



## Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing

John Hogland<sup>1\*</sup>, Nedret Billor<sup>2</sup> and Nathaniel Anderson<sup>1</sup>

<sup>1</sup> Rocky Mountain Research Station, USDA Forest Service,  
200 East Broadway, Missoula, MT 59807-7669, USA

<sup>2</sup> Department of Mathematics and Statistics, Auburn University,  
364C Parker Hall, Auburn, AL 36849-5168, USA

\*Corresponding author, e-mail address: jshogland@fs.fed.us

### Abstract

Discriminant analysis, referred to as maximum likelihood classification within popular remote sensing software packages, is a common supervised technique used by analysts. Polytomous logistic regression (PLR), also referred to as multinomial logistic regression, is an alternative classification approach that is less restrictive, more flexible, and easy to interpret. To assess the utility of PLR in image classification, we compared the results of 15 classifications using independent validation datasets, estimates of kappa and error, and a non-parametric analysis of variance derived from visually interpreted observations, Landsat Enhanced Thematic Mapper plus imagery, PLR, and traditional maximum likelihood classifications algorithms.

**Keywords:** discriminant, logistic, multinomial, polytomous, probabilistic, remote sensing.

### Introduction

Remote sensing analysts use a wide array of techniques to perform supervised classifications. Generally, these techniques can be described as either data mining or statistically based procedures depending on the assumptions of the model and the way the model performs the classification. While data mining techniques such as decision trees, neural networks, and K-means clustering have become quite popular [Lu and Weng, 2007], they can over fit data and produce inferior results when compared to statistical methods such as discriminant analysis [Malhotra et. al., 1999; Worth and Cronin, 2003; Kitsantas et. al., 2006; Chen and Huang, 2009].

Discriminant analysis [Press and Wilson, 1978; Johnson and Wichern, 2007], also referred to as maximum likelihood classification in the remote sensing literature [Lillesand and Kiefer, 1994; Smith and Brown, 1997; Jensen, 2000], is a long standing classification

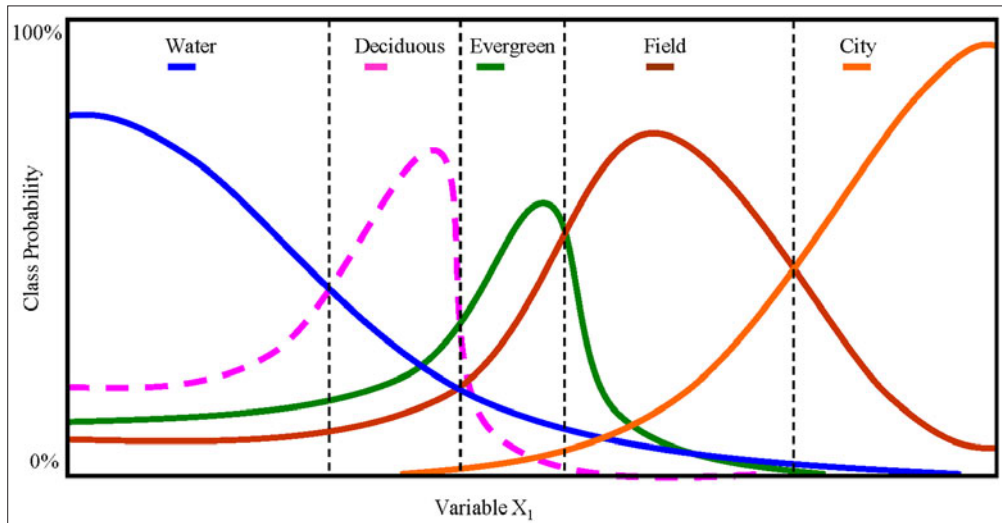
technique that distinguishes among classes by estimating multidimensional distances among classes, using either a linear or a quadratic discriminant function for a given set of explanatory variables [Smith and Brown, 1997; Johnson and Wichern, 2007; Metternicht, 2003]. From these distance measures, analysts either create a hard classification by generating rules that allocate class types to new observations based on minimizing class distance and maximizing posterior probability [Jensen, 1986; Johnson and Wichern, 2007; Sun et al., 2013], or by incorporating posterior probabilities into a fuzzy classification [Foody, 1996; Metternicht, 2003; Benz et al., 2004; Rocchini et al., 2013]. Discriminant analysis, though, assumes the data follow a multivariate normal distribution for quadratic discriminant analysis along with equal covariance for linear discriminant analysis. Because these assumptions are often difficult to satisfy [Press and Wilson, 1978; Foody, 1996], discriminant analysis tends to overestimate the magnitude of association among classes [Halpern et al., 1971; Press and Wilson, 1978; Hosmer et al., 1983; Hosmer and Lemeshow, 1989] and produces misleading posterior probabilities when assumptions are not met [Press and Wilson, 1978; Hosmer and Lemeshow, 2000; Johnson and Wichern, 2007].

To address these problems, analysts have employed a number of different techniques ranging from ignoring the statistical assumptions of discriminant analysis and performing a maximum likelihood classification [Ramírez-García et al., 1998; Keuchel et al., 2003; Galvão et al., 2005; Sun et al., 2013] to using data mining techniques that do not rely on the assumption of multivariate normality [Yoshida and Omatu, 1994; Gopal and Woodcock, 1996; Brown de Colstoun et al., 2003; Pal and Mather, 2003; Rodriguez-Galiano et al., 2012]. Classification accuracy is then assessed using a confusion matrix and comparing the proportions of times classes are correctly predicted to the number of times they are misclassified [Congalton 1991; Foody, 2002]. Assessing class accuracy using class proportions and kappa statistics [Agresti, 2002; Foody, 2002] assumes a multinomial distribution and large sample normality, respectively [Agresti, 2002], which are typically easier to satisfy.

Polytomous logistic regression (PLR), also referred to as multinomial logistic regression, is an alternative classification technique that assumes a multinomial distribution and large sample normality. This technique has been successfully used for both classification and deriving inference among response and explanatory variables in numerous fields including geography [Wrigley, 1985], engineering [Hasegawa and Kurita, 2002], biological and molecular sciences [Bailey et al., 2003], education [Peng and Nichols, 2003], environmental sciences [Mahopatra and Kant, 2005] and remote sensing [Kempen et al., 2009; Pal, 2012; Prabhakar et al., 2013]. Some of the benefits of using PLR include probabilistic outputs, less restrictive assumptions, the ability to use both continuous and categorical explanatory variables, hypothesis testing, straightforward model interpretation, and estimates of model error [Agresti, 2002]. Moreover, the probabilistic outputs of PLR are easy to interpret and can be used to generate a wide range of hard classifications by creating a series of probability threshold rules. These rules can range from identifying minimum or maximum probabilities for each class to using the classical maximum likelihood allocation rule (MLAR).

PLR hard classifications that use a MLAR, however, may not always provide good separation among classes when there is substantial overlap among explanatory variables

and can produce misleading estimates of area. For example, consider the hypothetical probability distribution of 5 class types illustrated in Figure 1.



**Figure 1 - Cross sectional view of a hypothetical, multidimensional, 5 class probability distribution. The vertical dashed lines identify the location of each class' maximum likelihood allocation rule while holding variables  $X_{1-1}$  constant and allowing variable  $X_1$  to vary.**

In this scenario using a MLAR (represented by vertical dashed lines) to classify each observation results in substantial misclassification (misclassification rate > 25%) between the classes Water and Deciduous, Deciduous and Evergreen, Evergreen and Field, and Field and City. Moreover, in this case the class area estimates derived from a MLAR can be misleading because classification errors are far from equal or symmetric for the majority of class comparisons. Using class probabilities, however, increases the ability to distinguish among class types at different ranges of the explanatory variable. In addition, accurate estimates of class area can be calculated by weighting the area of each observation by the probability of each class and then summing the weighted area estimates by class for all observations [McRoberts, 2011].

Many have made the argument that hard classifications do not always adequately describe class transitions [Foody, 1996; Metternicht, 2003; Benz et al., 2004]. To address this issue they have incorporated distance measures to assign new observations to multiple classes. However, these distance measures are inaccurate when statistical assumptions are not met [Hosmer and Lemeshow, 2000; Johnson and Wichern, 2007], potentially providing misleading results. In contrast, PLR does not have the same restrictive assumptions as discriminant analysis and directly models class probabilities, providing analysts with a more accurate estimate of class probabilities.

To illustrate the difference between discriminant analysis and PLR, it is useful to look at how each method derives estimates. In discriminant analysis, class probabilities are indirectly estimated based on class distance measurements and the assumption that

data for each class have multivariate normal distribution. Mahalanobis distances (MD), quadratic discriminant score (QDS), and posterior probabilities (PP) are calculated as follows:

$$MD = D_j(i) = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) \quad i = 1, \dots, n \quad j = 1, \dots, p \quad [1]$$

$$QDS = D_j(i) = -\frac{1}{2} \ln |S_j| - \frac{1}{2} (x_i - \bar{x}_j)' S_j^{-1} (x_i - \bar{x}_j) + \ln p_j \quad \begin{matrix} j = 1, \dots, p \\ p_j = \text{prior for class } j \end{matrix} \quad [2]$$

$$PP(j|i) = \frac{\exp\left(-\frac{1}{2} D_j(i)\right)}{\sum_{i=1}^j \exp\left(-\frac{1}{2} D_j(i)\right)} \quad [3]$$

where  $x_i$  and  $\bar{x}$  are the observed and mean vectors, respectively, and  $S$  and  $S_j$  are the pooled sampled covariance matrix and the sample covariances of class  $j$ , respectively [Johnson and Wichern, 2007].

In contrast, PLR directly estimates class probabilities and assumes that a multi-category response variable is multinomial distributed with asymptotic errors around the linear form of the natural log transformation of class odds (logits). Estimated class (response) probabilities  $\{\pi_j(x)\}$  are determined by manipulating baseline category logits as follows: given that

$$\sum_j \pi_j(x) = 1, \quad j = 1, \dots, J-1 \quad [4]$$

then

$$\pi_j(x) = 1 - \sum_{h=1}^{J-1} \pi_h(x) \quad [5]$$

Using the linear form of the logit for  $p$  covariates and a constant term, denoted by the vector,  $x$ , of length  $p + 1$

$$\ln\left(\frac{\pi_j(x)}{\pi_J(x)}\right) = x' \beta_j \quad x' = [1, x_i'] \quad i = 1, \dots, n, \quad j = 1, \dots, p \quad [6]$$

and substituting

$$1 - \sum_{h=1}^{J-1} \pi_h(x) \quad \text{for } \pi_J \quad [7]$$

$\pi_j$  and  $\pi_j$  are solved as follows:

$$\pi_j(x) = \frac{\exp(x' \beta_j)}{1 + \sum_{h=1}^{J-1} \exp(x' \beta_h)} \quad [8]$$

$$\pi_J(x) = \frac{1}{1 + \sum_{h=1}^{J-1} \exp(x' \beta_h)} \quad [9]$$

where  $\pi_j(x)$  is the mean probability of group  $j$ , with class means and variances equal to

$n\pi_j$  and  $n\pi_j(1 - \pi_j)$ , respectively [Agresti, 2002]. Maximum likelihood estimates,  $\hat{\beta}$ , are determined in an iterative fashion, typically using an optimization technique such as Newton-Raphson or Fisher scoring methods, and standard errors for each  $\hat{\beta}$  are obtained based on profile likelihood functions or using the delta method and assuming asymptotic normality [Agresti, 1990].

While the PLR has a number of benefits over the classical supervised approach, there are some drawbacks. The first drawback is that the optimization algorithms used to estimate PLR  $\hat{\beta}$  coefficients are more computationally intensive [Efron, 1975] than calculating MD or QDS when the assumption of multivariate normality holds [Bull and Donner, 1987]. Multivariate normality, however, rarely holds in remote sensing [Foody, 1996]. Secondly, PLR cannot obtain a maximum likelihood estimate when there is no overlap for classes.

While an unsolvable likelihood function for  $\hat{\beta}$  may be troubling in terms of mathematic complexity and model fit estimates [Agresti, 2002], viewed from a classification perspective this situation means that some of the class types can be separated from the rest of the class types with 100% accuracy given a set of rules. In this situation, a probabilistic classification is not required. Instead, class types can be assigned using means or Mahalanobis distance. For classes that do have overlap in the explanatory values, a maximum likelihood estimate can be obtained and a probabilistic classification can be generated.

From a theoretical standpoint, PLR is a very robust classification technique that should provide a better depiction of class distributions when compared with discriminant analysis. To test this assertion we compared PLR with discriminant analysis from the standpoint of a hard classification using a MLAR and a probabilistic classification. We then further

demonstrate the utility and flexibility of PLR using an example of 3 land cover types that share a considerable portion of spectral space.

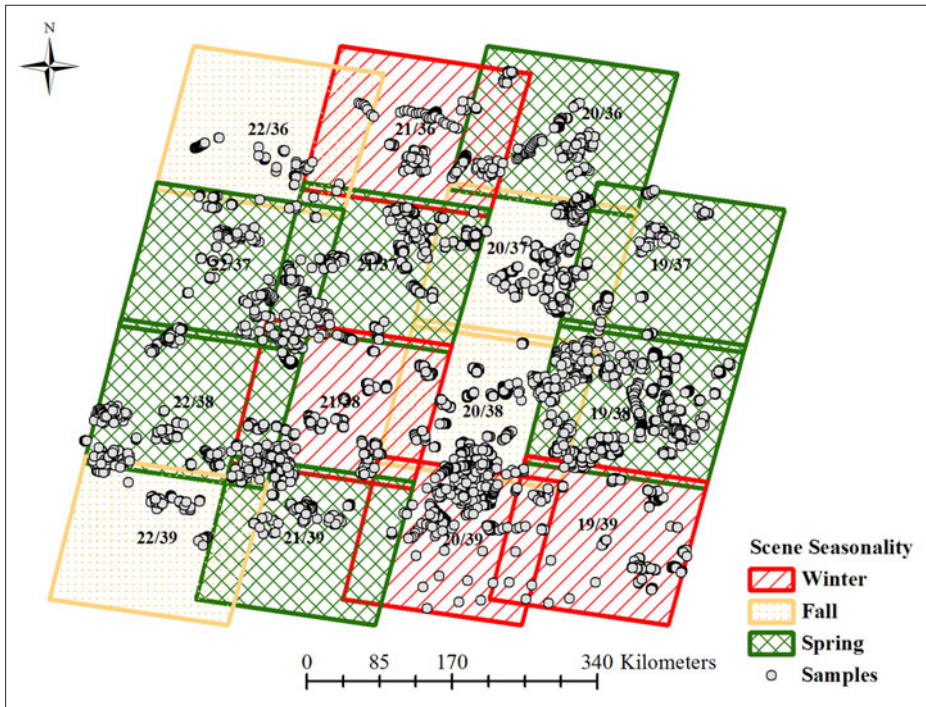
## Methods

To contrast PLR with discriminant analysis, both linear and quadratic, we classified 15 different Landsat enhanced thematic mapper plus (ETM+) scenes using a MLAR and compared measures of Kappa and error using an independent validation dataset. Landsat ETM+ scenes were preprocessed using Multi-Resource Land Characteristics Consortium data processing level 1t procedures, which radiometrically correct scenes using data generated by onboard computers during imaging events and geometrically correct scenes using ground control points and digital elevation models [NASA, 1998]. Class types consisted of generalized National Land-Cover Database categories (Tab. 1); after Homer et. al., [2004] and temporal features (i.e. clouds, burnt areas, shadows, and smoke) and were visually interpreted at the spatial scale of one ETM+ pixel using ETM+ imagery and digital orthophoto quarter quads.

To maintain consistency across the study area, one image interpreter identified all class types. Classification errors potentially caused by image acquisition dates representing different seasons were accounted for by randomly selecting one of three phenologies for each scene comparison: leaf-off winter season, leaf-on spring growing season, and leaf-on fall season (Fig. 2; Tab. 2). Digital number (DN) values occurring at the same spatial location as image interpreted samples were extracted on a nearest pixel basis, by band, using the sample command in Environmental Systems Research Institute's (ESRI) Spatial Analyst extension [ESRI, 2005].

**Table 1 - Cross walk between our stage 1 land cover classes and NLCD classes.**

| Land Cover Classes                   | NLCD Classes                    |
|--------------------------------------|---------------------------------|
| Water                                | 11 Open Water                   |
| Urban / Transportation / Bare Ground | 21 Developed, Open Space        |
|                                      | 22 Developed, Low Intensity     |
|                                      | 23 Developed, Medium Intensity  |
|                                      | 24 Developed, High Intensity    |
|                                      | 31 Barren Land (Rock/Sand/Clay) |
|                                      | 32 Unconsolidated Shore         |
| Forested *<br>Deciduous Evergreen    | 41 Deciduous Forest             |
|                                      | 42 Evergreen Forest             |
|                                      | 43 Mixed Forest                 |
|                                      | 52 Shrub/Scrub                  |
| Field<br>Wet Vegetated Area          | 71 Grassland/Herbaceous         |
|                                      | 81 Pasture/Hay                  |
|                                      | 82 Cultivated Crops             |
|                                      | 90 Woody Wetlands               |
|                                      | 95 Emergent Herbaceous Wetlands |



**Figure 2 - Seasonality of each Landsat ETM+ scene and the spatial location of each sample point (total of 15 different scenes).**

**Table 2 - Landsat ETM + scene, date, and number of classes used to compare maximum likelihood classification and PLR classifications.**

| Scene (Path/<br>Row) | Season | Date       | Number of<br>Samples | Number of<br>Classes |
|----------------------|--------|------------|----------------------|----------------------|
| 19/37                | Spring | 4/5/2000   | 251                  | 5                    |
| 19/38                | Spring | 4/5/2000   | 2155                 | 8                    |
| 19/39                | Winter | 12/20/2001 | 345                  | 5                    |
| 20/36                | Spring | 6/18/2001  | 579                  | 5                    |
| 20/37                | Fall   | 10/8/2001  | 1371                 | 7                    |
| 20/38                | Fall   | 10/8/2001  | 1421                 | 10                   |
| 20/39                | Winter | 1/25/2001  | 1005                 | 6                    |
| 21/36                | Winter | 3/5/2001   | 406                  | 4                    |
| 21/37                | Spring | 4/19/2000  | 897                  | 9                    |
| 21/38                | Winter | 2/15/2000  | 633                  | 8                    |
| 21/39                | Spring | 5/24/2001  | 318                  | 7                    |
| 22/36                | Fall   | 10/3/2000  | 88                   | 4                    |
| 22/37                | Spring | 5/15/2001  | 405                  | 8                    |
| 22/38                | Spring | 5/15/2001  | 282                  | 7                    |
| 22/39                | Fall   | 11/7/2001  | 194                  | 6                    |

Samples were then randomly partitioned into a training dataset (approximately 70% of the data for each scene) and validation dataset (approximately 30% of the data for each scene)

and imported into Statistical Analysis Software (SAS) to perform all statistical analyses. Within SAS the Discriminant and Logistic procedures were used to perform the maximum likelihood classification and PLR classifications using the training datasets. In situations where there was complete spectral separation among some of the classes, the logistic procedure was allowed to continue using the last maximum likelihood iteration to determine fit statistics. While fit statistics in these situations may be misleading, predicted class types, based on probabilities and a MLAR, can still be used to compare the class accuracy of the two methods. Total sample sizes, by class and scene, are listed in Table 3.

In several studies that have contrasted PLR and discriminant analysis [Bull and Donner, 1987; Hossain et al., 2002; McRoberts, 2011], only 3 classes were used to compare methods. In remote sensing, however, there are often many more than 3 classes. To determine whether the number of classes had an appreciable effect on the accuracy of the hard classification method, as many as 10 class types were sampled from some scenes, while other scenes had as few as 4 class types sampled. In all, 10,350 image-interpreted observations were collected across 15 different scenes.

Kappa statistics [Agresti, 1990; Smith and Brown, 1997; Foody, 2002] were estimated for each classification to assess the level of agreement between observed and predicted classes using the validation datasets. Each classification method's mean kappa estimate and corresponding lower and upper kappa 95% confidence limits were compared on a scene-by-scene basis to determine if the number of classes had an appreciable effect on the level of agreement of a particular classification method. To test general trends, we compared mean estimates of kappa for all scenes among all classification methods using a one-way nonparametric analysis of variance and the Kruskal-Wallis statistic.

**Table 3 - Number of samples per class type for each Landsat ETM+ scene. Approximately 70% of the samples were used to generate each classification model, while the remaining 30% of the data were used to test each classification methodology.**

| Scene<br>(Path<br>Row) | Burn | City | Clouds | Deciduous | Evergreen | Field | Shadow | Smoke | Water | Wetland |
|------------------------|------|------|--------|-----------|-----------|-------|--------|-------|-------|---------|
| 19/37                  | 44   | 46   | 0      | 0         | 48        | 0     | 0      | 47    | 66    | 0       |
| 19/38                  | 0    | 310  | 262    | 249       | 313       | 268   | 267    | 218   | 268   | 0       |
| 19/39                  | 59   | 0    | 0      | 0         | 71        | 73    | 0      | 0     | 71    | 71      |
| 20/36                  | 0    | 0    | 110    | 111       | 109       | 139   | 0      | 0     | 110   | 0       |
| 20/37                  | 192  | 195  | 191    | 197       | 204       | 196   | 0      | 0     | 196   | 0       |
| 20/38                  | 128  | 128  | 128    | 128       | 240       | 136   | 133    | 140   | 128   | 132     |
| 20/39                  | 170  | 157  | 0      | 169       | 172       | 0     | 0      | 0     | 168   | 169     |
| 21/36                  | 0    | 0    | 0      | 88        | 87        | 146   | 0      | 0     | 85    | 0       |
| 21/37                  | 86   | 240  | 80     | 77        | 81        | 78    | 0      | 74    | 101   | 80      |
| 21/38                  | 65   | 70   | 0      | 63        | 63        | 63    | 0      | 65    | 64    | 180     |
| 21/39                  | 44   | 44   | 0      | 50        | 45        | 44    | 0      | 0     | 47    | 44      |
| 22/36                  | 0    | 0    | 22     | 0         | 0         | 22    | 0      | 0     | 22    | 22      |
| 22/37                  | 69   | 44   | 78     | 44        | 34        | 45    | 0      | 0     | 37    | 44      |
| 22/38                  | 41   | 40   | 0      | 40        | 40        | 0     | 0      | 41    | 40    | 40      |
| 22/39                  | 33   | 0    | 29     | 41        | 33        | 0     | 29     | 0     | 0     | 29      |

After comparing the classification methods using a MLAR, the PLR procedure was performed again for a randomly chosen scene (path/row 21/37 in Fig. 2) and classes Evergreen, Deciduous, and Wetland to demonstrate how PLR can be used to test multiple

hypotheses, interpret a parsimonious probabilistic classification, and provide meaningful insights into biological processes. Training data for this scene were used to generate a suite of classification models, from which the best fitting, most parsimonious model was selected using Akaike's Information Criterion (AIC) [Akaike, 1973]. The top ranked model was then used to create a probabilistic surface for each of the three classes. To validate the probabilistic classification, we compared the summed class frequency estimates (lower and upper 95% confidence limits) against observed class frequencies using the validation dataset for that scene. Class frequency confidence intervals were estimated for each observation using variance derived from the delta method [Agresti, 1990; Agresti, 2002]. The delta method is a general method used to derive a variation based on a function, its derivative, and asymptotic normality. Under this scenario, summed lower and upper confidence limits should contain the actual number of counts associated with a confusion matrix within a specified level of accuracy (e.g., 95%), suggesting a good model fit and that the PLR model can be generalized to the rest of the population.

## Results

### Hard Classification

As they are used here, kappa statistics measure the level of agreement between the observed and predicted classes. Generally, kappa estimates among all classification methods followed similar trends. Kappa estimates for PLR and quadratic discriminant analysis were significantly larger than kappa estimates for linear discriminant analysis for scenes 19/38 and 20/38 (Fig. 3), indicating higher agreement with the observed classes for PLR and quadratic discriminant analysis.

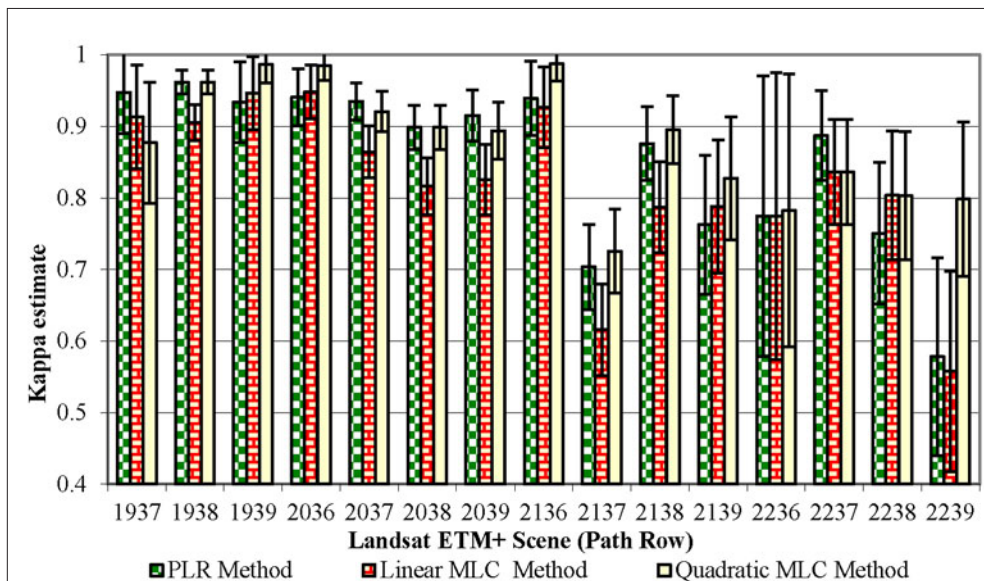


Figure 3 - Estimated mean kappa values (measure of agreement among classes) and 95% confidence intervals for each Landsat ETM+ scene (Path Row) for the PLR and MLC methods.

Additionally, PLR had a significantly larger kappa value when compared to the linear discriminant analysis for scenes 20/37 and 20/39. There were no significant differences in mean kappa estimates, however, between PLR and quadratic discriminant analysis. These results suggest that the number of classes and the sample size of each comparison did not have an appreciable effect on the accuracy of any particular classification method (Tab. 2; Fig. 3). In addition, there were no significant differences in hard classification accuracy trends among the different classification methods (all method comparisons: Kruskal-Wallis  $X^2_{df=2} = 1.6935$ ; p-value = 0.43, PLR vs linear maximum likelihood classification: Kruskal-Wallis  $X^2_{df=1} = 0.7230$ ; p-value = 0.40, PLR vs quadratic maximum likelihood classification Kruskal-Wallis  $X^2_{df=1} = 0.1897$ ; p-value = 0.66).

### ***Probabilistic Classification***

Applying the PLR classification method for the classes Evergreen, Deciduous, and Wetland in scene 21/37, we found that only 6 spectral bands were needed to sufficiently describe the probability transition among classes (Tab. 4).

**Table 4 - Model fit statistics for the PLR classification method using the Wetland class as the baseline response variable. Model one represents our top model.**

| Model # | Landsat ETM+ Bands Used in the Model | DF | AIC    | $\Delta$ AIC | AIC Model Weight | X <sup>2</sup> Nested Comparison | P-value            |
|---------|--------------------------------------|----|--------|--------------|------------------|----------------------------------|--------------------|
| 1       | bands 2-7                            | 12 | 182.28 | 0.000        | 0.500            | 2 vs 1                           | 0.14266            |
| 2       | bands 1-7*                           | 14 | 182.38 | 0.11         | 0.474            | NA                               | NA                 |
| 3       | bands 2, 4-7                         | 10 | 188.20 | 5.92         | 0.026            | 3 vs 2<br>3 vs 1                 | 0.07921<br>0.07048 |
| 4       | bands 1-6                            | 12 | 204.66 | 22.38        | 0.000            | 4 vs 2                           | 0.00000            |
| 5       | bands 2-6                            | 10 | 207.00 | 24.72        | 0.000            | 5 vs 2                           | 0.00000            |
| 6       | bands 3-7                            | 10 | 219.88 | 37.60        | 0.000            | 6 vs 2                           | 0.00000            |

Using Landsat ETM+ bands 2 through 7, we found a statistically significant relationship between our classes and the 6 spectral bands ( $X^2_{df, 12} = 227.59$ ; p-value < 0.0001) that explained the majority of information within our training data (maximum rescaled  $\hat{R}^2 = 0.8211$ ). Interpreting the  $\hat{\beta}$  estimates, in terms of odds ratios, we identified the effect each Landsat ETM+ band had on class probabilities (Tab. 5).

**Table 5 - Maximum likelihood  $\hat{\beta}$  estimates, standard errors (STE), chi-square values, p-values, and odds ratio for the top ranked probabilistic classification model. Band 3 for the deciduous and evergreen classes and band 6 for the evergreen class have p-values > 0.05. While these variables are not significant at  $\alpha = 0.05$  for their respective class, they are significant for the other class, thus making them significant in the overall model.**

| Variable  | Class     | DF | $\hat{\beta}$ | STE   | X <sup>2</sup> | P-value | Odds ratio<br>(Class /<br>Wetland) | Odds ratio<br>(Deciduous /<br>Evergreen) |
|-----------|-----------|----|---------------|-------|----------------|---------|------------------------------------|------------------------------------------|
| Intercept | deciduous | 1  | 263.20        | 59.22 | 19.74          | <0.0001 | Exp(263.2) *                       | Exp(196.75)<br>*                         |
| Intercept | evergreen | 1  | 66.45         | 23.31 | 8.12           | 0.0044  | Exp(66.45) *                       |                                          |
| Band 2    | deciduous | 1  | -1.24         | 0.35  | 12.42          | 0.0004  | 0.289 *                            | 0.820                                    |
| Band 2    | evergreen | 1  | -1.04         | 0.22  | 23.04          | <0.0001 | 0.352 *                            |                                          |
| Band 3    | deciduous | 1  | -0.51         | 0.28  | 3.30           | 0.0693  | 0.601                              | 0.483 *                                  |
| Band 3    | evergreen | 1  | 0.22          | 0.13  | 2.73           | 0.0987  | 1.243                              |                                          |
| Band 4    | deciduous | 1  | 0.38          | 0.11  | 12.48          | 0.0004  | 1.455 *                            | 1.211                                    |
| Band 4    | evergreen | 1  | 0.18          | 0.06  | 9.62           | 0.0019  | 1.202 *                            |                                          |
| Band 5    | deciduous | 1  | -0.36         | 0.16  | 4.98           | 0.0256  | 0.7 *                              | 1.070                                    |
| Band 5    | evergreen | 1  | -0.42         | 0.11  | 15.87          | <0.0001 | 0.654 *                            |                                          |
| Band 6    | deciduous | 1  | -1.83         | 0.47  | 14.85          | 0.0001  | 0.161 *                            | 0.215 *                                  |
| Band 6    | evergreen | 1  | -0.29         | 0.18  | 2.61           | 0.1062  | 0.748                              |                                          |
| Band 7    | deciduous | 1  | 1.16          | 0.33  | 12.60          | 0.0004  | 3.201 *                            | 1.401                                    |
| Band 7    | evergreen | 1  | 0.83          | 0.20  | 16.35          | <0.0001 | 2.284 *                            |                                          |

For example, with every incremental increase in DN values for band 4, while holding bands 2, 3, 5, 6, and 7 constant, the odds (chances) of the classes Deciduous and Evergreen increased by multiples of 1.455 and 1.202, respectively, when compared with the Wetland Class. Comparing the Deciduous and the Evergreen classes, however, was less straightforward. The odds ratio between the Deciduous and Evergreen classes was calculated as follows: Given the log odds of the Deciduous and Evergreen classes,

$$\ln\left(\frac{\pi_{Deciduous}}{\pi_{Wetland}}\right), \quad \text{and} \quad \ln\left(\frac{\pi_{Evergreen}}{\pi_{Wetland}}\right) = x' \beta_j \quad [10]$$

and that,

$$\ln\left(\frac{\pi_{Deciduous}}{\pi_{Evergreen}}\right) = \ln\left(\frac{\pi_{Deciduous}}{\pi_{Wetland}}\right) - \ln\left(\frac{\pi_{Evergreen}}{\pi_{Wetland}}\right) = (x' \beta_{Deciduous}) - (x' \beta_{Evergreen}) \quad [11]$$

Taking the exponent of equation [11] [Agresti, 2002], we estimated average odds of the class Deciduous increased by a multiple of 1.211 for every incremental increase in band 4 DN values (Tab. 5).

Using the validation dataset and our PLR model we predicted the total number of

observations for each class within the bounds of a 95% confidence interval (Tab. 6; row and column totals).

**Table 6 - Comparison between an observed hard classification accuracy assessment and a predicted hard classification accuracy assessment using the top ranked PLR probabilistic classification model and a MLAR (A = Lower Predicted (95%CI), B = Observed (predicted), C = Upper predicted (95%CI)).**

|                           |           | Predicted Classes |                  |       |           |                  |       |         |                  |       |        |                  |       |
|---------------------------|-----------|-------------------|------------------|-------|-----------|------------------|-------|---------|------------------|-------|--------|------------------|-------|
| Counts                    |           | Deciduous         |                  |       | Evergreen |                  |       | Wetland |                  |       | Totals |                  |       |
| R o w<br>Proportion       |           | A                 | B                | C     | A         | B                | C     | A       | B                | C     | A      | B                | C     |
| C o l u m n<br>Proportion |           |                   |                  |       |           |                  |       |         |                  |       |        |                  |       |
| Observed Classes          | Deciduous | 15.49             | 18.00<br>(18.48) | 20.98 | 0.17      | 0.00<br>(2.03)   | 4.43  | 0.00    | 4.00<br>(0.91)   | 2.32  | 15.66  | 22.00<br>(21.42) | 27.73 |
|                           |           | 0.72              | 0.82<br>(0.86)   | 0.98  | 0.01      | 0.00<br>(0.09)   | 0.21  | 0.00    | 0.18<br>(0.04)   | 0.11  | 0.73   | 1.00             | 1.30  |
|                           |           | 0.74              | 0.86<br>(0.88)   | 1.00  | 0.01      | 0.00<br>(0.11)   | 0.23  | 0.00    | 0.17<br>(0.04)   | 0.10  |        |                  |       |
|                           | Evergreen | 0.00              | 2.00<br>(1.56)   | 3.72  | 11.36     | 18.00<br>(14.60) | 17.42 | 0.57    | 4.00<br>(3.08)   | 5.81  | 11.93  | 24.00<br>(19.24) | 26.95 |
|                           |           | 0.00              | 0.08<br>(0.08)   | 0.19  | 0.59      | 0.75<br>(0.76)   | 0.91  | 0.03    | 0.17<br>(0.16)   | 0.30  | 0.62   | 1.00             | 1.40  |
|                           |           | 0.00              | 0.10<br>(0.07)   | 0.18  | 0.60      | 0.95<br>(0.77)   | 0.92  | 0.02    | 0.17<br>(0.13)   | 0.24  |        |                  |       |
|                           | Wetland   | 0.00              | 1.00<br>(0.96)   | 2.36  | 0.52      | 1.00<br>(2.37)   | 4.55  | 16.43   | 16.00<br>(20.01) | 23.08 | 16.95  | 18.00            | 29.99 |
|                           |           | 0.00              | 0.06<br>(0.04)   | 0.10  | 0.02      | 0.06<br>(0.10)   | 0.19  | 0.70    | 0.89<br>(0.86)   | 0.99  | 0.72   | (23.34)          | 1.28  |
|                           |           | 0.00              | 0.05<br>(0.05)   | 0.11  | 0.03      | 0.05<br>(0.12)   | 0.24  | 0.68    | 0.67<br>(0.83)   | 0.96  |        |                  |       |
|                           | Totals    | 15.49             | 21.00            | 27.06 | 12.05     | 19.00            | 26.4  | 17      | 24.00            | 31.21 | 44.54  | 64.00            | 84.67 |
|                           |           | 0.74              | 1.00             | 1.29  | 0.64      | 1.00             | 1.39  | 0.70    | 1.00             | 1.30  |        |                  |       |

Moreover, using the class probabilities, estimates of model error, a 95% confidence interval, and a MLAR for each observation, we predicted the cell counts of an independent accuracy assessment.

## Discussion

Our study focuses on comparing PLR with linear and quadratic discriminant analyses in a remote sensing context. While we did not specifically look at data mining techniques, discriminant analysis and PLR have been shown to perform as well if not better than those methods, especially when assessed against an independent validation dataset [Malhotra et al., 1999; Worth and Cronin, 2003; Kitsantas et al., 2006; Chen and Huang, 2009]. When comparing hard classification accuracies among PLR, linear discriminant analysis, and quadratic discriminant analysis we found no difference among the techniques. Similar results have been documented in the binary case [Press and Wilson, 1978] and multinomial case [Hossain et al., 2002; McRoberts, 2011], suggesting that a MLAR derived from MD, QDS, or PLR probabilities tend to find similar regions of overlap in explanatory variables. Under different circumstances, such as when categorical variables provide important class separating information, this would not necessarily be the case and PLR would have a significant advantage over the linear or quadratic discriminant analysis [Hosmer et al., 1983].

More importantly, PLR differs from discriminant analysis in its modeling assumptions, the ability to incorporate categorical and continuous variables, the derivation of estimates, the focus on directly modeling class probabilities, and the ability to estimate model error (i.e., error in the estimated probabilities). In essence, a PLR classification is not limited to one MLAR raster dataset. Instead, a series of probabilistic raster datasets, with corresponding lower and upper confidence limits can be produced.

For example, each observation in scene 21/37 represents a Landsat ETM+ pixel. Class probabilities for each pixel in that scene are interpreted as the mean proportion of times that one would expect to find each class at a pixel with given spectral values. Assume there were 100 pixels with DN values of 47, 36, 65, 69, 124, and 33 for bands 2 through

7, respectively. Using equations [8] and [9], the delta method, and  $\hat{\beta}$  estimates in Table 5, we would expect on average 28 (12-45, 95% confidence interval), 7 (0-17, 95% confidence interval), and 65 (47-83, 95% confidence interval) pixels out of the original 100 pixels to be Evergreen, Deciduous, and Wetland, respectively. These types of datasets not only depict the probability of each class for each pixel, but also maintain the errors associated with class probabilities, which can be directly incorporated into other predictive models.

Multiple techniques exist to validate PLR models [Neter et al., 1996]. For example, a classical model validation approach is to build models for both training and validation

datasets using the same variables and compare estimates of  $\hat{\beta}$  between models to see if they are statistically different from one another. The difficulty with this approach is that it requires a substantial amount of data. We demonstrated how conditional probabilities and probability confidence intervals can be used to assess the accuracy of a PLR model when sample size is limited (Tab. 6). One can then infer whether PLR probabilities can be generalized to the rest of the population (e.g., the rest of the pixels inside a Landsat ETM+ image) by comparing these predicted estimates with observed independent data.

Evaluating PLR fit statistics and taking into consideration model parsimony, Landsat ETM+ bands 2 through 7 were useful in separating the classes Evergreen, Deciduous, and Wetland in scene 21/37 (Tab. 4). ETM+ band 1 did not provide any additional information and was subsequently removed from the analysis. Interpreting the results for scene 21/37 in terms of scaled energy (i.e., DN values) and odds ratios, we see that the odds of the Evergreen class increase as the amounts of energy being reflected and transmitted from the earth's surface in ETM+ bands 3 and 6 increase while the energy being reflected and transmitted in band 2 is maintained at low levels (Tab. 5). Conversely, as the amounts of energy being reflected and transmitted in bands 3 and 6 decrease while maintaining band 2 at low levels, the odds of the Deciduous class increase.

These results have biological origins. For example, given the date of image acquisition and the fact that different plants often use different portions of the electromagnetic spectrum, these results may suggest that in the spring, Deciduous class types absorb a relatively large portion of visible light (bands 2 and 3) when performing photosynthesis, and that these class types maintain cooler daily temperatures (band 6) when compared with Evergreen and Wetland classes. Using PLR in this manner, analyst can generate and test a wide range of hypotheses in addition to creating a classification.

Whether remote sensing analysts are primarily focused on developing a hard classification, building a series of probability maps, or understanding the relative importance of variables

that drive a classification, it is critical that modeling assumptions are checked. Failing to do so may result in inaccurate classification models. If the assumptions of a classification method are met, such as in the PLR example for scene 21/37, then estimates of class probabilities can be used as an alternative to a hard classification. Moreover, these estimates of class probabilities (and error) can provide a more accurate depiction of class area (and area confidence intervals). This is not to imply that every pixel has a certain proportion of its area allocated to each class. Instead, it indicates that, on average, we expect a certain number of pixels to be allocated to each class type, given a set of explanatory values. This means that as the number of pixels used to estimate class area for a specified region increases, the accuracy of our class area estimate should also increase.

Identifying where classes are located across a landscape, however, is less straightforward. Given that class probabilities do not represent the amount of class area within each pixel, a class allocation rule must be specified to identify class location. Arguably, one of the easiest rules to apply is a MLAR. While this type of rule will ensure that every pixel is classified into one of the predefined categories, in some instances this rule may allocate class types to pixels where the probability of being a specific class could be as low as 100% divided by the number of classes. Alternatively, with PLR analysts can select class probability thresholds that identify class locations. While this approach may leave some pixels unclassified, it provides users of the dataset with a level of precision (in terms of probabilities) for each class location. By maintaining class probabilities for each pixel, dataset users have the flexibility to address numerous scenarios. For example, if a user is interested in identifying locations in scene 21/37 where there is a high probability of the class Evergreen ( $x > 66\%$ ) and medium to high probability of the class Wetland ( $33\% < x < 66\%$ ), they could easily perform such a query within a GIS.

## Conclusion

The PLR technique has desirable qualities for classifying remotely sensed data including relatively unrestrictive model assumptions, the ability to incorporate both continuous and categorical variables directly into the classification scheme, relatively easy techniques for model comparison (e.g., AIC), an intuitive relationship between class types and explanatory variables (odds ratios), and a focus on directly modeling class probabilities. Applying a PLR model to raster surfaces of explanatory variables creates a series of raster surfaces that depict class probability distributions. Using class probability distributions and estimates of error, users can accurately estimate the amount of area for each class type within a predefined level of confidence. In addition, they can query each class probability distribution to generate a wide variety of maps identifying class locations that are tailored to specific questions. These maps can include a MLAR map, but are not limited to this type of map. Overall, the PLR classification method provides an extremely flexible alternative to the classical supervised approach.

## Acknowledgements

Funding for this work was provided by Alabama Gap Analysis Program, the Rocky Mountain Research Station, and the Biomass Research and Development Initiative of the USDA National Institute of Food and Agriculture. Special thanks to Mark MacKenzie,

Melissa Reynolds-Hogland, and the reviewers for their help and constructive comments.

## References

- Agresti A. (1990) - *Categorical data analysis*. Wiley-Interscience, New York, pp. 558.
- Agresti A. (2002) - *Categorical data analysis*. Wiley-Interscience, Hoboken, New Jersey, pp. 710. doi: <http://dx.doi.org/10.1002/0471249688>.
- Akaike H. (1973) - *Information theory and an extension of the maximum likelihood principle*. Proceedings of the second international Symposium on information Theory, Aakademiai Kidao Budapest.
- Bailey N., Clemets T., Lee J.T., Thompson S. (2003) - *Modeling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach*, Journal of Environmental Management, 67: 395-407. doi: [http://dx.doi.org/10.1016/S0301-4797\(02\)00227-X](http://dx.doi.org/10.1016/S0301-4797(02)00227-X).
- Benz U.C., Hofmann P., Willhauck G., Lingenfelder I., Heynen M. (2004) - *Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information*. ISPRS Journal of Photogrammetry and Remote Sensing, 58: 239-258. doi: <http://dx.doi.org/10.1016/j.isprsjprs.2003.10.002>.
- Brown de Colstoun E.C., Story M.H., Thompson C., Commisso K., Smith T.G., Irons J.R. (2003) - *National park vegetation mapping using multitemporal landsat 7 data and a decision tree classifie*. Remote Sensing of Environment, 85: 316-327. doi: [http://dx.doi.org/10.1016/S0034-4257\(03\)00010-5](http://dx.doi.org/10.1016/S0034-4257(03)00010-5).
- Bull S. B., Donner A. (1987) - *The efficiency of multinomial logistic regression compared with multiple group discriminant analysis*. Journal of the American Statistical Association, 82 (400): 1118-1122. doi: <http://dx.doi.org/10.2307/2289389>.
- Chen L., Huang S. (2009) - *Accuracy and efficiency comparison of single-and multi-cycled software classification models*. Information and Software Technology, 51: 173-181. doi: <http://dx.doi.org/10.1016/j.infsof.2008.03.004>.
- Congalton R. (1991) - *A review of assessing the accuracy of classifications of remotely sensed data*. Remote Sensing of Environment, 37: 35-46. doi: [http://dx.doi.org/10.1016/0034-4257\(91\)90048-B](http://dx.doi.org/10.1016/0034-4257(91)90048-B).
- Efron B. (1975) - *The efficiency of logistic regression compared to normal discriminant analysis*. Journal of the American Statistical Association, 70 (352): 892-898. doi: <http://dx.doi.org/10.2307/2285453>.
- ESRI - Environmental Systems Resource Institute (2005) - *ArcGIS Desktop Help*. <http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?TopicName=welcome> (last accessed 7/6/2005).
- Foody G.M. (1996) - *Fuzzy modeling of vegetation from remotely sensed imagery*. Ecological Modeling, 85: 3-12.
- Foody G.M. (2002) - *Status of land cover classification accuracy assessment*. Remote Sensing of Environment, 80: 185-201. doi: [http://dx.doi.org/10.1016/S0034-4257\(01\)00295-4](http://dx.doi.org/10.1016/S0034-4257(01)00295-4).
- Galvão L.S., Formaggio A.R., Tistot D.A. (2005) - *Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data*. Remote Sensing of Environment, 94: 523-534. doi: <http://dx.doi.org/10.1016/j.rse.2004.11.012>.
- Gopal S., Woodcock C. (1996) - *Remote sensing of forest change using artificial neural networks*. IEEE Transactions on Geoscience and Remote Sensing, 34 (2): 398-404. doi: <http://dx.doi.org/10.1109/36.485117>.

- Halpern M., Blackwelder W.C., Verter J.I. (1971) - *Estimation of the multivariate logistic risk function: a comparison of the discriminant function and maximum likelihood approaches*. Journal of Chronic Disease, 24: 125-158. doi: [http://dx.doi.org/10.1016/0021-9681\(71\)90106-8](http://dx.doi.org/10.1016/0021-9681(71)90106-8).
- Hasegawa O., Kurita T. (2002) - *Face and non-face classification by multinomial logit models and kernel feature compound vectors*. Proceedings of the 9th International Conference on Neural Information Processing, 2: 996-1000. doi: <http://dx.doi.org/10.1109/ICONIP.2002.1198210>.
- Homer C., Huang C., Yang L., Wylie B., Coan M. (2004) - *Development of a 2001 National Land Cover Database for the United States*. Photogrammetric Engineering and Remote Sensing, 70 (7): 829-840.
- Hosmer T.D., Hosmer W., Fisher L.L. (1983) - *A comparison of the maximum likelihood and discriminant function estimators of the coefficients of the logistic regression model for mixed continuous and discrete variables*. Communications in Statistics, 12: 23-24. doi: <http://dx.doi.org/10.1080/03610918308812298>.
- Hosmer D.W., Lemeshow S. (1989) - *Applied logistic regression*. Wiley-Interscience New York, pp. 307.
- Hosmer D.W., Lemeshow S. (2000) - *Applied logistic regression, second edition*. Wiley-Interscience New York, pp. 375. doi: <http://dx.doi.org/10.1002/0471722146>.
- Hossain M., Wright S., Petersen L.A. (2002) - *Comparing performance of multinomial logistic regression and discriminant analysis for monitoring access to care for acute myocardial infarction*. Journal of Clinical Epidemiology, 55: 400-406. doi: [http://dx.doi.org/10.1016/S0895-4356\(01\)00505-4](http://dx.doi.org/10.1016/S0895-4356(01)00505-4).
- Jensen J.R. (1986) - *Introductory digital image processing: a remote sensing perspective*. Prentice-Hall, Englewood Cliffs, N.J., pp. 378. doi: <http://dx.doi.org/10.1080/10106048709354084>.
- Jensen J.R. (2000) - *Remote sensing of the environment: an earth resource perspective*. Prentice Hall, 2nd edition, Upper Saddle River, N.J., pp. 544.
- Johnson R.A., Wichern D.W. (2007) - *Applied multivariate statistical data analysis*. Prentice Hall, sixth edition, Upper Saddle River, N.J. pp. 773.
- Kempen B., Brus D., Heuvelink G., Stoorvogel J. (2009) - *Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach*. Geoderma, 151 (3): 311-326. doi: <http://dx.doi.org/10.1016/j.geoderma.2009.04.023>.
- Keuchel J., Naumann S., Heiler M., Siegmund A. (2003) - *Automatic land cover analysis for Tenerife by supervised classification using remotely sensed data*. Remote Sensing of Environment, 86: 530-541. doi: [http://dx.doi.org/10.1016/S0034-4257\(03\)00130-5](http://dx.doi.org/10.1016/S0034-4257(03)00130-5).
- Kitsantas P., Holander M., Lei L. (2006) - *Using classification trees to assess low birth weight outcomes*. Artificial Intelligence in MEDICINE, 38 (3): 275-269. doi: <http://dx.doi.org/10.1016/j.artmed.2006.03.008>.
- Lillesand T.M., Kiefer R.W. (1994) - *Remote sensing and image interpretation*. Wiley and Sons; third edition, New York, pp. 750.
- Lu D., Weng Q. (2007) - *A survey of image classification methods and techniques for improving classification performance*. International Journal of Remote Sensing, 28 (5): 823-870. doi: <http://dx.doi.org/10.1080/01431160600746456>.

- Mahopatra K., Kant S. (2005) - *Tropical deforestation: a multinomial logistic model and some country-specific policy prescriptions*. Forest Policy and Economics, 7: 1-24. doi: [http://dx.doi.org/10.1016/S1389-9341\(03\)00064-9](http://dx.doi.org/10.1016/S1389-9341(03)00064-9).
- Malhotra M.K., Sharma S., Nair S.S. (1999) - *Decision making using multiple models*. European Journal of Operational Research, 114: 1-14. doi: [http://dx.doi.org/10.1016/S0377-2217\(98\)00037-X](http://dx.doi.org/10.1016/S0377-2217(98)00037-X).
- McRoberts R.E. (2011) - *Satellite image-based maps: Scientific inference or pretty pictures?* Remote Sensing of Environment, 115 (2011): 715-724. doi: <http://dx.doi.org/10.1016/j.rse.2010.10.013>.
- Metternicht G.I. (2003) - *Categorical fuzziness: a comparison between crisp and fuzzy class boundary modeling for mapping salt-affected soils using Landsat TM data and a classification based on anion ratios*. Ecological Modeling, 168: 371-389. doi: [http://dx.doi.org/10.1016/S0304-3800\(03\)00147-9](http://dx.doi.org/10.1016/S0304-3800(03)00147-9).
- NASA - National Aeronautics and Space Administration (1998) - *Landsat 7 Science Data User's Handbook*, Available at: [http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook\\_toc.html](http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_toc.html) (last date accessed: 5 July 2005).
- Neter J., Kutner M., Nachtsheim C.J., Wasserman W. (1996) - *Applied linear statistical models*. McGraw-Hill, Boston, Massachusetts, pp. 1408.
- Pal M., Mather P.M. (2003) - *An assessment of the effectiveness of decision tree methods for land cover classification*. Remote Sensing of Environment, 86: 554-556. doi: [http://dx.doi.org/10.1016/S0034-4257\(03\)00132-9](http://dx.doi.org/10.1016/S0034-4257(03)00132-9).
- Pal M. (2012) - *Multinomial logistic regression-based feature selection for hyperspectral data*. International Journal of Applied Earth Observation and Geoinformation, 14 (1): 214-220. doi: <http://dx.doi.org/10.1016/j.jag.2011.09.014>.
- Prabhakar M., Prasad Y., Desai S., Thirupathi M., Gopika K., Rao G., Venkateswarlu B. (2013) - *Hyperspectral remote sensing of yellow mosaic severity and associated pigment losses in Vigna mungo using multinomial logistic regression models*. Crop Protection, 45: 132-140. doi: <http://dx.doi.org/10.1016/j.cropro.2012.12.003>.
- Peng C.J., Nichols R.N. (2003) - *Using multinomial logistic models to predict adolescent behavioral risk*. Journal of Modern Applied Statistical Methods, 2 (1): 177-188.
- Press S.J., Wilson S. (1978) - *Choosing between logistic regression and discriminant analysis*. Journal of the American Statistical Association, 73 (364): 699-705. doi: <http://dx.doi.org/10.2307/2286261>.
- Ramírez-García P., López-Blanco J., Ocaña D. (1998) - *Mangrove vegetation assessment in the Santiago River Mouth, Mexico, by means of supervised classification using Landsat TM imagery*. Forest Ecology and Management, 105: 217-229. doi: [http://dx.doi.org/10.1016/S0378-1127\(97\)00289-2](http://dx.doi.org/10.1016/S0378-1127(97)00289-2).
- Rocchini D., Foody G., Ricotta H., Anand M., He K., Amici V., Kleinschmit B., Förster M., Schmidtlein S., Feilhauer H., Ghisla A., Metz M., Neteler M. (2013) - *Uncertainty in ecosystem mapping by remote sensing*. Computers & Geosciences, 50: 128-135. doi: <http://dx.doi.org/10.1016/j.cageo.2012.05.022>.
- Rodriguez-Galiano V., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J. (2012) - *An assessment of the effectiveness of a random forest classifier for land-cover classification*. ISPRS Journal of Photogrammetry and Remote Sensing, 67: 93-104 doi: <http://dx.doi.org/10.1016/j.isprsjprs.2012.05.002>.

- org/10.1016/j.isprsjprs.2011.11.002.
- Smith C., Brown N. (1997) - *ERDAS field guide; fourth edition*. ERDAS Inc. Atlanta, Georgia. pp. 656.
- Sun J., Yang J., Zhang C., Yun W., Qu J. (2013) - *Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method*. Mathematical and Computer Modelling, 58 (3-4): 573-581. doi: <http://dx.doi.org/10.1016/j.mcm.2011.10.063>.
- Worth A.P., Cronin M.T.D. (2003) - *The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects*. Journal of Molecular Structure, 622: 97-111. doi: [http://dx.doi.org/10.1016/S0166-1280\(02\)00622-X](http://dx.doi.org/10.1016/S0166-1280(02)00622-X).
- Wrigley N. (1985) - *Categorical data analysis for geographers and environmental scientists*. Longman New York, pp. 385.
- Yoshida T., Omatu S. (1994) - *Neural network approach to land cover mapping*. IEEE Transactions on Geoscience and Remote Sensing, 32 (5): 1103-1109. doi: <http://dx.doi.org/10.1109/36.312899>.