Quantification Error Versus Location Error in Comparison of Categorical Maps

R. Gil Pontius, Jr.

Abstract

This paper analyzes quantification error versus location error in a comparison between two cellular maps that show a categorical variable. Quantification error occurs when the quantity of cells of a particular category in one map is different from the quantity of cells of that category in the other map. Location error occurs when the location of a category in one map is different from the location of that category in the other map. The standard Kappa index of agreement is usually not appropriate for map comparison. This paper offers alternative statistics: (1) proportion correct with perfect ability to specify location, (2) proportion correct with perfect ability to specify quantity, (3) Kappa for no ability, (4) Kappa for location, and (5) Kappa for quantity. These statistics can help scientists improve classification. This paper applies these theoretical concepts to the validation of a land-use change model for the Ipswich Watershed in Massachusetts.

Introduction

The growth of GIS and remote sensing has made it increasingly necessary to develop statistical measurements of agreement between cellular maps (Congalton and Green, 1993). During a comparison, one map is usually considered reality and the other map is simulated by a model or satellite. The goal of the comparison is to measure the agreement between the two maps. This paper addresses a situation where both maps show a categorical variable, such as land type.

Creation of a contingency table is usually the first step in the objective comparison of maps (Congalton and Green, 1999). Table 1 shows the layout for a J-by-J contingency table, where J is the number of categories in each map. Each map grid cell is classified according to both its category in the reality map and its category in the simulated map. The entries of Table 1 give proportions of the study area. The "total" row at the bottom shows the marginal distribution of the quantity of grid cells in reality. The "total" column on the right shows the marginal distribution of the quantity of grid cells in the simulation. The sum of the main diagonal gives the proportion correct classification for the entire study area.

At first glance, the proportion correct classification seems to be a straightforward statistic, but it is tricky to interpret because a surprisingly high number of cells can be classified correctly due to chance. Therefore, it is necessary to incorporate the expected proportion correct classification due to chance in an index of agreement. Equation 1 gives one of the most popular indices: i.e.,

$$Kappa = \frac{(Po - Pc)}{(Pp - Pc)} \qquad (1)$$

where Po is the observed proportion correct, Pc is the expected

proportion correct due to chance, and Pp is the proportion correct when classification is perfect.

Kappa is usually attributed to Cohen (1960); however, Smeeton (1985) traces its history to Galton (1892). Pontius (1994) is probably one of many investigators who independently derived the Kappa. Kappa is a member of a family of indices (Landis and Koch, 1977; Aickin, 1990) that have the following desirable properties: (1) if classification is perfect, then Kappa $= 1$; (2) if observed proportion correct is greater than expected proportion correct due to chance, then Kappa > 0; (3) if observed proportion correct is equal to expected proportion correct due to chance, then Kappa $= 0$; and (4) if observed proportion correct is less than expected proportion correct due to chance, then Kappa < 0.

Many modelers, remote sensing specialists, and statisticians endorse the Kappa and encourage colleagues to adopt it (Rosenfield, 1986; Cartersen, 1987; Hudson and Ramm, 1987; Stehman, 1996). However, the appropriateness of the standard Kappa depends on whether or not the marginal distributions are fixed by the scientist (Cohen, 1960; Brennan and Prediger, 1981; Foody, 1992). The standard Kappa is appropriate for contingency tables when the scientist does not have control over the marginal distributions, such as a test of the agreement between smoking and respiratory illness. In contrast, usually a goal of a satellite or spatially explicit model is to obtain similar marginal distributions; hence, the standard method to compute the expected proportion correct classification by chance is usually not appropriate for classification schemes that attempt to specify accurately both quantity and location. A GIS-based classification scheme should be judged on its ability to produce accurately both quantities and locations of categories of grid cells in a map (Card, 1982; Congalton et al., 1983; Congalton, 1991). For such classification schemes, the methods section describes some appropriate ways to compute alternatives to the standard Kappa. This paper also gives new methods for scientists to validate and to improve a wide variety of spatially explicit classification schemes such as those described in Turner (1988), Monserud and Leemans (1992), Hall (1995), Lambin (1996), Veldkamp and Fresco (1996), Liverman (1998), Wu (1998), etc.

Methods

Quantification versus Location

When comparing maps for validation, we can distinguish between quantification error and location error. Quantification error occurs when the quantity of cells of a category in

Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA 01610 (rpontius@clarku.edu).

Simulation	Reality				
	1	2	...	J	total
1	p_{11}	p_{12}		p_{1J}	$S_1 = \Sigma p_{1j}$
2	p_{21}	p_{22}		p_{2J}	$S_2 = \Sigma p_{2j}$
...					...
J	p_{J1}	p_{J2}		p_{JJ}	$S_J = \Sigma p_{Jj}$
total	$R_1 = \Sigma p_{j1}$	$R_2 = \Sigma p_{j2}$	...	$R_J = \Sigma p_{jJ}$	1

one map is different than the quantity of cells of that category in the other map. Even if there is no quantification error, location error can occur where the location of a category in one map is different than the location of that category in the other map.

Each spatially explicit simulation can be categorized according to its ability to specify accurately both quantity and location. Table 2 shows the expected proportion correct classification of simulations according to their idealized ability to specify quantity and location. Simulations that have no ability to specify location accurately are in the NL column of Table 2. Simulations in the PL column have perfect ability to specify location accurately. Simulations that have no ability to specify quantity accurately are in the NQ row. Simulations in the PQ row have perfect ability to specify quantity accurately. The MQ row and ML column show medium levels of ability to specify accurately quantity and location, respectively. Most simulations have some medium ability to predict both quantity and location, so the central entry of Table 2 is the observed proportion correct classification, denoted *Po*. The following subsections derive the equations in Table 2 and define "no ability," "medium ability," and "perfect ability."

Specify Location at Random

Let us consider the expected proportion correct classification of a simulation that has no ability to specify quantity accurately and no ability to specify location accurately. The simulation assigns $1/J$ of the cells to each of the J categories because the simulation has no ability to distinguish among the categories (Foody, 1992). The simulation distributes those cells at random locations on the map; therefore, the expected proportion correct with no information of quantity and no information of location (NQNL) is $1/J$ as given in the NQ row NL column of Table 2. Equation 2 gives the derivation; i.e.,

$$\text{NQNL} = \sum_{j=1}^{J} (1/J) \, R_j = (1/J) \sum_{j=1}^{J} R_j = 1/J \qquad (2)$$

where $1/J$ is the proportion of cells assigned to category j by simulation and R_j is the proportion of cells in category j in reality.
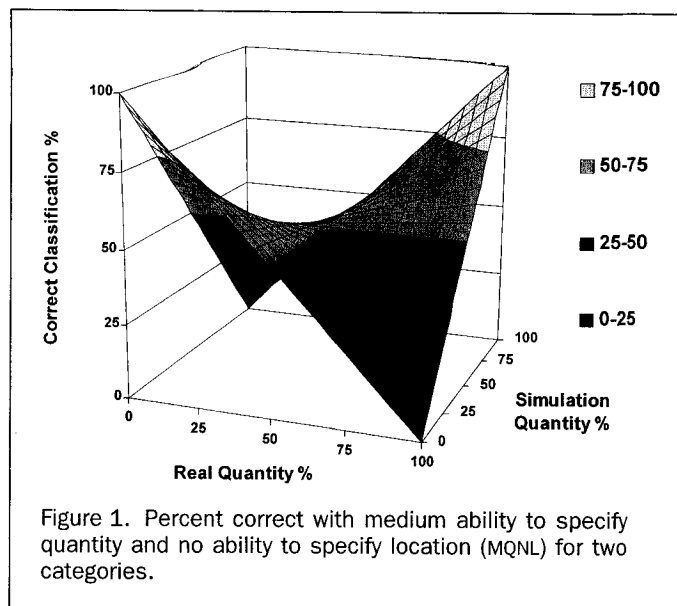

Figure 1. Percent correct with medium ability to specify quantity and no ability to specify location (MQNL) for two categories.

Thus, $Po - \text{NQNL}$ is the success attributable to the simulation.

Let s_j be the proportion of cells in category j in the simulation. Using the same mathematical reasoning, the MQ row NL column of Table 2 gives the proportion classified correctly associated with a simulation that has some medium ability to specify quantity accurately, but assigns locations at random (MQNL). For a fixed distribution of S_j, $\text{MQNL} - \text{NQNL}$ is the success due to the simulation's ability to specify quantity, and $Po - \text{MQNL}$ is the success due to the simulation's ability to specify location.

Figure 1 shows the surface of MQNL for the case where there are two categories, j and not j. The success space is defined by the reality axis (R-axis), the simulation axis (S-axis), and the correct axis (C-axis). The R-axis shows the proportion of cells of category j in reality, the S-axis shows the proportion of cells of category j in the simulation, and the C-axis shows the proportion of cells classified correctly by the simulation. The magnitude of the quantification error is S minus R, so any point outside the S = R plane has quantification error. Figure 1 is symmetric about the S = R plane which illustrates the fact that it makes no difference mathematically which map is considered simulated and which map is considered reality. The ability to specify quantity accurately is valuable when the proportion of category j in the real landscape is near either 0 or 1.

In the next case, if the simulation has perfect ability to specify quantity and no ability to specify location, then simulation assigns $S_j = R_j$, and locations at random. The PQ row NL column of Table 2 gives the expected proportion correct classification with perfect ability of quantity and no ability of location PQNL.

Ability to Specify Quantity	Ability to Specify Location		
	No (NL)	Medium (ML)	Perfect (PL)
No (NQ)	$1/J$	$(1/J) + \text{Klocation}(\text{NQPL} - (1/J))$	$\sum_{j=1}^{J} \min((1/J), R_j)$
Medium (MQ)	$\sum_{j=1}^{J} (S_j \bullet R_j)$	Proportion correct observed, denoted *Po*	$\sum_{j=1}^{J} \min(S_j, R_j)$
Perfect (PQ)	$\sum_{j=1}^{J} (R_j^2)$	$\text{PQNL} + \text{Klocation}(1 - \text{PQNL})$	1

$\rightarrow$ multiplication

Specify Location Perfectly

Now let's turn our attention to simulations that have perfect ability to specify location accurately, found in the PL column of Table 2. These simulations do the best job possible at assigning a specified quantity of cells to the correct locations within the map, such that the total proportion correct classification is limited only by quantification error. The proportion correct classification for any particular category is at most the smaller of (a) the total proportion of that category in the simulated map and (b) the total proportion of that category in the reality map.

For the case where we have no information of quantity, the simulation sets the proportion in each category to $1/J$. The NQ row PL column of Table 2 gives the proportion correct for a simulation with no information of quantity and perfect information of location (NQPL).

Next, consider a simulation that has a medium ability to specify quantity accurately and perfect ability to specify location (MQPL). The MQ row PL column of Table 2 gives MQPL by using the same mathematical logic as NQPL. Figure 2 shows MQPL for $J = 2$. A quantification error of ± 1 unit creates a decrease of one unit in proportion correct. For a fixed distribution of S_j, MQPL $-$ Po is the error due to a simulation's specification of location, and $1 -$ MQPL is the error due to a simulation's specification of quantity.

Finally, the PQ row PL column of Table 2 shows that a simulation that has perfect ability to specify accurately both quantity and location attains perfect classification; thus, PQPL $= 1$.

Kstandard

Recall that Kappa compares the observed proportion correct to the expected proportion correct due to chance. According to Table 2, the expected proportion correct due to chance of a simulation with no information of quantity and no information of location is $1/J$. However, the standard Kappa does not use $1/J$ as the expected proportion correct due to chance; the standard Kappa uses MQNL from Table 2. Table 3 gives the equation for standard Kappa, denoted as Kstandard. Kstandard measures a simulation's ability to attain perfect classification, given a fixed marginal distribution of S_j.

Kno

Let us now define an alternative to the standard Kappa. Kappa for no ability, denoted Kno, indicates the proportion classified correctly relative to the expected proportion classified correctly by a simulation with no ability to specify accurately
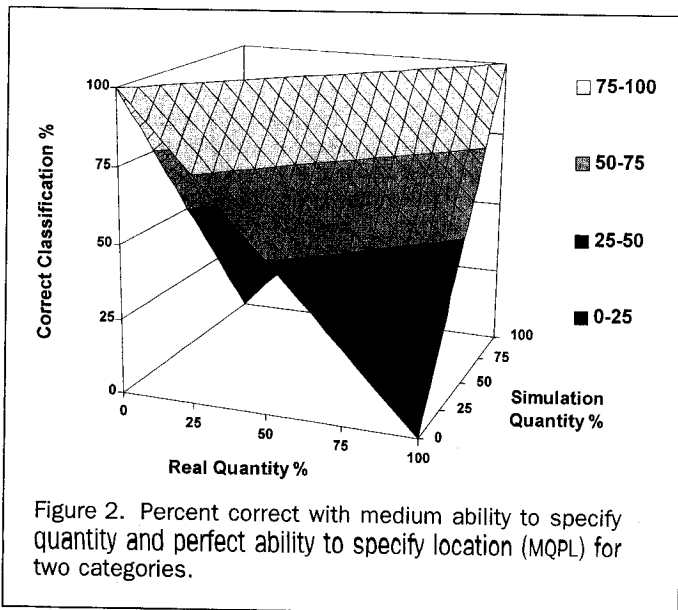

Figure 2. Percent correct with medium ability to specify quantity and perfect ability to specify location (MQPL) for two categories.

TABLE 3. FORMULAS AND SIMULATION RESULTS FOR VARIATIONS OF KAPPA

Kappa Definition		Ipswich Results	
Variations	Formula	Formula	Kappa (%)
Kstandard	$\dfrac{Po\ -\ \text{MQNL}}{1\ -\ \text{MONL}}$	$\dfrac{0.91\ -\ 0.90}{1.00\ -\ 0.90}$	12
Kno	$\dfrac{Po\ -\ \text{NQNL}}{1\ -\ \text{NQNL}}$	$\dfrac{0.91\ -\ 0.50}{1.00\ -\ 0.50}$	82
Klocation	$\dfrac{Po\ -\ \text{MQNL}}{\text{MQPL}\ -\ \text{MQNL}}$	$\dfrac{0.91\ -\ 0.90}{0.94\ -\ 0.90}$	28
Kquantity	$\dfrac{Po\ -\ \text{NQML}}{\text{PQML}\ -\ \text{NQML}}$	$\dfrac{0.91\ -\ 0.51}{0.97\ -\ 0.51}$	87

quantity or location. Table 3 gives Kno where Po = observed proportion of cells classified correctly, Pc = NQNL = $1/J$, and $Pp = 1$. Kno is the success attributable to the simulation divided by the maximum possible success attributable to a perfect simulation. Both Kstandard and Kno attain 1 only when $R_j = S_j$ and location is perfect. However, there are three major problems with Kstandard. Kno fixes the first two of these problems listed below.

The first major problem is that Kstandard fails to penalize for large quantification error. Kstandard can be positive over all combinations of R and S. Kstandard can be positive even when the observed proportion of cells classified correctly is less than $1/J$, whereas Kno is negative when there is large quantification error.

The second problem is that Kstandard fails to reward sufficiently when S is near R, especially when R is near 0 or 1. Figure 1 shows that MQNL is near 1 when both R and S are near 0 or 1. The simulation should receive reward for specifying quantity accurately, but Kstandard attributes those correct classifications to chance. Kno properly rewards the simulation for specifying quantity accurately.

The third major problem is that Kstandard fails to distinguish clearly between quantification error and location error. For example, if quantification error exists, then Kstandard is constrained to be less than 1, even when a simulation could be otherwise perfect at specifying location. Conversely, if location error exists, then Kstandard is constrained to be less than 1, even when a simulation could be perfect at specifying quantity. Kno also has this third problem.

Klocation

Therefore, let us define a Kappa for location, denoted Klocation, as shown in Table 3, where Po = observed proportion of cells classified correctly, Pc = MQNL, and Pp = MQPL. Given a fixed distribution of S_j, Klocation is the success due to the simulation's ability to specify location divided by the maximum possible success due to a simulation's ability to specify location perfectly. The upper range of Klocation is 1, regardless of quantification error.

Klocation is helpful in defining a success space similar to Figures 1 and 2. In Figures 3 and 4, the Klocation axis replaces the proportion simulated (S) axis of Figures 1 and 2. Figure 3 shows the proportion correct classification of a simulation for which there is no information concerning quantity; therefore, the proportion of map grid cells assigned to each category is $1/J$. If the simulation assigns location at random (that is, Klocation = 0), then the expected proportion correct is $1/J$. If the simulation is perfect at specifying location (that is, Klocation = 1), then the proportion correct is NQPL. The ability to specify location accurately is valuable when the proportion of category j in the real landscape is near $1/J$.

Figure 4 shows the proportion correct classification for a simulation in which there is perfect information concerning
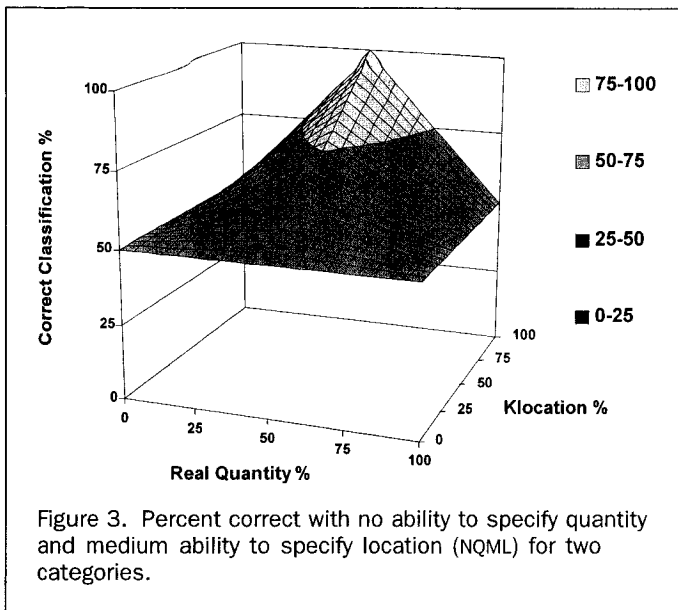
Figure 3. Percent correct with no ability to specify quantity and medium ability to specify location (NQML) for two categories.
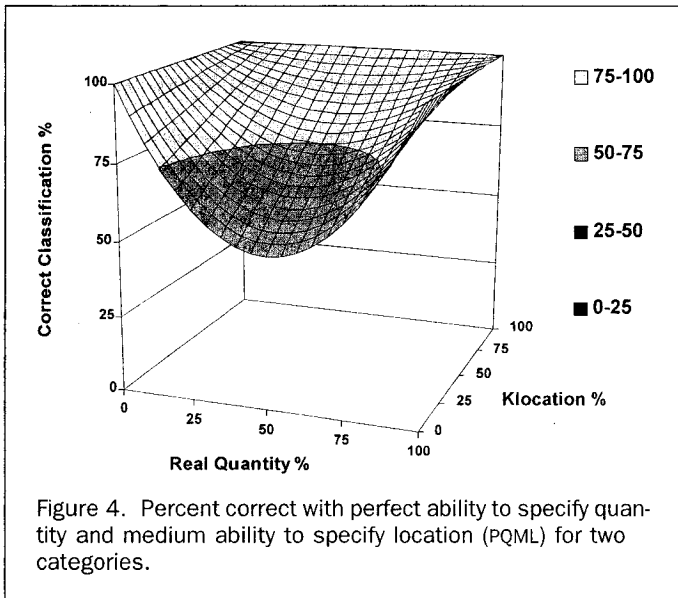

Figure 4. Percent correct with perfect ability to specify quantity and medium ability to specify location (PQML) for two categories.

quantity, that is, $S_j = R_j$. If the simulation assigns location at random (that is, Klocation = 0), then the expected proportion correct is PQNL. If the simulation specifies locations perfectly (i.e., Klocation = 1), then the proportion correct is 1. The ability to specify quantity accurately is valuable when the proportion of category j in the real landscape is near 0 or 1.

Kquantity
Given a fixed Klocation, column ML of Table 2 defines (a) proportion correct with no ability to specify quantity and medium ability to specify location (NQML) and (b) proportion correct with perfect ability to specify quantity and medium ability to specify location (PQML). For a fixed Klocation, NQML−NQNL is the success due to a simulation's specification of location, Po − NQML is the success due to a simulation's specification of quantity, PQML − Po is the error due to a simulation's specification of quantity, and 1 − PQML is the error due to a simulation's specification of location.

NQML and PQML are helpful in defining a Kappa for quantity, denoted Kquantity. Table 3 defines Kquantity, where $Po =$

proportion correct observed, Pc = NQML, and Pp = PQML. Given a fixed Klocation, Kquantity is the success due to the simulation's ability to specify quantity divided by the maximum possible success due to a simulation's ability to specify quantity perfectly.

Application to the Ipswich Watershed, Massachusetts
To illustrate the theoretical concepts, I apply the above statistics to a land-use change model that uses a multi-criteria evaluation to predict human induced disturbance to a landscape. The model predicts the quantity and location of new residential land versus other land from 1985 to 1991 in the watershed of Ipswich, Massachusetts. The model is calibrated with maps of socio-physical characteristics and residential areas in 1971 and 1985. The model predicts locations of land-use change due to new residential development from 1985 to 1991 based on suitability for human habitation due to socio-physical characteristics, such as proximity to roads. The model predicts that the quantity of new residential land in Ipswich Massachusetts between 1985 and 1991 is 8 percent of the watershed, based on an exponential extrapolation from 1951 to 1985. MassGIS (1999) supplies the maps for calibration and validation.

Results
Figure 5 shows the spatial arrangement of those grid cells that were classified correctly and incorrectly in the simulation for the Ipswich watershed. Table 3 shows Kno is 82 percent, which is a result of poor ability to specify location (Klocation = 28 percent) and good ability to specify quantity (Kquantity = 87 percent). Kstandard of 12 percent indicates that specification of location is poor, given a fixed quantity. The quantity of disturbance in the simulated map is 8 percent, whereas the quantity of disturbance in reality is 2 percent.

The overall proportion correct is 91 percent. MQPL = 94 percent, which indicates that if the scientist improves specification of location only, then proportion correct can increase by at most 3 percentage points, as shown in Figure 6. However, PQML = 97 percent, which indicates that if the scientist improves specification of quantity, then the proportion correct can increase by 6 percentage points.

Discussion
The Ipswich watershed example shows that the statistics given in this paper can be useful for analysis and improvement of classification. PQML and MQPL show that, if the purpose of the model is to maximize percent correct, then it is more important to improve the specification of quantity than to improve specification of location. Perhaps, the scientist should consider determining quantity by a linear rather than exponential extrapolation. Only 2 percent of the watershed underwent change from 1985 to 1991, so an accurate specification of quantity can yield a large percent correct, even when locations are distributed at random. If the quantity of change were close to 50 percent of the watershed, then it would be important to specify location accurately.

However, a large percent correct is not necessarily an important criterion to judge classification schemes because a large portion of percent correct can be attributable to chance. If a scientist is interested in location, then Klocation can be a more important criterion than either Kno or percent correct. For the Ipswich watershed example, Klocation shows that specification of location is poor, but MQPL shows that, if specification of location were good, then percent correct would increase only modestly.

As a warning, NQML, PQML, and Kquantity are based on the assumption that Klocation is constant across various scenarios of quantification. However, Klocation can change when quantification changes. Nevertheless, NQML, PQML, and Kquantity serve as reasonable guide posts for the scientist.
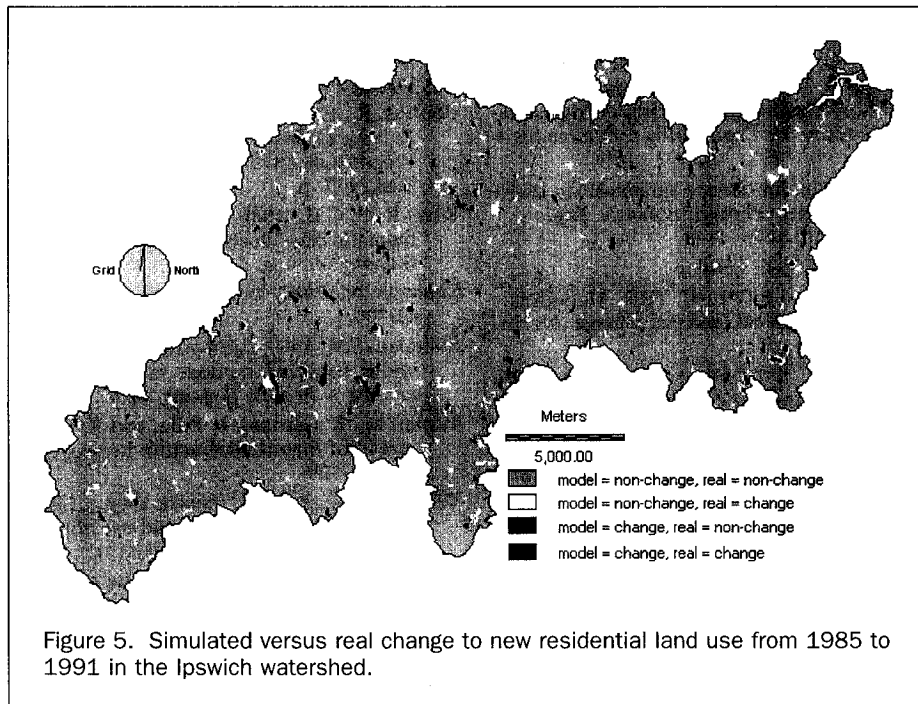
Figure 5. Simulated versus real change to new residential land use from 1985 to 1991 in the Ipswich watershed.
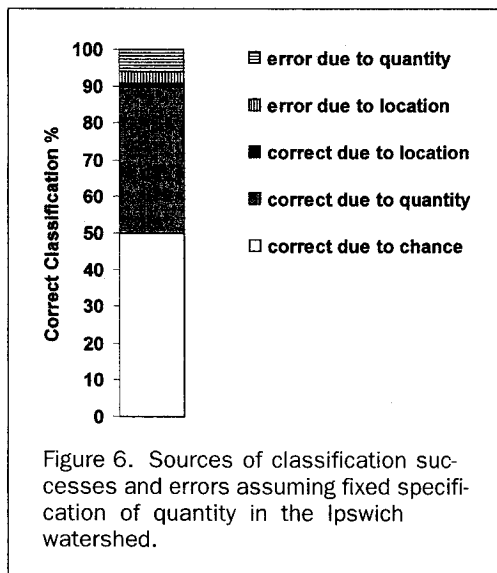

Figure 6. Sources of classification successes and errors assuming fixed specification of quantity in the Ipswich watershed.

The standard Kappa offers almost no useful information because it confounds quantification error with location error. A Kstandard of 12 percent would seem to indicate that the classification is poor, when in fact 91 percent of the cells are classified correctly. Kstandard fails to reward the simulation for good specification of quantity.

Conclusions

This paper gives equations to enable scientists to separate overall classification error into quantification error and location error. This insight can help scientists to decide whether to dedicate energy to improve a simulation's ability to specify quantity versus location. As a rule of thumb, if the quantity of grid cells in the map of reality is distributed equally among the categories, then percent correct classification is large when location is specified accurately. If the quantity of grid cells in the map of reality is not distributed equally among the categories, then percent correct classification is large when quantity is specified accurately. Specifically, when PQML − Po is large, then scientists can improve percent correct by improving the simulation's ability to specify quantity. When MQPL − Po is large, then scientists can improve percent correct by improving the simulation's ability to specify location. Scientists should use Kno to evaluate the simulation's overall success, Klocation to evaluate the simulation's ability to specify location, and Kquantity to evaluate the simulation's ability to specify quantity.

Acknowledgments

References

Aickin, M., 1990. Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's Kappa, *Biometrics*, 46:293–302.

Brennan, R.L., and D.J. Prediger, 1981. Coefficient Kappa: Some uses, misuses and alternatives, *Educational and Psychological Measurement*, 41:687–698.

Card, D., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy, *Photogrammetric Engineering & Remote Sensing*, 48(3):431–439.

Cartersen, L.W., Jr., 1987. A measure of similarity for cellular maps, *The American Geographer*, 14(4):345–358.

Cohen, J., 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1):37–46.

Congalton, R., 1991. A review of assessing the accuracy of classification of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R., and K. Green, 1993. A practical look at the sources of confusion in error matrix generation, *Photogrammetric Engineering & Remote Sensing*, 59(5):641–644.

———, 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York, 137 p.

Congalton, R., R. Oderwald, and R. Mead, 1983. Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques, *Photogrammetric Engineering & Remote Sensing*, 49(12):1671–1678.

Foody, G., 1992. On the compensation for chance agreement in image classification accuracy assessment, *Photogrammetric Engineering & Remote Sensing*, 58(10):1459–1460.

Galton, F., 1892. *Finger Prints*, Macmillan, London.

Hall, C., H. Tian, Y. Qi, G. Pontius, and J. Cornell, 1995. Modelling spatial and temporal patterns of tropical land use change, *Journal of Biogeography*, 22:753–757.

Hudson, W.D., and C.W. Ramm, 1987. Correct formulation of the Kappa coefficient of agreement, *Photogrammetric Engineering & Remote Sensing*, 53(4):421–422.

Lambin, E.F., 1996. Change detection at multiple temporal scales: Seasonal and annual variations in landscape variables, *Photogrammetric Engineering & Remote Sensing*, 62(8):931–938.

Landis, J.R., and G.C. Koch, 1977. The measurement of observer agreement for categorical data, *Biometrics*, 33:159–174.

Liverman, D., E. Moran, R. Rindfuss, and P. Stern, (editors), 1998. *People and Pixels*, National Academy Press, Washington, D.C., 244 p.

MassGIS, 1999. *Land Use Map*, Executive Office of Environmental Affairs, Boston, Massachusetts.

Monserud, R., and R. Leemans, 1992. Comparing global vegetation maps with the Kappa statistic, *Ecological Modelling*, (62):275–293.

Pontius, R., 1994. *Modeling Tropical Land Use Change and Assessing Policies to Reduce Carbon Dioxide Release from Africa*, Graduate Program in Environmental Science. SUNY-ESF, Syracuse, New York, 177. p.

Rosenfield, G.H., 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52(2):223–227.

Smeeton, N., 1985. Early history of the kappa statistic, *Biometrics*, 41:795.

Stehman, S.V., 1996. Estimating the Kappa coefficient and its variance under stratified random sampling, *Photogrammetric Engineering & Remote Sensing*, 62(4):401–407.

Turner, M., R. Gardner, V. Dale, and R. O'Neill, 1988. Predicting the spread of disturbance across heterogeneous landscapes, *Oikos*, 55:121–129.

Veldkamp, A., and L.O. Fresco, 1996. CLUE-CR: An integrated multiscale model to simulate land use change scenarios in Costa Rica, *Ecological Modelling*, 91(1/3):231.

Wu, F., 1998. SimLand: A prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules, *International Journal of Geographical Information Science*, 12(1):63–82.