

PENNSTATE



Center for **S**tatistical **E**cology and **E**nvironmental **S**tatistics

MODELING AND INTERPRETING THE ACCURACY ASSESSMENT ERROR MATRIX FOR A DOUBLY CLASSIFIED MAP

G. P. Patil and C. Taillie

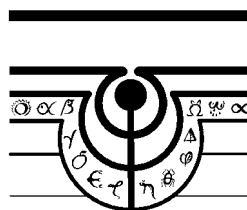
Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

EPA Project Officer: N. Phillip Ross

Prepared with partial support from the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Invited Paper to appear in *Environmental and Ecological Statistics*
Special Issue on Map Accuracy, S. Stehman and R. Czaplewski, guest editors

Technical Report Number 2000-0502
TECHNICAL REPORTS AND REPRINTS SERIES
May 2000



Modeling and Interpreting the Accuracy Assessment Error Matrix for a Doubly Classified Map

G. P. Patil and C. Taillie

Center for Statistical Ecology and Environmental Statistics

Department of Statistics

The Pennsylvania State University

University Park, PA 16802

Abstract. This paper considers two maps having the same spatial extent and the same mapping categories but where each map is subject to classification error. An overlay of the maps yields a (dis)similarity matrix whose (i, j) -entry is the areal proportion placed into category i by the first map and into category j by the second map. A parametric model, called the latent truth model, is proposed which specifies the dissimilarity matrix in terms of the true (but unknown) proportions for the mapping categories as well as the unknown error rates for the two maps. The number of parameters in the model exceeds the degrees of freedom in the dissimilarity matrix. However, a method of regularization is applied to effectively reduce the dimension of the parameter space and to permit model fitting. From the fitted model, one obtains estimates for the true mapping proportions as well as estimated error matrices for each of the maps. Accuracy assessment characteristics for each map (such as user's accuracy, producer's accuracy, overall accuracy, and the kappa coefficient) can be computed from the estimated error matrices. Methods are illustrated with two landcover maps of Wicomico County, Maryland.

Keywords: Dissimilarity matrix; Kullback-Liebler distance; Latent truth model; Producer's accuracy; Regularization; User's accuracy.

Prepared with partial support from the United States Environmental Protection Agency Cooperative Agreement Number CR-825506. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

While much attention is placed on overall accuracy percentages, by far the more interesting analysis concerns learning why sites do not fall on the diagonal of the error matrix. To both effectively use the map and to make better maps in the future, we need to know what causes the differences in the matrix. . . Analysis of the causes of differences in the error matrix can be one of the most important steps in the creation of a map from remotely sensed data. In the past, too much emphasis has been placed on the overall accuracy of the map, without delving into the conditions that give rise to that accuracy. By understanding what causes the reference and map data to differ, we can use the map more reliably, and produce both better maps and better accuracy assessment in the future.

R. G. Congalton and K. Green (1999)

Photo-interpretation errors should be addressed during the analysis and estimation phase. . . A statistical device may be designed to measure the accuracy of photo-interpretations.

Z. Zhu, L. Yang, S. V. Stehman, and R. L. Czaplewski (1999)

The results dispel the myth of ‘ground truth,’ and argue against the notion that field observation is inherently more accurate than aerial photographic interpretation or that aerial photographic interpretation is inherently more accurate than satellite image analysis. . . The data derived through aerial photographic interpretation are not always more reliable than data derived through classification of digital satellite data. The two approaches result in different types of error because each has its strengths and weaknesses. . . Accuracy assessment should also be viewed as a means of enhancing the major digital coverage, rather than purely as a means of critiquing error.

S. Khorram *et al.* (1999)

1 Introduction

Consider a raster map whose pixels have been classified into K categories using two independent classification methods which we refer to as the I -method and the J -method. We suppose that both methods are attempting to arrive at the same underlying “truth” but

each classification is potentially subject to error. Method I , for example, might be based on remote sensing and method J on photography or ground truthing. Let π_{ij} be the proportion of pixels assigned to category i by method I and to category j by method J . Here, i and j range from 1 to K and $\boldsymbol{\pi} = [\pi_{ij}]$ is the dissimilarity matrix for the two classification methods. Although the language of pixels and raster maps is convenient, the methods of this paper apply equally to vector maps where the (i, j) -entry of the dissimilarity matrix is the proportionate area assigned to category i and j by the overlaid maps.

Often, only a sample rather than the full universe of pixels is doubly classified and on the basis of the sample one arrives at an estimate $\hat{\pi}_{ij}$ of the dissimilarity matrix as well as an estimated dispersion matrix $\hat{\boldsymbol{\Sigma}}$ where

$$\Sigma_{(i,j),(i',j')} = \text{Cov}(\hat{\pi}_{ij}, \hat{\pi}_{i'j'}).$$

Here we are not concerned with questions of sampling design or formation of the estimators (which depend upon the design). Instead, we ask what, if anything, can be said about the accuracy of the two classifications, as opposed to their agreement or disagreement? Implicit in this question is the supposition that there is some latent classification T which is the “true” classification. Let p_t , $t = 1, \dots, K$, be the proportion of pixels which are assigned to category t by classification T . Also, for given t , let α_{ti} be the conditional probability that a pixel whose true class is t is assigned to category i by method I . Let β_{tj} be the corresponding conditional probability for method J . Observe that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are row stochastic matrices, whose entries are the so-called producer’s conditional accuracy rates (diagonal entries) and producer’s conditional error rates (off-diagonal entries). for

the two classification methods. Likewise, $[p_t\alpha_{ti}]$ is the (overall) error matrix for method I while $[p_t\beta_{tj}]$ is the (overall) error matrix for method J .

Conceptually, each pixel has a threefold classification and, assuming conditional independence of methods I and J , the three-way contingency table has relative frequencies given by

$$\pi_{tij} = p_t\alpha_{ti}\beta_{tj}, \quad t, i, j = 1, \dots, K.$$

Since classification T is latent, we only get to observe (or estimate) the dissimilarity matrix which is the marginal

$$\pi_{ij} = \pi_{+ij} = \sum_t p_t\alpha_{ti}\beta_{tj}, \quad i, j = 1, \dots, k. \quad (1)$$

We refer to (1) as the *latent truth model* for the given dissimilarity matrix. One of the reviewers points out that the model has close connections to latent structure analysis (Clogg and Goodman, 1984; Goodman, 1974; Green, 1951; McCutcheon, 1984).

The question now becomes: given the left hand side of this system of equations, can we solve the system to obtain estimates for \mathbf{p} , $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$? Intuitively, the answer is no, and we can formally see the reason for this answer by noting that there are

$$K^2 - 1 = (K + 1)(K - 1)$$

degrees of freedom on the left hand side. This is the number of independent equations.

On the other hand, the number of independent parameters on the right is

$$(K - 1) + K(K - 1) + K(K - 1) = (2K + 1)(K - 1). \quad (2)$$

Here, we have used the fact that the components of \mathbf{p} sum to unity and that the rows of $\boldsymbol{\alpha}$ and of $\boldsymbol{\beta}$ sum to unity. In all cases, the system is underdetermined (i.e., fewer equations than unknowns) and an infinitude of solutions can be expected with a corresponding infinitude of explanations for the same observed dissimilarity matrix.

2 Regularization of the Latent Truth Model

The foregoing situation of an underdetermined system of equations is known as non-identifiability of parameters in a model, i.e., distinct sets of parameter values yield exactly the same model outcomes for all observed quantities. There are only two solutions to problems of non-identifiability: (i) expand the set of observed variables so that all parameters in the model are identifiable or (ii) impose constraints on the set of parameters so as to reduce the dimension of the parameter space. It is the second option, sometimes called regularization, that we wish to explore here.

The conventional solution to the error matrix problem, though it is seldom acknowledged to be regularization, is to declare that one of the two classifications—say, J —is correct and to call it as the reference classification. To quote Khorram *et al.* (1999, top of p. 46) this can create “a self-fulfilling prophecy in which all measured error is assigned to the source of assumed lower quality.” Also, on the same page of Khorram *et al.* see the discussion of a study by Shapiro (1995). We analyze one of the Shapiro data sets below in Section (3). Assuming classification J is declared as the reference, this regularization of equation (1) sets $\beta_{tj} = \delta_{tj}$ (Kronecker delta) and reduces the number of parameters by

$K(K - 1)$ to the needed number, $(K + 1)(K - 1)$; see equation (2). Equation (1) becomes

$$\pi_{it} = p_t \alpha_{ti},$$

equating the dissimilarity matrix to the error matrix for classification I .

We would suggest an alternative, or perhaps supplemental, approach in which the needed parameter reduction is spread across both matrices α and β instead of concentrated in β . We want to reduce the dimension of the parameter space so that the number of independent parameters on the right hand side of equation (1) is $(K + 1)(K - 1)$ or fewer. If there are fewer than $(K + 1)(K - 1)$ parameters, the system is over-determined but it can still be solved by some minimization criterion such as least squares (see Section 3).

If the dissimilarity matrix is estimated, then the system of equations (1) becomes

$$\hat{\pi}_{ij} = \sum_t p_t \alpha_{ti} \beta_{tj}, \quad i, j = 1, \dots, K. \quad (3)$$

The solution will yield estimates for the parameters in the model and these estimates are implicitly functions of the $\hat{\pi}_{ij}$ so that the estimated dispersion matrix $\hat{\Sigma}$ together with the δ -method will yield an approximate dispersion matrix for the estimated model parameters. Diagonal entries in the dispersion matrix could be used to assess whether particular model parameters are significantly different from zero. However, the focus of this paper is on the modeling and model-fitting issues and we do not attempt to assess the precision of estimation.

Now let us consider how the parameter reduction might be achieved. In all, there are $(K + 1)(K - 1)$ degrees of freedom to spread across the three quantities \mathbf{p} , α and β in

equation (3). The components of \mathbf{p} , i.e., the marginal landcover proportions for the true classification, is of central interest and there is little that we can do in general to reduce the dimension of \mathbf{p} from $K - 1$. This leaves $K(K - 1)$ degrees of freedom to spread across the two matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For the parametric modeling, we do not have to treat $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ symmetrically. If we have reason to think that classification J is more accurate, we could assign only a few parameters to $\boldsymbol{\beta}$ to model its possible departure from the identity matrix and the remaining parameters would be available for a more robust modeling of the error structure represented by $\boldsymbol{\alpha}$. As pointed out above, the dispersion matrix could be used to test if the fitted departure of $\boldsymbol{\beta}$ from an identity matrix is statistically significant.

Although it is highly non-representative, we start with the simplest case of $K = 2$ categories. The true marginal landcover vector $[p, 1 - p]'$ involves one unknown parameter p . This leaves $K(K - 1) = 2 \cdot 1 = 2$ parameters to model the 2×2 row stochastic matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. With only one parameter per matrix, about the only possible models are

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 - \delta & \delta \\ \delta & 1 - \delta \end{bmatrix} \quad (4)$$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{bmatrix} \quad (5)$$

where the parameters δ and ϵ determine the error rates for the I and J classifications, respectively. The model requires that $0 \leq \epsilon, \delta \leq 1$, but in practice ϵ and δ will have to be fairly small (say, $\epsilon, \delta < \frac{1}{2}$) or else the two landcover categories become completely confused. Our model with $K = 2$ now has 3 independent equations with 3 unknown parameters (p, δ, ϵ). The system (1) can be solved analytically to obtain closed form expressions for

(p, δ, ϵ) in terms of the π_{ij} ; see Patil and Taillie (2000).

The unattractive feature of the model (4) is that each row of the matrix α has the same error rate δ . Likewise, for β . It would be much better if each row could have its own error rate. This can be accomplished (with K categories) if α is modeled as

$$\alpha = \begin{bmatrix} 1 - \delta_1 & \frac{\delta_1}{K-1} & \frac{\delta_1}{K-1} & \cdots & \frac{\delta_1}{K-1} \\ \frac{\delta_2}{K-1} & 1 - \delta_2 & \frac{\delta_2}{K-1} & \cdots & \frac{\delta_2}{K-1} \\ \frac{\delta_3}{K-1} & \frac{\delta_3}{K-1} & 1 - \delta_3 & \cdots & \frac{\delta_3}{K-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\delta_K}{K-1} & \frac{\delta_K}{K-1} & \frac{\delta_K}{K-1} & \cdots & 1 - \delta_K \end{bmatrix} \quad (6)$$

This involves K parameters $\delta_1, \dots, \delta_K$. Row t has the value $1 - \delta_t$ in the diagonal entry and $\delta_t/(K-1)$ in the off-diagonal entries. The analogous expression for β involves another K parameters $\epsilon_1, \dots, \epsilon_K$. We refer to these regularized models for α and β as the *diagonal* models. The special cases with $\delta_1 = \delta_2 = \dots = \delta_K \equiv \delta$ or $\epsilon_1 = \epsilon_2 = \dots = \epsilon_K \equiv \epsilon$ will be referred to as the *scalar* models. We were forced to use the scalar models for α and β when $K = 2$. For which values of K , then, can we use the more attractive diagonal model? In all, we would need $2K$ parameters to model α and β whereas we have $K(K-1)$ available degrees of freedom. Thus, we need $K-1 \geq 2$ or $K \geq 3$. If $K > 3$, there would still be extra degrees of freedom left over for more refined modeling of α and/or β .

In general, we would expect the diagonal entries of α and β to be large and the off-diagonal entries to be fairly small (diagonal dominance). The diagonal model given by equation (6) sets all the off-diagonal entries in a given row equal to one another; this is the sense in which it regularizes or smooths an otherwise insolvable problem. Certainly, we

do not expect the off-diagonal entries within a row to be exactly equal but it may not be a bad approximation since these entries are small and positive. Alternatively, it may be unreasonable to think that one could estimate the minor variations among these entries with acceptable precision.

We have taken a very simple approach to regularization in the above discussion. The general method of regularization employs a penalty function (Good and Gaskins, 1971; O’Sullivan, 1986; Press *et al.*, 1994, Section 18.4) and fits the model by minimizing an expression of the form,

$$\mathcal{D}(\boldsymbol{\pi}, \boldsymbol{\nu}) + \lambda \mathcal{P}(\boldsymbol{\nu}), \tag{7}$$

where $\boldsymbol{\pi}$ is the observed dissimilarity matrix, $\boldsymbol{\nu}$ is the modeled dissimilarity matrix, $\mathcal{D}(\cdot, \cdot)$ is a distance function, λ is a user-specified smoothing parameter, and $\mathcal{P}(\cdot)$ is the penalty function. In the present context, $\mathcal{P}(\cdot)$ should be chosen to penalize large values and/or large variability among the off-diagonal entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The user has to specify a numerical value for λ and there are no generally accepted guidelines for making this choice. Our approach to regularization sets certain of the off-diagonal entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to be equal to one another.

3 Example: Wicomico County Landcover Maps

We illustrate the methods with an overlay of two landcover maps of Wicomico County, Maryland, a low-lying region on the eastern shore of the Chesapeake Bay. Much of the wetland in the county is Palustrine forested and is difficult to identify as wetland with

Table 1: Similarity/Dissimilarity matrix for an overlay of two photointerpreted landcover maps of Wicomico County, Maryland. Entries are in acres. The row classification (I-method) is by the Maryland Water Resource Administration (MD-WRA) while the column classification (J-method) is by the Fish and Wildlife Service National Wetland Inventory (FWS-NWI). *Source:* Shapiro, 1995.

	Pal	Upl	Lac	Riv	Est	Total
Pal	7124	7311	33	8	15	14581
Upl	3193	136705	49	29	75	140050
Lac	32	58	458	0	0	548
Riv	178	61	5	567	41	852
Est	24	48	0	1	1044	1117
Total	10641	144183	545	605	1174	157148

remotely sensed data. The dissimilarity matrix for the two maps is given in Table 1. There are five landcover categories represented in the maps (aggregated at the system level of the Cowardin system):

- Palustrine (Pal)
- Upland (Upl)
- Lacustrine (Lac)
- Riverine (Riv)
- Estuarine (Est).

It is evident from Table 1 that there is substantial disagreement in the Palustrine versus Uplands categories. Shapiro (1995) points out that both MD-WRA and FWS-NWI are conservative in their wetlands delineation, that is, an area is classified as wetlands only if

Table 2: Observed relative similarity/dissimilarity matrix corresponding to Table 1. The entry in row i and column j is denoted by π_{ij} . Overall observed agreement (sum of the diagonal entries) is 0.929.

	Pal	Upl	Lac	Riv	Est	Total
Pal	0.0459	0.0465	0.0002	0.0001	0.0001	0.0928
Upl	0.0203	0.8699	0.0003	0.0002	0.0005	0.8912
Lac	0.0002	0.0004	0.0029	0.0000	0.0000	0.0035
Riv	0.0011	0.0004	0.0000	0.0036	0.0003	0.0054
Est	0.0002	0.0003	0.0000	0.0000	0.0066	0.0071
Total	0.0677	0.9175	0.0035	0.0038	0.0075	1.0000

Table 3: Fitted relative similarity/dissimilarity matrix. The entry in row i and column j is denoted by ν_{ij} and the fit is obtained by minimizing the Kullback-Liebler distance $\mathcal{D}_{\text{KL}}(\boldsymbol{\pi}, \boldsymbol{\nu})$. The minimized Kullback-Liebler distance is 0.000618.

	Pal	Upl	Lac	Riv	Est	Total
Pal	0.0459	0.0465	0.0001	0.0001	0.0001	0.0928
Upl	0.0203	0.8699	0.0003	0.0003	0.0003	0.8912
Lac	0.0003	0.0003	0.0029	0.0000	0.0000	0.0036
Riv	0.0006	0.0005	0.0002	0.0036	0.0002	0.0052
Est	0.0003	0.0003	0.0000	0.0000	0.0066	0.0073
Total	0.0675	0.9176	0.0036	0.0040	0.0073	1.0000

it is explicitly rather than potentially identified as wetlands through photointerpretation. The large size of the (1,2) entry compared with the (2,1) entry suggests that MD-WRA may be less conservative than FWS-NWI. This is also reflected by the size of the Palustrine row total compared with the Palustrine column total.

Dividing the entries of Table 1 by the grand total, 157148, gives the relativized dissimilarity matrix (Table 2), whose entries will be denoted by π_{ij} . We fit a latent truth model,

$$\nu_{ij} = \sum_{t=1}^5 p_t \alpha_{ti} \beta_{tj}$$

with appropriate regularization of α and β , to this relativized dissimilarity matrix. Our parametrization of α is given by

$$\alpha = \begin{bmatrix} 1 - \delta_1 - \delta_1^* & \delta_1^* & \delta_1/3 & \delta_1/3 & \delta_1/3 \\ \delta_2^* & 1 - \delta_2 - \delta_2^* & \delta_2/3 & \delta_2/3 & \delta_2/3 \\ \delta_3/4 & \delta_3/4 & 1 - \delta_3 & \delta_3/4 & \delta_3/4 \\ \delta_4/4 & \delta_4/4 & \delta_4/4 & 1 - \delta_4 & \delta_4/4 \\ \delta_5/4 & \delta_5/4 & \delta_5/4 & \delta_5/4 & 1 - \delta_5 \end{bmatrix}.$$

This parametrization is similar to that of the K -degrees of freedom model except that we have assigned separate parameters δ_1^* and δ_2^* to the (1,2) and (2,1) entries to account for the high confusion between the Palustrine and Uplands categories. In all, seven parameters are used to model α , namely $\delta_1, \dots, \delta_5, \delta_1^*$, and δ_2^* . We model the matrix β in exactly the same way as α using ϵ instead of δ to denote the error rate parameters:

$$\beta = \begin{bmatrix} 1 - \epsilon_1 - \epsilon_1^* & \epsilon_1^* & \epsilon_1/3 & \epsilon_1/3 & \epsilon_1/3 \\ \epsilon_2^* & 1 - \epsilon_2 - \epsilon_2^* & \epsilon_2/3 & \epsilon_2/3 & \epsilon_2/3 \\ \epsilon_3/4 & \epsilon_3/4 & 1 - \epsilon_3 & \epsilon_3/4 & \epsilon_3/4 \\ \epsilon_4/4 & \epsilon_4/4 & \epsilon_4/4 & 1 - \epsilon_4 & \epsilon_4/4 \\ \epsilon_5/4 & \epsilon_5/4 & \epsilon_5/4 & \epsilon_5/4 & 1 - \epsilon_5 \end{bmatrix}.$$

Our model involves $4 + 7 + 7 = 18$ parameters, including the four free parameters among p_1, \dots, p_5 . Since the relativized dissimilarity matrix provides $25 - 1 = 24$ degrees of freedom, we will not get a perfect fit but there will be 6 residual degrees of freedom.

We fit the model by minimizing the “distance” between the π_{ij} and the ν_{ij} . A variety of “distance” functions are possible. Perhaps the most obvious is the sum of squared differences:

$$\sum_{i,j=1}^5 (\pi_{ij} - \nu_{ij})^2.$$

This has the disadvantage that it regards the distance between .001 and .003 exactly the same as the distance between .201 and .203. Relativized distances which correct this deficiency include the Pearson chi-square distance,

$$\sum_{i,j=1}^5 \frac{(\pi_{ij} - \nu_{ij})^2}{\pi_{ij}},$$

and the Neyman chi-square distance,

$$\sum_{i,j=1}^5 \frac{(\pi_{ij} - \nu_{ij})^2}{\nu_{ij}}.$$

We will use the Kullback-Liebler distance given by

$$\mathcal{D}_{\text{KL}}(\boldsymbol{\pi}, \boldsymbol{\nu}) = \sum_{i,j=1}^5 \pi_{ij} \log \left(\frac{\pi_{ij}}{\nu_{ij}} \right).$$

Here, natural logarithms are used and we follow the usual convention that $0 \cdot \log(0) = 0$.

The Kullback-Liebler distance has close connections with maximum likelihood estimation.

In addition, if $\mathcal{D}_{\text{KL}}(\boldsymbol{\pi}, \boldsymbol{\nu})$ is expanded in powers of the differences $\pi_{ij} - \nu_{ij}$ and only the lowest order terms are retained then either the Pearson or the Neyman chi-square results

depending on how the expansion is performed (the Neyman and Pearson chi-square are equivalent to lowest order in $\pi_{ij} - \nu_{ij}$). An additional attractive feature of Kullback-Liebler fitting is that whenever two cells have the same modeled frequency ν_{ij} then pooling those cells gives the same fitted model parameters. Our method of regularization effectively sets selected off-diagonal frequencies equal and would be equivalent to pooling those cells under Kullback-Liebler fitting.

The error rate parameters, ϵ and δ , in the model are expected to be small. Since most minimization algorithms have difficulty converging when parameters are close to boundary constraints, we have performed the minimization with respect to the logarithms of these parameters in order to remove the constraint. The estimated parameter values are given in Table 4 and the resulting dissimilarity matrix ν appears in Table 3. The match between π and ν is excellent except possibly in cell (4,1) where the observed value is 0.0011 and the fitted value is 0.0006.

Referring to Table 4, the model parameters δ_1^* , δ_2^* , ϵ_1^* , ϵ_2^* are intended to account for the confusion between the Palustrine and Upland categories. Specifically, δ_2^* and ϵ_2^* are the rates at which Upland is incorrectly classified as Palustrine wetland by MD-WRA and FWS-NWI, respectively. The relatively small estimated values for these two parameters shows that the model has detected the fact both classifications are conservative in classifying areas as wetland. In contrast, δ_1^* and ϵ_1^* are the rates at which Palustrine is incorrectly classified as Upland and the large values for these parameters again reflects the conservativeness of both classifications. That ϵ_1^* is 1.6 times as large as δ_1^* suggests either

Table 4: Fitted parameters. The δ -parameters determine the estimated error rates for the MD-WRA classification and the ϵ parameters determine the estimated error rates for the FWS-NWI classification. The parameters p_1, \dots, p_5 are the estimated “true” landcover proportions.

δ_1	0.017629	ϵ_1	0.004609	p_1	0.089
δ_2	0.000410	ϵ_2	0.000998	p_2	0.897
δ_3	0.004317	ϵ_3	0.002273	p_3	0.003
δ_4	0.001460	ϵ_4	0.199153	p_4	0.005
δ_5	0.003331	ϵ_5	0.001325	p_5	0.007
δ_1^*	0.205206	ϵ_1^*	0.335445		
δ_2^*	0.026271	ϵ_2^*	0.009457		

Table 5: Comparison of landcover proportions. The second column gives the “true” landcover proportions as estimated from the model. The third and fourth columns give the landcover proportions for the MD-WRA and FWS-NWI classifications, respectively. These are the column and row marginal totals from Table 1.

Landcover Category	True p Model Estimate	Marginal p for MD-WRA	Marginal p for FWS-NMI
Pal	0.089	0.093	0.068
Upl	0.897	0.891	0.918
Lac	0.003	0.004	0.004
Riv	0.005	0.005	0.004
Est	0.007	0.007	0.008

that FWS-NWI is more conservative than MD-WRA or that photointerpretation is less effective in detecting wetlands.

Table 5 compares the true landcover proportions as estimated from the model with the landcover proportions for the two maps. For the Palustrine and Upland categories, the estimated true proportions are closer to those of MD-WRA than FWS-NWI, suggesting that MD-WRA is more accurate in discriminating between these two categories.

The fitted model also allows us to produce estimated error matrices for each classification method. Consider the MD-WRA classification. The (t, i) component of the estimated error matrix is given by

$$p_t \cdot \alpha_{ti}.$$

Numerical values for these components appear in Table 6. Note that the rows refer to the “true” category t while the columns refer to the MD-WRA classified category. Normalizing this error matrix so the rows sum to unity gives the producer’s error matrix (Table 7). The (t, i) component of this matrix is α_{ti} and gives the proportionate area of the map whose true class is t but which is labeled as class i by MD-WRA. The diagonal entries of this matrix are the so-called producer’s accuracy values. Finally, when the error matrix in Table 6 is normalized so that the columns sum to unity, we obtain the user’s error matrix (Table 8). The (t, i) entry in this matrix gives the proportionate area of the map which is labeled as class i by MD-WRA but which is actually of class t . Diagonal entries of this matrix are the user’s accuracy values.

The corresponding error matrices for the FWS-NWI map appear in Tables 9–11. Note that the Upland producer’s accuracy is slightly better for FWS-NWI versus MD-WRA ($\beta_{22} = 0.9895$ versus $\alpha_{22} = 0.9733$). This would be the expected result of the conservatism of the FWS-NWI map. But it has a major impact on the comparative overall accuracy of the two maps ($\sum_t p_t \beta_{tt}$ versus $\sum_t p_t \alpha_{tt}$) since the dominant Upland class in these two sums is heavily weighted with $p_2 = .897$ while the other components of \mathbf{p} are all less than 0.09. The result is that both maps have about the same overall accuracy (95–96 percent)

even though examination of two estimated producer's error matrices would indicate that MD-WRA map has superior accuracy in several respects. This weighting by the dominant category is a major shortcoming of the overall accuracy measure as a comparative accuracy assessment tool and emphasizes the need to examine all components of the error matrices.

Table 6: Estimated error matrix for the MD-WRA map. The rows correspond to the “true” category t , while the columns are the MD-WRA classification. The estimated overall accuracy (sum of diagonal entries) is 0.9562.

	Pal	Upl	Lac	Riv	Est	Total
Pal	0.0692	0.0183	0.0005	0.0005	0.0005	0.0891
Upl	0.0236	0.8729	0.0001	0.0001	0.0001	0.8968
Lac	0.0000	0.0000	0.0029	0.0000	0.0000	0.0029
Riv	0.0000	0.0000	0.0000	0.0045	0.0000	0.0045
Est	0.0000	0.0000	0.0000	0.0000	0.0066	0.0067
Total	0.0928	0.8912	0.0036	0.0052	0.0073	1.0000

Table 7: Estimated producer’s error matrix for the MD-WRA map. The rows correspond to the “true” category t , while the columns are the MD-WRA classification. Each row sums to unity

	Pal	Upl	Lac	Riv	Est
Pal	0.7772	0.2052	0.0059	0.0059	0.0059
Upl	0.0263	0.9733	0.0001	0.0001	0.0001
Lac	0.0011	0.0011	0.9957	0.0011	0.0011
Riv	0.0004	0.0004	0.0004	0.9985	0.0004
Est	0.0008	0.0008	0.0008	0.0008	0.9967

Table 8: Estimated user’s error matrix for the MD-WRA map. The rows correspond to the “true” category t , while the columns are the MD-WRA classification. Each column sums to unity.

	Pal	Upl	Lac	Riv	Est
Pal	0.7460	0.0205	0.1468	0.1015	0.0717
Upl	0.2539	0.9795	0.0344	0.0238	0.0168
Lac	0.0000	0.0000	0.8168	0.0006	0.0004
Riv	0.0000	0.0000	0.0005	0.8730	0.0002
Est	0.0001	0.0000	0.0016	0.0011	0.9108

Table 9: Estimated error matrix for the FWS-NWI map. The rows correspond to the “true” category t , while the columns are the FWS-NWI classification. The estimated overall accuracy (sum of diagonal entries) is 0.9594.

	Pal	Upl	Lac	Riv	Est	Total
Pal	0.0588	0.0299	0.0001	0.0001	0.0001	0.0891
Upl	0.0085	0.8874	0.0003	0.0003	0.0003	0.8968
Lac	0.0000	0.0000	0.0029	0.0000	0.0000	0.0029
Riv	0.0002	0.0002	0.0002	0.0036	0.0002	0.0045
Est	0.0000	0.0000	0.0000	0.0000	0.0067	0.0067
Total	0.0675	0.9176	0.0036	0.0040	0.0073	1.0000

Table 10: Estimated producer’s error matrix for the FWS-NWI map. The rows correspond to the “true” category t , while the columns are the FWS-NWI classification. Each row sums to unity

	Pal	Upl	Lac	Riv	Est
Pal	0.6599	0.3354	0.0015	0.0015	0.0015
Upl	0.0095	0.9895	0.0003	0.0003	0.0003
Lac	0.0006	0.0006	0.9977	0.0006	0.0006
Riv	0.0498	0.0498	0.0498	0.8008	0.0498
Est	0.0003	0.0003	0.0003	0.0003	0.9987

Table 11: Estimated user’s error matrix for the FWS-NWI map. The rows correspond to the “true” category t , while the columns are the FWS-NWI classification. Each column sums to unity.

	Pal	Upl	Lac	Riv	Est
Pal	0.8710	0.0326	0.0382	0.0338	0.0187
Upl	0.1257	0.9672	0.0833	0.0737	0.0408
Lac	0.0000	0.0000	0.8152	0.0004	0.0002
Riv	0.0033	0.0002	0.0627	0.8916	0.0307
Est	0.0000	0.0000	0.0006	0.0005	0.9097

References

- Clogg, C. C. and Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, **79**, 762–771.
- Congalton, R. G. and Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, Boca Raton, FL.
- Good, I. J. and Gaskins, R. A. (1971). Non-parametric roughness penalties for probability densities. *Biometrika*, **21**, 255–277.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215–231.
- Green, B. F. (1951) A general solution for the latent class model of latent structure analysis. *Psychometrika*, **16**, 151–166.
- Khorram, S., Biging, G. S., Chrisman, N. R., Colby, D. R., Congalton, R. G., Dobson, J. E., Ferguson, R. L., Goodchild, M. F., Jensen, J. R., and Mace, T. H. (1999). *Accuracy Assessment of Remote Sensing Derived Change Detection*. ASPRS Monograph Series, American Society for Photogrammetry and Remote Sensing. Bethesda, MD.
- McCutcheon, A. L. (1987). *Latent Class Analysis*. Sage, Newbury Park, California.

- O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science*, **1**, 502–527.
- Patil, G. P. and Taillie, C. (2000). *Analytic Solution of the Regularized Latent Truth Model for Binary Maps*. Technical Report Number 2000-0601, Center for Statistical Ecology and Environmental Statistics, Pennsylvania State University, University Park, PA 16802.
- Press, W. H., Teukolsky, A. A., Vetterling, W. T., and Flannery, B. P. (1994). *Numerical Recipes in C: The Art of Scientific Computing* (second edition). Cambridge University Press, New York.
- Shapiro, C. (1995). *Coordination and Integration of Wetland Data for Status and Trends and Inventory Estimates*. Technical Report 2. Federal Geographic Data Committee, Wetlands Subcommittee, U.S. Geological Survey. The report is available in HTML format at <http://www.usgs.gov/wetlands/>
- Wahba, G. (1986). Comment on O’Sullivan (1986). *Statistical Science*, **1**, 521–522.
- Zhu, Z., Yang, L., Stelman, S. V., and Czaplewski, R. L. (1999). Designing an accuracy assessment for a USGS regional land cover mapping program. In *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, K. Lowell and A. Jaton, eds. Ann Arbor Press, Chelsea, MI. pp. 393–398.