International Journal of Remote Sensing
Vol. 26, No. 3, 10 February 2005, 563–577

# Using logit models to classify land cover and land-cover change from Landsat Thematic Mapper

K. C. SETO*† and R. K. KAUFMANN‡

†Department of Geological and Environmental Sciences, and Stanford Institute for International Studies, Stanford University, Encina Hall, E413, Stanford, California 94305-6055, USA

‡Center for Energy and Environmental Studies, Department of Geography, Boston University, Boston, Massachusetts 02215, USA

(*Received 3 June 2003; in final form 30 June 2004*)

In this paper, we use logit models to classify data from Landsat Thematic Mapper (TM) among 23 land-cover and land-cover change classes. The logit model is a simple statistical technique that is designed to analyse categorical data. Diagnostic statistics indicate that the logit model can classify remotely sensed data in a statistically significant fashion. User accuracies for individual land-cover classes range between 50 and 92%, with an overall accuracy of 79%. To assess these accuracies, we compare them to those generated by a Bayesian maximum likelihood classifier. While the overall accuracies are similar, the accuracies for individual land-cover categories differ. These differences may be associated with the size of the training data for each land-cover class. There is some evidence that the logit models generate higher accuracies for land-cover categories for which relatively few training pixels are available. Finally, a comparison of classification results using a 12-band composite of the six reflective TM bands and their change vectors versus a six-band composite of the three Tasselled Cap bands and their change vectors indicates that the latter reduces classification accuracies.

## 1. Introduction

Monitoring land-cover change is one of the most important applications of remote sensing. The urgent need to develop accurate assessments of the rate and scale of land-cover changes has lead to the generation of a plethora of change detection and classification techniques. These classification techniques fall into two general categories: parametric and non-parametric classifiers. Parametric classifiers assume that the data for individual classes are distributed normally. The most widely used parametric classifier is the maximum likelihood decision rule, which creates decision surfaces based on the mean and covariance of each class (Richards and Jia 1999). Non-parametric techniques make no assumptions about the statistical nature of the data. Among these, Artificial Neural Networks (ANNs) and decision trees have gained attention (Gopal and Woodcock 1996, Friedl and Brodley 1997, Seto and Liu 2003).

In this paper, we classify data from Landsat TM images with a logit model, a simple statistical technique designed to analyse categorical data. In summary, the

*Corresponding author. Email: kseto@stanford.edu

*International Journal of Remote Sensing*
ISSN 0143-1161 print/ISSN 1366-5901 online 
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160512331299270

logit model uses continuous variables to determine a binary dependent variable. For remote sensing applications, where ground surface reflectance (continuous data) is quantized into discrete spectral bands, the logit model translates spectral data on reflectance into a binary variable that indicates whether a pixel is or is not a member of a land-cover category. Strahler and Maynard (1980) first suggest the use of logit models to classify remotely sensed images, but this paper is the first attempt to classify these data with logit models. We use logit models to classify land cover and land-cover change from two Landsat Thematic Mapper (TM) images of the Pearl River Delta, China. To assess the logit models' ability to classify TM data, we calculate out-of-sample accuracies for 23 categories of land cover. These accuracies are compared with those generated by a more traditional technique, the Bayesian maximum likelihood classifier.

## 2. The logit model

Many statistical techniques have been developed to analyse categorical data (Lloyd 1999). Of these, the logit model is among the simplest. Here, we describe the logit model with special emphasis on its use to classify Landsat TM data. A more detailed description of the logit model is available in most econometric textbooks (Pindyck and Rubinfeld 1991). Much of the discussion below is adapted from Aldrich and Nelson (1984).

Our use of statistical techniques to classify Landsat TM data is based on the assumption that the probability that an individual pixel $i$ ($P_i$) is a member of a land-cover category can be described as follows:

$$P_i = \sum_{j=1}^{k} \beta_j X_{ij} \tag{1}$$

in which $X_{ij}$ is a vector of $k$ independent variables that is used to explain membership in a given land-cover category and $\beta$ is a vector of $k$ regression coefficients whose elements represent the effect of an individual variable on the probability that pixel $i$ is a member of a given land-cover category. For the purpose of analysing TM images, we hypothesize that membership in a land-cover category is determined by a pixel's spectral signature in either the six reflective bands or in the three Tasselled Cap transformed bands, Brightness (B), Greenness (G), and Wetness (W) (Crist and Cicone 1984). The Brightness band is a weighted sum of all six reflective bands and can be interpreted as the overall brightness at the surface. The Greenness band primarily measures the contrast between the visible bands and near-infrared bands and is similar to a vegetation index. The Wetness band measures the difference between the weighted sum of the visible, near-infrared bands, and mid-infrared bands, and is a useful measure of soil and plant moisture. Generally, there is a high degree of correlation among TM bands, particularly in the visible region of the spectrum. One of the main benefits of the BGW transformation is that it is a convenient form of data reduction and the BGW bands are directly associated with physical scene attributes. Therefore, the BGW bands are easily interpreted. The Tasselled Cap emphasizes the most spectrally perceptible aspects of crops and vegetation. Because of these qualities, the Tasselled Cap is a popular method of data transformation for change detection (Collins and Woodcock 1996, Seto *et al.* 2002). However, one of the assumptions that underlie the use of transformed data for classification is that the BGW bands capture most of the spectral information in the data. Therefore, we test the assumption that classification with Tasselled Cap transformed data will yield results similar to an analysis that uses all six reflective bands.

The linear relation given by equation (1) is continuous between $-\infty$ and $\infty$, but assigning a pixel membership in a land-cover category is a binary choice, $Y_i=1$ for pixels that are members of the land-cover category and $Y_i=0$ for pixels that are not members of the land-cover category. This contradiction is reconciled by transforming the dependent variable $P_i/(1-P_i)$. This division allows the dependent variable to approach infinity as the value of $P_i$ approaches 1. The lower bound, zero, is removed by taking the natural log; the value of $\log(P_i/(1-P_i))$ approaches negative infinity as the value of $P_i$ approaches zero. These transformations generate the following:

$$\log[(P_i/(1-P_i)]=\sum_{j=1}^{k}\beta_j X_{ij}=R \tag{2}$$

in which $R$ is shorthand that represents the sum of the product of regression coefficients ($\beta$) and independent variables ($X$).

Solving equation (2) for $P_i$ yields:

$$P_i=\frac{e^R}{(1+e^R)} \tag{3}$$

which is known as the logistic function. This function varies nonlinearly between 0 and 1. For very negative values of $R$, the logistic function asymptotically approaches 0. As $R$ approaches infinity, the logistic function approaches 1. This nonlinear relation between $R$ and $P_i$ is symmetric around a value of zero.

It is possible to estimate equation (2) using ordinary least squares (OLS), but OLS will generate heteroscedastic error terms, those with non-constant variance, reducing the efficiency of the estimates. Under these conditions, the estimates for the regression coefficientsthinsp $\hat{\beta}$ will not have minimum variance. Heteroscedasticity does not bias the point estimates for $\beta$, but it does obscure the interpretation of the statistical significance of the regression results.

To avoid this difficulty, the logistic model is estimated using maximum likelihood techniques. These techniques choose the elements of $\beta$ that maximize the likelihood of having observed the sample pixels' membership in a given land-cover category. The probability of this membership is given by the general likelihood function:

$$P(Y|X)=\prod_{i=1}^{N}P_i^{Y_i}(1-P_i)^{1-Y_i} \tag{4}$$

in which $P_i$ is the probability of membership in a given land-cover class ($Y_i=1$) and $(1-P_i)$ is the probability of non-membership in a given land-cover class ($Y_i=0$). Substituting the logistic equation for the probabilities yields the logit likelihood function:

$$L(Y|X,\beta)=\prod_{i=1}^{N}\left[\frac{e^R}{(1+e^R)}\right]^{Y_i}\left[\frac{1}{(1+e^R)}\right]^{1-Y_i} \tag{5}$$

This equation can be simplified by taking the log of both sides, which yields an equation with the general form:

$$\log L(Y|X, \beta) = \sum_{i=1}^{N} [Y_i \log P_i + (1 - Y_i)\log(1 - P_i)] \tag{6}$$

To identify the elements of $\beta$ that maximize the likelihood function, the first derivative with respect to each element of $\beta$ is set equal to zero. The likelihood equations for the logit model are given by:

$$\sum_{i=1}^{N}\left[Y_i - \frac{e^R}{(1+e^R)}\right]X_{ij} = 0 \quad j = 1, ..., K \tag{7}$$

Because the likelihood equations do not vary linearly with the coefficients in $R$, it is not possible to solve the equations for the elements of $\beta$. Rather, the elements of $\beta$ are identified using iterative techniques. The technique developed by Berndt *et al.* (1974) is used to estimate the results reported in section 4.

Estimation results can be evaluated using a variety of diagnostic statistics. The statistical significance of the individual regression coefficients can be evaluated by testing the null hypothesis $\beta_j = 0$ with a $t$ statistic. Values that exceed a critical threshold (e.g. $p < 0.05$) indicate that the variable $X$ has a statistically meaningful effect on the probability that a pixel is a member of an individual land-cover category. The statistical significance of the $\beta$ vector can be evaluated with a likelihood ratio statistic ($\omega$), which is distributed as a chi-square with $k-1$ degrees of freedom (the elements of $\beta$ other than the constant). The formula for the test statistic is given by:

$$\omega = -2(\log L_0 - \log L_1) \tag{8}$$

in which $L_1$ is the value of the likelihood function for the full model (as determined by estimating equation (7)) and $L_0$ is the value of the likelihood function if all the elements of $\beta$ are zero (other than the intercept). This latter is given by:

$$\log L_0 = N_0\left(\log\frac{N_0}{N}\right) + N_1\left(\log\frac{N_1}{N}\right) \tag{9}$$

in which $N_0$ is the number of pixels that are not members of a particular land-cover category ($Y_i = 0$) and $N_1$ is the number of pixels that are members of a particular land-cover category ($Y_i = 1$).

The maximum likelihood estimator chooses values of $\beta$ where the probability of membership in a given class equals 0.5 when the variables in $X$ equal the sample mean. Values of $X$ for pixel $i$ that are different from the sample mean increase/decrease the probability that pixel $i$ is a member of a given land-cover category. If the probability calculated by equation (2) is less than 0.5, pixel $i$ is classified as a non-member of the land-cover class. Conversely, if the probability is equal to or greater than 0.5, pixel $i$ is assigned membership in that land-cover category.

The logarithmic transformation of equation (1) alters the interpretation of the regression coefficients and their effect on the probability of membership relative to a linear model. The effect of a change in $X$ on the probability that a pixel is a member of a given land-cover category depends on both the regression coefficient ($\beta_k$) and the elements of $X$ associated with a given pixel, which is given as:

$$\frac{dP(Y=1)}{dX_k} = \frac{e^R}{(1+e^R)} \times \frac{1}{(1+e^R)} \times \beta_k \tag{10}$$

This implies that changes in $X$ have the greatest effect on the probability of membership at a probability of $P_i$=0.5. As the value of $X_k$ moves away from its mean, the effect of changes in $X_k$ on the probability of membership diminishes.

## 3. Data and methods

For this study, we acquire two predominantly cloud-free Landsat TM images of the Pearl River Delta in southern China (WRS path 122, row 44). Cloud-free images are not available throughout the entire calendar year because the Delta is located in the tropics, between 21° N and 23° N, which is mostly cloudy during the rainy season. As a result, the images are obtained in December 1988 and March 1996. The images are georeferenced with a Universal Transverse Mercator (UTM) map projection provided by the Institute for Remote Sensing Application in Beijing, and resampled to $30\,\mathrm{m} \times 30\,\mathrm{m}$ pixels. For both images, the root mean square error (RMSE) of the registration process is less than a third of a pixel.

Some of the difficulties that arise when comparing non-anniversary images are associated with atmospheric conditions, illumination angles, and seasonal variations. To correct for these effects, several relative atmospheric normalization techniques are available. They include methods that calibrate images to a master image based on mean radiances from stable dark and bright features (Schott *et al.* 1988, Hall *et al.* 1991). These techniques are not suitable for images of the Pearl River Delta because there is a dearth of well-defined spectrally stable dark and bright ground features. Features that are normally used as stable dark features, such as water bodies, are not suitable because of high turbidity in the Delta. Similarly, stable bright features are difficult to locate in the two images. Therefore, we use a relative radiometric normalization technique that uses a two-date density plot of DN values. Stable features in the images form a natural 'ridge' which is fit with a line. This line can be interpreted as a simple relation between the two images, and is used to match each band for each image to a reference image (Song *et al.* 2001).

Based on fieldwork in China during February 1998 and visual interpretation of the images by three analysts, we identified 809 ground data sites, 151 of which were visited in the field, with a total of 7807 pixels for training and testing that are distributed among the 23 classes (table 1). At each site, we recorded four types of

Table 1. Land-cover classes.

| Class no. | Pixels | Land-cover class | Class no. | Pixels | Land-cover class |
|---|---|---|---|---|---|
| 1 | 327 | Water | 13 | 478 | Agriculture-to-Urban |
| 2 | 319 | Forest | 14 | 761 | Agriculture-to-Transition |
| 3 | 1208 | Agriculture | 15 | 313 | Agriculture-to-Fish Pond |
| 4 | 371 | Urban | 16 | 109 | Fish Pond-to-Transition |
| 5 | 178 | Fish Pond | 17 | 238 | Transition-to-Urban |
| 6 | 227 | Transition | 18 | 362 | Shrub-to-Transition |
| 7 | 411 | Shrub | 19 | 100 | Shrub-to-Water |
| 8 | 506 | Water-to-Agriculture | 20 | 317 | Forest-to-Transition |
| 9 | 236 | Water-to-Urban | 21 | 147 | Forest-to-Water |
| 10 | 197 | Water-to-Fish Pond | 22 | 153 | Forest-to-Urban |
| 11 | 392 | Water-to-Transition | 23 | 202 | Shrub-to-Urban |
| 12 | 255 | Agriculture-to-Water | **Total** | **7807** | |

information: the primary land cover and land use (e.g. bare soil), the secondary land cover and land use (e.g. agriculture), likely misclassified class (e.g. urban), and the year when the site was converted. The latter information was obtained by interviewing land users during fieldwork and by visual interpretation of the images. The 809 sites are separated randomly five times to create five data splits. In each split, 80% of the sites are used to train the classifier while the remaining 20% are used to test the classification results. The site-based 80%/20% splits of the ground data allows us to perform a true out-of-sample analysis that can be used to evaluate the performance of the logit models and the maximum likelihood methodology (Friedl *et al.* 2000).

Seven of 23 categories represent stable land covers: Water, Forests, Shrub, Fish Pond, Agriculture, Transition, and Urban. Agriculture is a broad class that includes rice paddies, field crops, and fruit orchards. The Transition class includes tracts of land where the previous land cover has been removed, but on which there has not been extensive construction. Sites that are Transition in the 1988 image and either Urban or Transition in the 1996 image are labelled stable Urban areas. The remaining 16 categories represent changes in land cover that occur between the 1988 image and the 1996 image. To classify pixels among these categories, we estimate 23 logit models, one for each land cover and land-cover change class. In each of these models, the value of $P_i$ for pixels in that category is assigned a value of 1—all other pixels are assigned a value of 0.

Each equation uses the same vector of explanatory variables ($X$): either the DN values for the six reflective TM bands in the 1988 image and their change vectors, or the three Tasselled Cap bands and their change vectors. To calculate the change vectors, we subtract the 1988 values from the 1996 images for each band. Therefore, the input to the models is either a 12-band composite of the 1988 values and the 1996–1988 change vectors (for the six TM bands) or a six-band composite (for the BGW transformations). The effect of these variables on the probability of membership is specified using a quadratic relation as follows:

$$X = \left[\sum_{i=1}^{6}\sum_{t=1}^{2} \mathrm{Band}_{it} + (\mathrm{Band}_{it})^2\right] + \sum_{i=1}^{6} (\mathrm{Band}_{i1988} - \mathrm{Band}_{i1996})^2 \tag{11}$$

This quadratic specification allows four possible relationships between a variable and the probability of membership in an individual class: positive (increasing values of $X$ increase the probability of membership); negative (increasing values of $X$ decrease the probability of membership); U-shaped (intermediate values of $X$ minimize the probability of membership); and inverted U-shaped (intermediate values of $X$ maximize the probability of membership). Of these specifications, we expect an inverted U-shaped relation. This relation would be generated by a positive coefficient for the element of $\beta$ associated with the linear element of $X$ and a negative coefficient for the element of $\beta$ associated with the squared value of $X$.

Determining membership among 23 classes for land cover and land-cover change complicates classification relative to the binary choice of membership that is indicated by a single equation. For each pixel $i$, we use the equations to calculate the probability that it belongs to each of the 23 classes. This creates a vector of 23 probabilities for each pixel, in which each element represents the probability that pixel $i$ belongs to each of the 23 land-cover categories. Unlike the criterion for membership used by a single equation, we assign pixel $i$ to the land-cover category for which it has the highest probability of membership. We use this criterion because the probability of membership may be greater than 0.5 for more than one land-cover category or none of the probabilities of membership may exceed 0.5. Assigning a pixel to the land-cover category for which it has the highest probability of membership ensures that all pixels are classified.

## 4. Results and discussion

We estimate logit models for each of the 23 land cover and land-cover change classes in each of the five data splits (115 models). Generally, we reject the null hypothesis that the vector $\beta$ is equal to zero (table 2). The hypothesis generally is not rejected for five land-cover classes: Transition, Water-to-Fish Pond, Fish Pond-to-Transition, Transition-to-Urban, and Forest-to-Water. This result is caused largely by the failure of the non-linear iterative search technique to satisfy the convergence criterion within 1000 iterations. Despite this failing, the user accuracies for these land-cover classes are not noticeably lower than those for other land-cover classes. Furthermore, the signs associated with the elements of $\beta$ generally are consistent with an inverted U-shaped relationship between the DN value for each of the six reflective bands (or the three BGW bands) and the probability of membership in a given land-cover category. We summarize the logit models' classification results by summing the confusion matrices for the testing portion of the five data splits (table 3). The results indicate that the

Table 2. Values of $\omega$ (equation (8)) that tests the statistical significance of the vector of regression coefficients estimated using the logit models.

| Class no. | Land-cover category | Data Split 1 | Data Split 2 | Data Split 3 | Data Split 4 | Data Split 5 |
|---|---|---|---|---|---|---|
| 1 | Water | **202.2** | **225.3** | **210.6** | **106.8** | **93.2** |
| 2 | Forest | **56.1** | **106.7** | **111.1** | **128.1** | 12.3 |
| 3 | Agriculture | **690.6** | **674.0** | **714.6** | **707.1** | **702.8** |
| 4 | Urban | **53.2** | **106.1** | **119.9** | **73.6** | **122.7** |
| 5 | Fish Pond | **112.9** | **112.7** | **118.6** | **128.3** | **88.5** |
| 6 | Transition | 0.0 | 10.2 | 11.4 | 11.1 | 9.4 |
| 7 | Shrub | **437.4** | **438.7** | **434.4** | **369.7** | **301.4** |
| 8 | Water-to-Agriculture | **126.3** | **82.7** | **132.4** | **131.6** | **130.2** |
| 9 | Water-to-Urban | **112.6** | **145.8** | **47.6** | **148.1** | **138.9** |
| 10 | Water-to-Fish Pond | 19.5 | 0 | 20.6 | 14.7 | 5.4 |
| 11 | Water-to-Transition | **114.1** | **122.9** | 21.5 | **113.5** | **71.6** |
| 12 | Agriculture-to-Water | **107.7** | **117.7** | **117.0** | **115.7** | **113.4** |
| 13 | Agriculture-to-Urban | **369.2** | **419.7** | **474.8** | **461.1** | **437.2** |
| 14 | Agriculture-to-Transition | **565.0** | **531.6** | **541.5** | **568.8** | **597.5** |
| 15 | Agriculture-to-Fish Pond | **350.3** | **357.0** | **369.8** | **365.0** | **368.9** |
| 16 | Fish Pond-to-Transition | 33.4 | 36.7 | 38.0 | 39.6 | *42.1* |
| 17 | Transition-to-Urban | 2.8 | 17.4 | 14.8 | 12.5 | 16.6 |
| 18 | Shrub-to-Transition | **238.8** | **218.9** | **230.7** | **223.1** | **242.6** |
| 19 | Shrub-to-Water | **88.0** | **94.3** | **97.6** | **95.0** | **95.2** |
| 20 | Forest-to-Transition | **170.9** | **185.2** | **182.0** | **162.8** | **179.2** |
| 21 | Forest-to-Water | 13.3 | 18.3 | 22.2 | 21.4 | 19.0 |
| 22 | Forest-to-Urban | **125.8** | **132.8** | **122.2** | **131.6** | **143.0** |
| 23 | Shrub-to-Urban | **191.7** | **210.0** | **225.6** | **227.1** | **167.3** |

Test statistic is distributed as a chi-square with 30 degrees of freedom.
Values that exceeded the 0.01 threshold are given in bold; values that exceeded the 0.1 threshold are given in italics.

Table 3. Confusion matrix for classification generated by the logit models.

| | Ground data | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | User accuracy |
| **Classified** | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 273 | 1 | 0 | 0 | 17 | 0 | 0 | 1 | 0 | 2 | 0 | 3 | 0 | 2 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.89 |
| 2 | 9 | 278 | 4 | 0 | 0 | 4 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.91 |
| 3 | 7 | 1 | 1000 | 11 | 13 | 2 | 39 | 22 | 1 | 0 | 0 | 0 | 59 | 2 | 61 | 0 | 4 | 0 | 0 | 0 | 0 | 10 | 15 | 0.80 |
| 4 | 0 | 0 | 12 | 337 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 1 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 2 | 0.91 |
| 5 | 10 | 0 | 30 | 1 | 146 | 0 | 2 | 0 | 4 | 0 | 0 | 0 | 1 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.73 |
| 6 | 0 | 8 | 1 | 0 | 0 | 204 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0.92 |
| 7 | 0 | 27 | 25 | 0 | 0 | 3 | 337 | 6 | 0 | 0 | 1 | 0 | 1 | 3 | 1 | 0 | 0 | 5 | 0 | 1 | 0 | 2 | 3 | 0.81 |
| 8 | 0 | 0 | 37 | 1 | 0 | 0 | 2 | 455 | 3 | 1 | 0 | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.91 |
| 9 | 0 | 0 | 15 | 1 | 0 | 0 | 0 | 6 | 189 | 0 | 12 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0.81 |
| 10 | 1 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 184 | 0 | 0 | 0 | 0 | 17 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0.88 |
| 11 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 1 | 349 | 0 | 6 | 5 | 0 | 10 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0.89 |
| 12 | 3 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 186 | 0 | 0 | 17 | 0 | 0 | 0 | 28 | 0 | 12 | 0 | 0 | 0.73 |
| 13 | 0 | 0 | 38 | 4 | 0 | 0 | 1 | 1 | 11 | 0 | 0 | 0 | 342 | 42 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 13 | 31 | 0.70 |
| 14 | 0 | 0 | 6 | 0 | 0 | 0 | 9 | 0 | 1 | 0 | 26 | 0 | 14 | 584 | 9 | 8 | 0 | 67 | 0 | 19 | 0 | 3 | 3 | 0.78 |
| 15 | 13 | 0 | 20 | 3 | 1 | 0 | 0 | 14 | 2 | 6 | 0 | 21 | 0 | 0 | 178 | 0 | 1 | 0 | 10 | 0 | 1 | 0 | 1 | 0.66 |
| 16 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 0 | 3 | 25 | 0 | 90 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0.71 |
| 17 | 0 | 0 | 1 | 11 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 16 | 0 | 0 | 215 | 0 | 0 | 0 | 0 | 0 | 1 | 0.84 |
| 18 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 44 | 0 | 0 | 0 | 233 | 0 | 69 | 0 | 0 | 0 | 0.66 |
| 19 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 0 | 0 | 11 | 0 | 0 | 0 | 46 | 0 | 8 | 0 | 0 | 0.50 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 16 | 1 | 0 | 0 | 56 | 0 | 220 | 0 | 11 | 0 | 0.72 |
| 21 | 2 | 4 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 5 | 0 | 0 | 0 | 0 | 16 | 2 | 124 | 0 | 0 | 0.72 |
| 22 | 1 | 0 | 3 | 0 | 0 | 0 | 6 | 0 | 0 | 3 | 0 | 0 | 12 | 4 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 83 | 25 | 0.59 |
| 23 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 9 | 0 | 0 | 3 | 20 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 121 | 0.62 |
| Producer accuracy | | | | | | | | | | | | | | | | | | | | | | | | |
| | 0.83 | 0.87 | 0.83 | 0.91 | 0.82 | 0.90 | 0.82 | 0.90 | 0.80 | 0.93 | 0.89 | 0.73 | 0.72 | 0.77 | 0.57 | 0.83 | 0.90 | 0.64 | 0.46 | 0.69 | 0.84 | 0.54 | 0.60 | 0.79 |

logit models classify the out-of-sample test portion of the five datasets with a high degree of accuracy. Overall, 78.9% of the pixels are classified correctly. Accuracy varies among land-cover categories. Of the pixels that the logit models classify as Shrub-to-Water (class 19), only 50% are correctly labelled. While this is a low user's accuracy, an assessment of the confusion matrix indicates that most of the errors of commission are with the Agriculture-to-Water class (class 12). 27% of the pixels classified as Shrub-to-Water are in fact Agriculture-to-Water. Because the Agriculture class includes field crops, orchards, and rice paddies—shrub-like ground cover—it is not surprising that the logit models have difficulty differentiating between these two land-cover change classes. In contrast, the logit models' user accuracy is high (92%) for the Transition class (class 6).

The results described in the previous section beg the question, how do logit models perform relative to other methods used to classify land cover in satellite images? To answer this question, we compare the logit models' classification to the classification generated by a more conventional remote sensing approach, the Bayesian maximum likelihood classifier. The maximum likelihood classifier evaluates the statistical probability of a pixel's membership in a land-cover class based on the probability density functions of each class. As with the logit models, we classify the pixels using multivariate statistics from two inputs: either the 12-band composite of the six TM bands and their change vectors or the six-band composite of the three BGW bands and their change vectors. For each of the five data splits described in section 3, we calculate multivariate statistics for each of the land cover and land-cover change classes. The training statistics are then used to classify the test data with a Bayesian maximum likelihood classifier (Seto *et al.* 2002).

The classification results generated by the six reflective bands and their change vectors are summarized by the confusion matrix given in table 4. The results indicate that the maximum likelihood technique classifies the data with a high degree of accuracy. Overall, 79.6% of the pixels are classified correctly. As with the logit results, the accuracy varies significantly among classes. Of the pixels classified as Shrub-to-Water (class 19), only 35% are classified correctly. Similar to the logit models, the main source of misclassification is confusion with the Agriculture-to-Water class (class 12). In contrast to the low user's accuracy of this class, the Water-to-Fish Pond class has a high user's accuracy of 96% (class 10). This result is surprising considering that the Water and Fish Pond classes are both water bodies. One would expect that pixels in the Water-to-Fish Pond change class would be easily confused with either the Water or Fish Pond class.

The significance of differences in the user accuracies generated by the logit models and the maximum likelihood technique is evaluated using statistical techniques. The classification accuracy for each land-cover category in tables 3 and 4 is an average of the five datasets therefore, the statistical significance of differences in the accuracy of the logit classification relative to the maximum likelihood technique can be evaluated with a $Z$ statistic:

$$Z = \frac{L - M}{\sqrt{\frac{\sigma_L^2}{N} + \frac{\sigma_M^2}{N}}} \tag{12}$$

in which $L$ is the average accuracy of the logit classification for a given land-cover category, $M$ is the average accuracy of the maximum likelihood classifier, $\sigma^2{}_L$ is the variance of the logit classification accuracy, $\sigma^2{}_M$ is the variance of the maximum

Table 4. Confusion matrix for classification generated by the maximum likelihood estimator.

| | Ground data | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | User accuracy |
| **Classified** | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 287 | 2 | 8 | 0 | 18 | 0 | 1 | 1 | 0 | 10 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.85 |
| 2 | 9 | 285 | 27 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.86 |
| 3 | 11 | 1 | 803 | 5 | 3 | 0 | 17 | 5 | 2 | 0 | 0 | 0 | 24 | 2 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 5 | 0.90 |
| 4 | 0 | 0 | 6 | 337 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0.95 |
| 5 | 12 | 0 | 39 | 0 | 150 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.74 |
| 6 | 0 | 8 | 0 | 0 | 0 | 209 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0.94 |
| 7 | 0 | 23 | 59 | 0 | 0 | 2 | 370 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 3 | 0.79 |
| 8 | 0 | 0 | 60 | 0 | 0 | 0 | 2 | 477 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.88 |
| 9 | 0 | 0 | 8 | 0 | 3 | 0 | 0 | 7 | 190 | 0 | 24 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.81 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 184 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.96 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 0 | 340 | 0 | 0 | 3 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.91 |
| 12 | 3 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 169 | 0 | 0 | 23 | 0 | 0 | 0 | 23 | 0 | 9 | 0 | 1 | 0.71 |
| 13 | 0 | 0 | 72 | 10 | 0 | 0 | 0 | 2 | 5 | 0 | 0 | 0 | 357 | 54 | 8 | 1 | 2 | 0 | 0 | 0 | 0 | 10 | 11 | 0.67 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 11 | 0 | 24 | 0 | 21 | 585 | 10 | 7 | 0 | 19 | 0 | 23 | 0 | 7 | 2 | 0.82 |
| 15 | 4 | 0 | 81 | 3 | 4 | 0 | 0 | 14 | 0 | 3 | 0 | 3 | 18 | 0 | 226 | 0 | 1 | 0 | 7 | 0 | 1 | 2 | 1 | 0.61 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 20 | 1 | 93 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.78 |
| 17 | 0 | 0 | 2 | 15 | 0 | 16 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 228 | 0 | 3 | 0 | 0 | 0 | 0 | 0.85 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 52 | 0 | 0 | 0 | 290 | 0 | 56 | 0 | 0 | 0 | 0.71 |
| 19 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 67 | 0 | 0 | 10 | 0 | 0 | 0 | 51 | 0 | 16 | 0 | 0 | 0.35 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 9 | 0 | 0 | 0 | 53 | 0 | 228 | 0 | 11 | 1 | 0.75 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 1 | 0 | 0 | 0 | 16 | 0 | 121 | 0 | 0 | 0.81 |
| 22 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 21 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 85 | 24 | 0.47 |
| 23 | 0 | 0 | 13 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 27 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 152 | 0.64 |
| Producer accuracy | 0.88 | 0.89 | 0.66 | 0.91 | 0.84 | 0.92 | 0.90 | 0.94 | 0.81 | 0.93 | 0.87 | 0.66 | 0.75 | 0.77 | 0.72 | 0.85 | 0.96 | 0.80 | 0.51 | 0.72 | 0.82 | 0.56 | 0.75 | 0.80 |

likelihood classification accuracy, and $N$ the number of observations used to calculate the average, in this case, five. The $Z$ statistic tests the null hypothesis that $L - T = 0$ i.e. the average accuracy of the logit classification is equal to the average accuracy of the multi-date maximum likelihood technique. Values of $Z$ that exceed the critical threshold ($p<0.05$) indicate that the averages are different. The more accurate classification is identified by the sign on $Z$. A positive sign indicates that the logit classification is more accurate than the maximum likelihood technique (equation (12)). Conversely, a negative sign indicates that the maximum likelihood technique is more accurate.

The overall accuracy of the logit classification (78.9%) is slightly lower than the 79.6% accuracy generated by the maximum likelihood technique. As indicated in figure 1(*a*), this difference is not statistically significant ($Z=-0.46$, $p<0.67$). Despite the similarity of the overall accuracies, the individual accuracies vary among classes. However, the difference in accuracy is only statistically significant for the Agriculture class. For Agriculture (class 3), the average accuracy generated by maximum likelihood technique, 89.7%, is greater than the 80% accuracy generated by the logit model ($Z=-3.45$, $p<0.02$). For the remaining 22 classes of land cover, the differences in user accuracies classification generated by the logit models and maximum likelihood technique are not statistically distinguishable. The statistical significance of differences in accuracies generated from the three BGW bands are similar (figure 1(*b*)). The maximum likelihood technique generates a greater use accuracy 93.7%, for agriculture (category 3), than the logit model's 73.4% ($Z=-6.55$, $p<0.002$). But for category 15, Agriculture-to-Fish Pond, the accuracy of the logit model 71.2% is greater than the 50.9% accuracy generated by the maximum likelihood technique ($Z=-2.72$, $p<0.041$).

The differences in user accuracies among categories may be related to the number of pixels available to train the classification technique. Using either set of remotely sensed information, the maximum likelihood technique generates higher accuracies for the *agriculture* land-cover category, which has the greatest number of pixels, 1208. On the other hand, the logit models tend to generate higher accuracies (as indicated by positive values for $Z$ statistics) for land-cover categories that have relatively few pixels (figure 2(*a*),(*b*)). The negative regression coefficient for the relationship between the number of pixels in the training set for a land-cover category and the $Z$ score for the difference in accuracy between the logit model and the maximum likelihood technique is significant for the classifications generated from the six DN bands ($t=-3.21$, $p<0.002$) and the three BGW bands ($t=-3.64$, $p<0.0003$). The relationship weakens when the results for the Agriculture land-cover category are excluded from the classifications using the reflective bands ($t=-0.66$, $p<0.51$) and the BGW bands ($t=-1.44$, $p<0.16$). These results hint at the possibility that the logit model may generate more accurate classification when the size of the training dataset is small or there are relatively few pixels in each land-cover category.

While the differences in classification accuracy between the logit models and the maximum likelihood classifier are not definitive, the logit model provides additional statistical information on the magnitude and effect of each input band to a particular land cover or land-cover change class. That is, the statistical significance of the logit models can be tested, and the effect of each input band (e.g. BGW or reflective) on a land cover or land-cover change class can be interpreted through the sign and the size of the band coefficients.

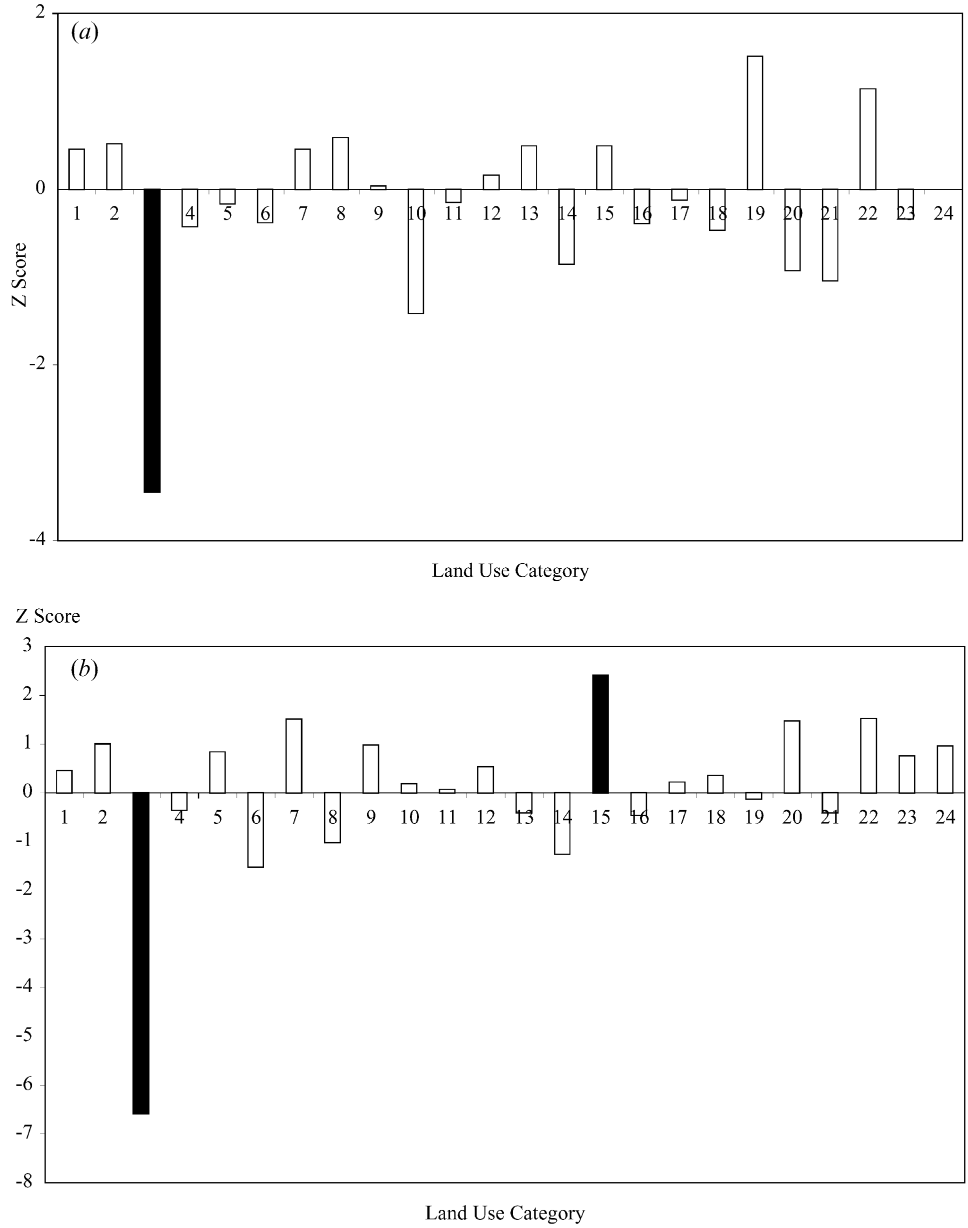


Figure 1. (*a*) *Z* values for the difference in accuracy between the logit model and maximum likelihood classifier using values for the six reflective TM bands and their change vectors. Values of *Z* that exceed the $p<0.05$ threshold are in black. (*b*) *Z* values for the difference in accuracy between the logit model and maximum likelihood classifier using values from the three BGW bands and their change vectors. Values of *Z* that exceed the $p<0.05$ threshold are in black.

Using the six reflective bands and their change vectors instead of the three BGW bands and their change vectors allows the maximum likelihood technique to generate higher accuracies. For all 23 categories, the average accuracy generated by the six reflective bands and their change vectors, 79.6%, is greater than the average

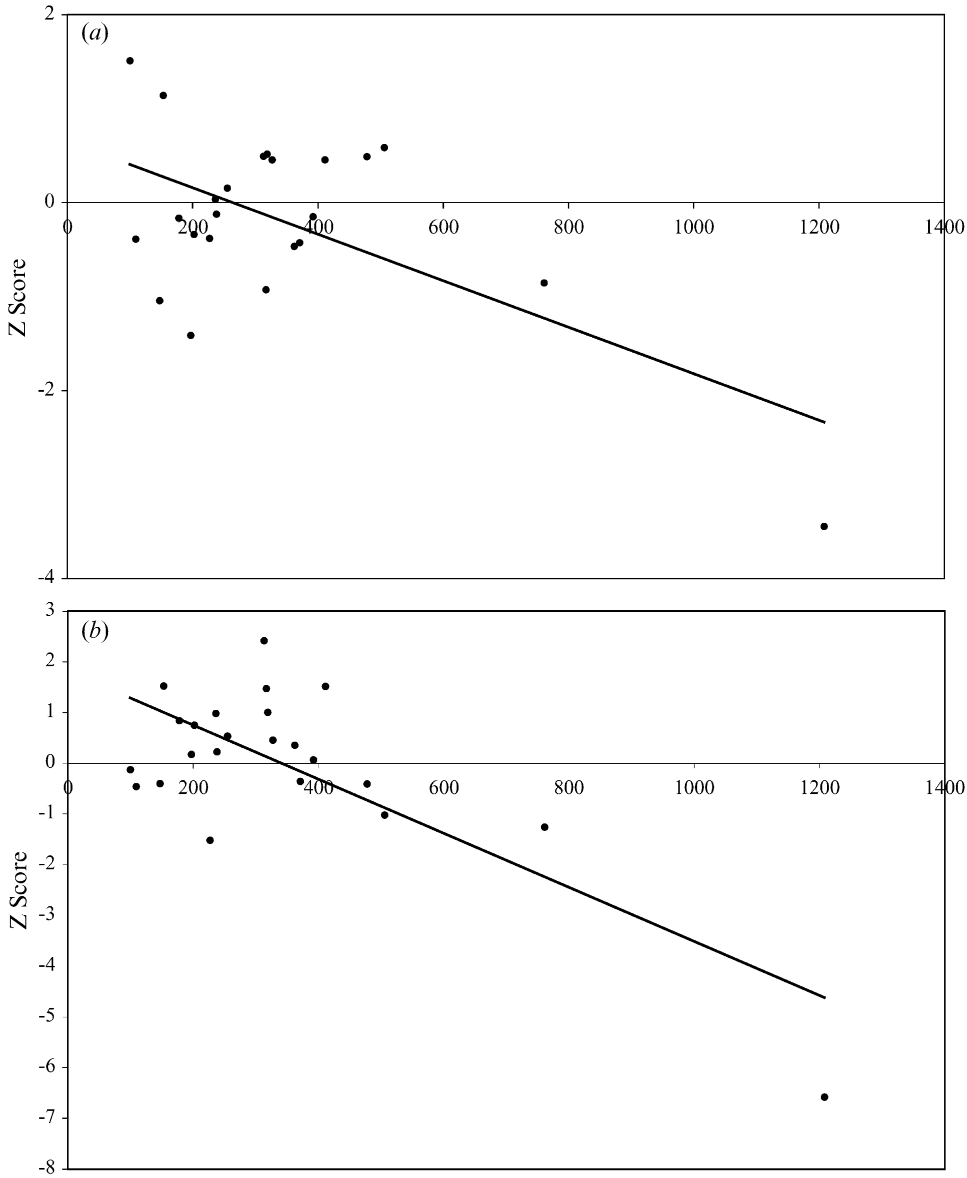


Figure 2. (*a*) The relation between the number of pixels in the training data for individual land-cover classes and $Z$ values for the difference in accuracy between the logit model and maximum likelihood classifier using values for the six reflective bands and their change vectors. The trend line is generated using OLS. (*b*) The relation between the number of pixels in the training data for individual land-cover categories and $Z$ values for the difference in accuracy between the logit model and maximum likelihood classifier using values for the three BGW bands and their change vectors. The trend line is generated using OLS.

accuracy generated from the three BGW bands and their change vectors, 73.3%. As indicated in figure 3, this difference is statistically significant ($Z$=4.28, $p$<0.006). This statistically significant difference appears in one individual category, Shrub-to-Urban ($Z$=4.4, $p$<0.009). For another 19 categories, the accuracies generated from

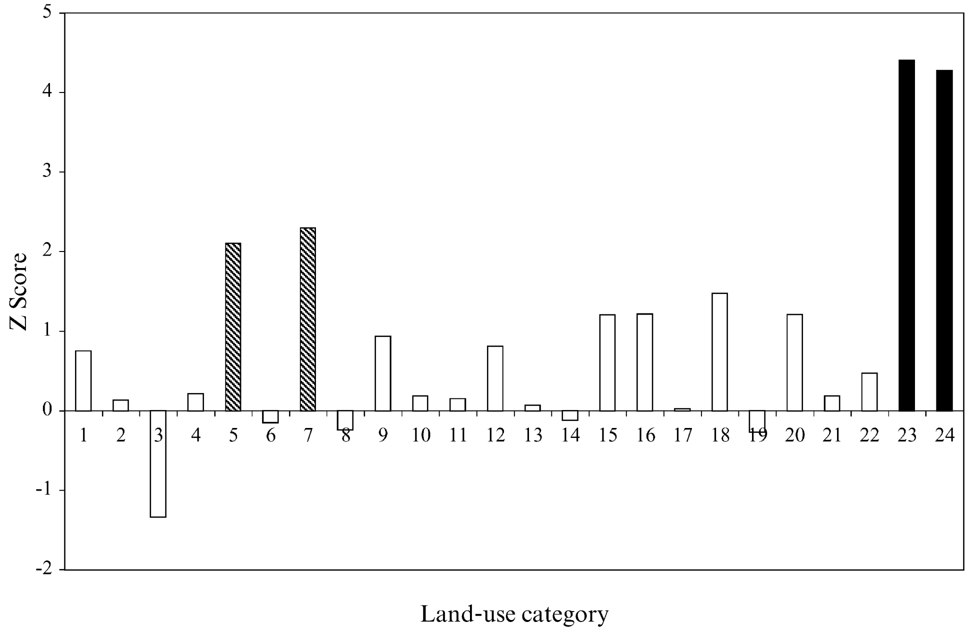


Figure 3. $Z$ values for the difference in accuracy generated by the maximum likelihood classifier using values for the six reflective bands and their change vectors, and values for the three BGW bands and their change vectors. Values of $Z$ that exceed the $p<0.05$ threshold are in black. Values of $Z$ that exceed the $p<0.10$ threshold are striped.

the six reflective bands are greater than the accuracies generated from three BGW bands, but not in a statistically significant fashion ($p<0.05$). The accuracies generated from the three BGW bands and their change vectors are greater than those generated from the six reflective bands and their change vectors for the three land-cover categories, but again none are statistically distinguishable. Together, these results indicate that using BGW bands rather than the six reflective TM bands eliminates information and reduces the accuracy of the resulting classification.

## 5. Conclusion

The results indicate that it is possible to estimate statistically meaningful logit models for membership in land-cover categories from remotely sensed images. The models can be used to classify pixels among multiple categories of land cover with a high degree of accuracy. Although the average accuracy generated by the logit models is similar to that generated by a more conventional remote sensing approach, the accuracy for individual land-cover categories may vary between techniques based on the number of pixels in the training set. The similarity of the results between the logit model and the maximum likelihood classifier indicates that while the former method works, it is not a substantial improvement to the latter. Finally, the results demonstrate that collapsing the six reflective TM bands into three BGW bands reduces the accuracy of the classification.

We recognize that these conclusions are preliminary. The ability of statistical techniques can be evaluated completely only after they are used to classify an entire image, not only training pixels selected from the scene. This additional criterion is needed because the sites included in the datasets are chosen because they are the 'purest' representation of the 23 categories of land cover. We cannot be sure how the accuracies of the two techniques will change when they are forced to classify pixels from sites that are less 'pure'.

**Acknowledgements**

This research was supported by NASA Land Cover/Land Use Change Program grant NAG5-6214 and NASA New Investigator Program grant NAG5-10 534. The authors would like to thank Alan Strahler for comments on an earlier draft.

**References**

ALDRICH, J.H. and NELSON, F.D., 1984, *Linear Probability, Logit, and Probit Models.* (Beverly Hills, California: Sage Publications).

BERNDT, E.K., HALL, B.H., HALL, R.E. and HAUSMAN, J.A., 1974, Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, **3**, pp. 653–665.

COLLINS, J.B. and WOODCOCK, C.E., 1996, An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data. *Remote Sensing of Environment*, **56**, pp. 66–77.

CRIST, E.P. and CICONE, R.C., 1984, A physically-based transformation of Thematic Mapper data—the TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, **22**, pp. 256–263.

FRIEDL, M.A. and BRODLEY, C.E., 1997, Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, **61**, pp. 399–409.

FRIEDL, M.A., WOODCOCK, C., GOPAL, S., MUCHONEY, D., STRAHLER, A.H. and BARKER SCHAAF, C., 2000, A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data. *International Journal of Remote Sensing*, **21**, pp. 1073–1077.

GOPAL, S. and WOODCOCK, C., 1996, Remote sensing of forest change using artificial neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, pp. 398–404.

HALL, F.G., STREBEL, D.E., NICKESON, J.E. and GOETZ, S.J., 1991, Radiometric rectification: toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment*, **35**, pp. 11–27.

LLOYD, C.J., 1999, *Statistical Analysis of Categorical Data*. (New York: John Wiley & Sons).

PINDYCK, R.S. and RUBINFELD, D.L., 1991, *Econometric Models and Economic Forecasts.* (New York: McGraw-Hill).

RICHARDS, J.A. and JIA, X., 1999, *Remote Sensing Digital Image Analysis: an Introduction*. (New York: Springer Verlag).

SCHOTT, J., SALVAGGIO, C. and VOLCHOK, W., 1988, Radiometric scene normalization using pseudo invariant features. *Remote Sensing of Environment*, **26**, pp. 1–16.

SETO, K.C. and LIU, W.G., 2003, Comparing ARTMAP neural network with Maximum Likelihood classifier for detecting urban change. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 981–990.

SETO, K.C., WOODCOCK, C.E., SONG, C., HUANG, X., LU, J. and KAUFMANN, R.K., 2002, Monitoring land-use change in the Pearl River Delta using Landsat TM. *International Journal of Remote Sensing*, **23**, pp. 1985–2004.

SONG, C., WOODCOCK, C.E., SETO, K.C., PAX LENNEY, M. and MACOMBER, S.A., 2001, Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? *Remote Sensing of Environment*, **75**, pp. 230–244.

STRAHLER, A.H. and MAYNARD, P.F., 1980, A logit classifier for multi-image data, *Proceedings of the Workshop on Picture Data Description and Management* (IEEE).