n	P_{LDF}	P_{QDF}	PSSE
G1a	pure	1164	.4808	.5194	.5721
G1a	all	2500	.4871	.5244	.5803¤
G1d	pure	1127	.4644	.5249	.5158
G1d	all	2500	.4836	.5337¤	.5205
G2a	pure	4919	.4473	.5599	.5495
G2a	all	10000	.4512	.5616¤	.5340
G3	pure	17623	.4783	.5431	.5496
G3	all	40000	.4868	.5430	.5500¤

¤ denote the best values for unmixing all cells in T1 based on respective training data.

It is shown that the use of mixels improves the most of classi¯cation results signi¯cantly. Further, P_{QDF} and PSSE are superior to P_{LDF}.

Fig. 5 illustrates that the water region in the test area T1 is unmixed by PSSE based on pure cells only, or pure and mixed cells of the training areas G1a and G1d. Numerals in legends give corresponding accuracy measures.

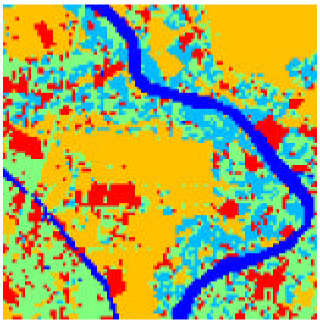
The di®erence of accuracies due to G1a and G1d is around 0.06, however, it may be seen that the ¯gures due to G1a are far superior to those due to G1d.

VI. Concluding remarks

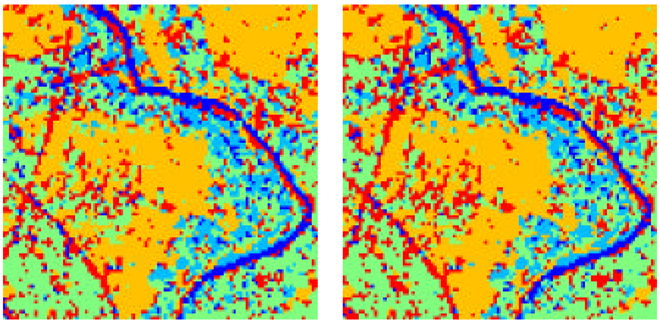
We have considered the parameter estimation method based on training data with pure and mixed cells. Next, the classi¯er based on penalized SSE are reviewed. The unmixing procedure without penalty term is very close to the ordinary linear spectral unmixing. The di®erence is just a weight matrix for evaluating the quadratic term. Based on the experiments on the actual Landsat data, we observed the followings:

² For most of discrimination procedures, training data including mixels improve the discrimination results based on pure-cells only, especially, in unmixing case.

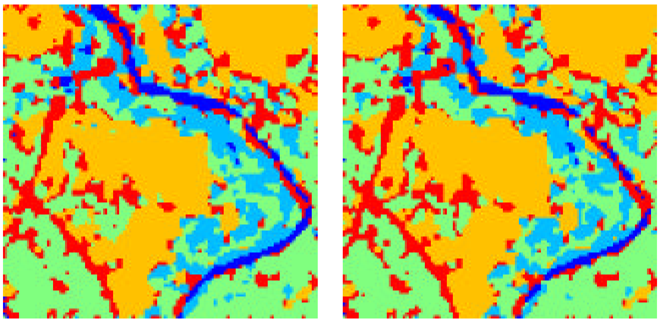
² The method PSSE de¯ned by minimizing (9) is useful for discrimination as well as unmixing. It gives spatially-stable classi¯cation results.



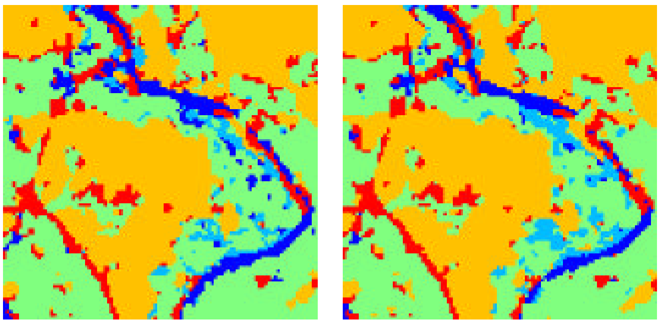
True categories of the test area T1 by the detailed digital land-use data



LDF: pure pixels (left)(6), and all pixels (right)(5)



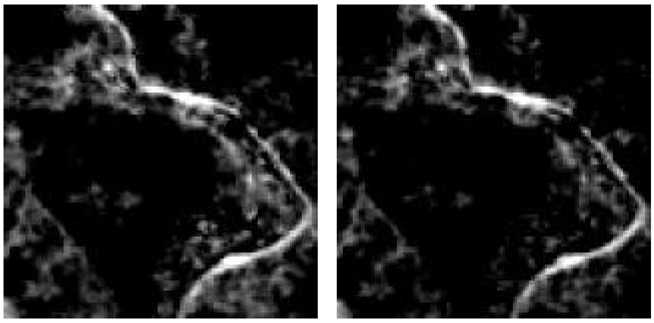
Switzer 8: pure pixels (left)(3), and all pixels (right)(4)



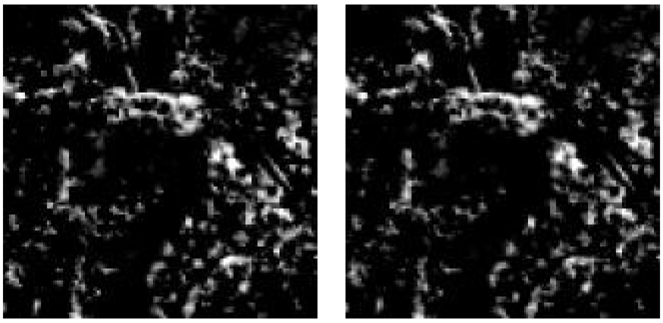
PSSE: pure pixels (left)(2), and all pixels (right)(1)

Fig. 4. Discrimination of the test area T1 due to pure pixels only or all pixels of the training area G1a. Numerals denote rank of accuracy assessment

Water region of the test area T1 by the detailed digital land-use data



G1a: pure pixels .5721 (left), all pixels .5803 (right)



G1d: pure pixels .5158 (left), all pixels .5205 (right)

Fig. 5. Unmixing of water region in T1 by PSSE due to pure pixels only or all pixels of the training areas G1a and G1d. Numerals denote the extended over-all accuracies

² In discrimination, PSSE works well when training data is small, and QDF overcomes PSSE when training data is large. In unmixing, PSSE shows a better performance than P_{LDF} and P_{QDF} de¯ned by (13) and (14) respectively.

Further investigation in the case that categories have individual variance-covariance matrices is still required.

References

[1] N. Cressie, Statistics for Spatial Data. New York: John Wiley & Sons, 1991.

[2] M. L. Eaton, Multivariate Statistics. New York: John Wiley & Sons, ch. 9, pp. 334{402, 1983.

[3] Y. H. Hu, H. B. Lee and F. L. Scarpace, \Optimal linear spectral unmixing," Transactions on Geoscience and Remote Sensing, vol. 37, no. 1, pp. 639{645, 1999.

[4] K. V. Mardia, \Spatial discrimination and classi¯cation map," Commun. Statist.{Theory Meth., vol. 13, no. 18, pp. 2181{2197, 1984.
(erratum: Commun. Statist.{Theory Meth., vol. 16, no. 12, p. 3749, 1987.)

[5] G. J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition. New York: John Wiley & Sons, ch. 13, pp. 413{446, 1992.

[6] R. Nishii, \Spatial discriminant analysis applied to multispectral satellite images," Technical Report 2000-1, Statistical Research Group, Hiroshima University, 2000.

[7] R. Nishii and S. Tanaka \Accuracy and inaccuracy assessments in land-cover classi¯cation," IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 1, pp. 491{498, 1999.

[8] P. C. Smits and S. G. Dellepiane, \Discontinuity-adaptive Markov random ¯eld model for segmentation of intensity SAR images," Transactions on Geoscience and Remote Sensing, vol. 37, no. 1, pp. 627{631, 1999.

[9] P. Switzer, \Extensions of linear discriminant analysis for statistical classi¯cation of remotely sensed satellite imagery," Mathematical Geology, vol. 12, no. 4, pp. 367{376, 1980.