

CHAPTER 17

Components of Agreement between Categorical Maps at Multiple Resolutions

R. Gil Pontius, Jr. and Beth Suedmeyer

CONTENTS

17.1	Introduction.....	233
17.1.1	Map Comparison	233
17.1.2	Puzzle Example	234
17.2	Methods	236
17.2.1	Example Data	236
17.2.2	Data Requirements and Notation	236
17.2.3	Minimum Function.....	239
17.2.4	Agreement Expressions and Information Components	239
17.2.5	Agreement and Disagreement	242
17.2.6	Multiple Resolutions	244
17.3	Results.....	245
17.4	Discussion.....	248
17.4.1	Common Applications	248
17.4.2	Quantity Information	249
17.4.3	Stratification and Multiple Resolutions	250
17.5	Conclusions.....	250
17.6	Summary.....	251
	Acknowledgments	251
	References	251

17.1 INTRODUCTION

17.1.1 Map Comparison

Map comparisons are fundamental in remote sensing and geospatial data analysis for a wide range of applications, including accuracy assessment, change detection, and simulation modeling. Common applications include the comparison of a reference map to one derived from a satellite image or a map of a real landscape to simulation model outputs. In either case, the map that is

considered to have the highest accuracy is used to evaluate the map of questionable accuracy. Throughout this chapter, the term *reference map* refers to the map that is considered to have the highest accuracy and the term *comparison map* refers to the map that is compared to the reference map. Typically, one wants to identify similarities and differences between the reference map and the comparison map.

There are a variety of levels of sophistication by which to compare maps when they share a common categorical variable (Congalton, 1991; Congalton and Green, 1999). The simplest method is to compute the proportion of the landscape classified correctly. This method is an obvious first step; however, the proportion correct fails to inform the scientist of the most important ways in which the maps differ, and hence it fails to give the scientist information necessary to improve the comparison map. Thus, it would be helpful to have an analytical technique that budgets the sources of agreement and disagreement to know in what respects the comparison map is strong and weak. This chapter introduces map comparison techniques to determine agreement and disagreement between any two categorical maps based on the quantity and location of the cells in each category; these techniques apply to both hard and soft (i.e., fuzzy) classifications (Foody, 2002).

This chapter builds on recently published methods of map comparison and extends the concept to multiple resolutions (Pontius, 2000, 2002). A substantial additional contribution beyond previous methods is that the methods described in this chapter support stratified analysis. In general, these new techniques serve to facilitate the computation of several types of useful information from a generalized confusion matrix (Lewis and Brown, 2001). The following puzzle example illustrates the fundamental concepts of comparison of quantity and location.

17.1.2 Puzzle Example

Figure 17.1 shows a pair of maps containing two categories (i.e., light and dark). At the simplest level of analysis, we compute the proportion of cells that agree between the two maps. The agreement is 12/16 and the disagreement is 4/16. At a more sophisticated level, we can compute the disagreement in terms of two components: (1) disagreement due to quantity and (2) disagreement due to location. A disagreement of quantity is defined as a disagreement between the maps in terms of the quantity of a category. For example, the proportion of cells in the dark category in the comparison map is 10/16 and in the reference map is 12/16; therefore, there is a disagreement of 2/16. A disagreement of location is defined as a disagreement such that a swap of the location of a pair of cells within the comparison map increases overall agreement with the reference map. The disagreement of location is determined by the amount of spatial rearrangement possible in the comparison map, so that its agreement with the reference map is maximized. In this example, it would be possible to swap the #9 cell with the #3, #10, or #13 cell within the comparison map to increase its agreement with the reference map (Figure 17.1). Either of these is the only swap we

1	2	3	4	1	2	3	4
5	6	7	8	5	6	7	8
9	10	11	12	9	10	11	12
13	14	15	16	13	14	15	16

Comparison (forgery) Reference (masterpiece)

Figure 17.1 Demonstration puzzle to illustrate agreement of location vs. agreement of quantity. Each map shows a categorical variable with two categories: dark and light. Numbers identify the individual grid cells.

can make to improve the agreement, given the quantity of the comparison map. Therefore, the disagreement of location is $2/16$. The distinction between information of quantity and information of location is the foundation of this chapter's philosophy of map comparison.

It is worthwhile to consider in greater detail this concept of separation of information of quantity vs. information of location in map comparison before introducing the technical methodology of the analysis. The remainder of this introduction uses the puzzle example of Figure 17.1 to illustrate the concepts that the Methods section then formalizes in mathematical detail.

The following analogy is helpful to grasp the fundamental concept. Imagine that the reference map of Figure 17.1 is an original masterpiece that has been painted with two colors: light and dark. A forger would like to forge the masterpiece, but the only information that she knows for certain is that the masterpiece has exactly two colors: light and dark. Armed with partial information about the masterpiece (reference map), the forger must create a forgery (comparison map).

To create the forgery, the forger must answer two basic questions: What proportion of each color of paint should be used? Where should each color of paint be placed? The first question requires information of quantity and the second question requires information of location.

If the forger were to have perfect information about the quantity of each color of paint in the masterpiece, then she would use $4/16$ light paint and $12/16$ dark paint for the forgery, so that the proportion of each color in the forgery would match the proportion of each color in the masterpiece. The quantity of each color in the forgery must match the quantity of each color in the masterpiece in order to allow the potential agreement between the forgery and the masterpiece to be perfect. At the other extreme, if the forger were to have no information on the quantity of each color in the masterpiece, then she would select half light paint and half dark paint, since she would have no basis on which to treat either category differently from the other category. In the most likely case, the forger has a medium level of information, which is a level of information somewhere between no information and perfect information. Perhaps the forger would apply $6/16$ light paint and $10/16$ dark paint to the forgery, as in Figure 17.1.

Now, let us turn our attention to information of location. If the forger were to have perfect information about the location of each type of paint in the masterpiece, then she would place the paint of the forgery in the correct location as best as possible, such that the only disagreement between the forgery and the masterpiece would derive from error (if any) in the quantity of paint. If the forger were to have no information about the location of each color of paint in the masterpiece, then she would spread each color of paint evenly across the canvas, such that each grid cell would be covered smoothly with light paint and dark paint. In the most likely case, the forger has a medium level of information of location about the masterpiece, so perhaps the forgery would have a pair of grid cells that are incorrect in terms of location, as in Figure 17.1. However, in the case of Figure 17.1, the error of location is not severe, since the error could be corrected by a swap of neighboring grid cells.

After the forger completes the forgery, we compare the forgery directly to the masterpiece in order to find the types and magnitudes of agreement between the two. There are two basic types of comparison, one based on information of quantity and another based on information of location. Each of the two types of comparisons leads to a different follow-up question.

First, we could ask, Given its medium level of information of quantity, how would the forgery appear if the forger would have had perfect information on location during the production of the forgery? For the example, in Figure 17.1, the answer is that the forger would have adjusted the forgery by swapping the location of cell #9 with cell #3, #10, or #13. As a result, the agreement between the adjusted forgery and the masterpiece would be $14/16$, because perfect information on location would imply that the only error would be an error of quantity, which is $2/16$.

Second, we could ask, Given its medium level of information of location, how would the forgery appear if the forger would have had perfect information of quantity during the production of the forgery? In this case, the answer is that the forger would have adjusted the forgery by using more dark paint and less light paint, but each type of paint would be in the same location as in Figure

17.1. Therefore, the adjusted forgery would appear similar to Figure 17.1; however, the light cells of Figure 17.1 would be a smooth mix of light and dark, while the dark cells would still be completely dark. Specifically, the light cells would be adjusted to be $2/3$ light and $1/3$ dark; hence, the total amount of light and dark paint in the forgery would equal the total amount of light and dark paint in the masterpiece. As a result, the agreement between the adjusted forgery and the masterpiece would be larger than $12/16$. The exact agreement would require that we define the agreement between the light cells of the masterpiece and the partially light cells of the adjusted forgery.

The above analogy prepares the reader for the technical description of the analysis in the Methods section. In the analogy, the reference map is the masterpiece that represents the ground information, and the comparison map is the forgery that represents the classification of a remotely sensed image. The classification rule of the remotely sensed image represents the scientist's best attempt to replicate the ground information. In numerous conversations with our colleagues, we have found that it is essential to keep in mind the analogy of painting a forgery. We have derived all the equations in the Methods section based on the concepts of the analogy.

17.2 METHODS

17.2.1 Example Data

Categorical variables consisting of "forest" and "nonforest" are represented in three maps of example data (Figure 17.2). Each map is a grid of 12×12 cells. The 100 nonwhite cells represent the study area and the remaining 44 white cells are located out of the study area. We have purposely made a nonsquare study area to demonstrate the generalized properties of the methods. The methods apply to a collection of any cells within a grid, even if those cells are not contiguous, as is typically the case in accuracy assessment. Each map has the same nested stratification structure. The coarser stratification consists of two strata (i.e., north and south halves) separated by the thick solid line. The finer stratification consists of four substrata quadrates of 25 cells each, defined as the northeast (NE), northwest (NW), southeast (SE), and southwest (SW). The set of three maps illustrates the common characteristics encountered when comparing map classification rules. Imagine that Figure 17.2 represents the output maps from a standard classification rule (COM1), alternative classification rule (COM2), and the reference data (REF). Typically, a statistical test would be applied to assess the relative performance of the two classification approaches and to determine important differences with respect to the reference data. However, it would also be helpful if such a comparison would offer additional insights concerning the sources of agreement and disagreement.

Table 17.1a and Table 17.1b represent the standard confusion matrix for the comparison of COM1 and COM2 vs. REF. The agreement in Table 17.1a and Table 17.1b is 70% and 78%, respectively. Note that the classification in COM2 is identical to the reference data in the south stratum. In the north stratum, COM2 is the mirror image of REF reflected through the central vertical axis. Therefore, the proportion of forest in COM2 is identical to that in REF in both the north and south strata. For the entire study area, REF is 45% forest, as is COM2. COM1 is 47% forest. A standard accuracy assessment ends with the confusion matrices of Table 17.1.

17.2.2 Data Requirements and Notation

We have designed COM1, COM2, and REF to illustrate important statistical concepts. However, this chapter's statistical techniques apply to cases that are more general than the sample data of Figure 17.2. In fact, the techniques can compare any two maps of grid cells that are classified as any combination of soft or hard categories.

This means that each grid cell can have some membership in each category, ranging from no membership (0) to complete membership (1). The membership is the proportion of the cell that

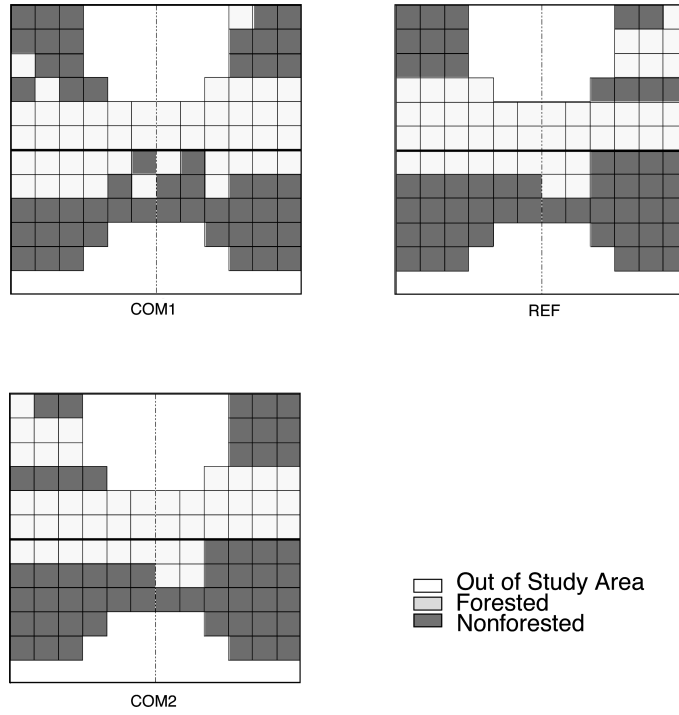


Figure 17.2 Three maps of example data.

Table 17.1a Confusion Matrix for COM1 vs. Reference

		Reference Map		
		Forest	Nonforest	Total
Comparison Map	Forest	31	16	47
	Nonforest	14	39	53
Total		45	55	100

Table 17.1b Confusion Matrix for COM2 vs. Reference

		Reference Map		
		Forest	Nonforest	Total
Comparison Map	Forest	34	11	45
	Nonforest	11	44	55
Total		45	55	100

belongs to a particular category; therefore, the sum of the membership values over all categories is 1. In addition, each grid cell has a weight to denote its membership in any particular stratum, where the stratum weight can also range from 0 to 1. The weights do not necessarily need to sum to 1. For example, if a cell's weights are 0 for all strata, then that cell is eliminated from the analysis. These ideas are expressed mathematically in Equation 17.1 through Equation 17.4, where j is the category index, J is the number of categories, R_{dnj} is the membership of category j in cell n of stratum d of the reference map, S_{dnj} is the membership of category j in cell n of stratum d of the comparison map, and W_{dn} is the weight for the membership of cell n in stratum d :

$$0 \leq R_{dnj} \leq 1 \quad (17.1)$$

$$0 \leq S_{dnj} \leq 1 \quad (17.2)$$

$$\sum_{j=1}^J R_{dnj} = \sum_{j=1}^J S_{dnj} = 1 \quad (17.3)$$

$$0 \leq W_{dn} \leq 1 \quad (17.4)$$

Just as each cell has some proportional membership to each category, each stratum has some proportional membership to each category. We define the membership of each stratum to each category as the proportion of the stratum that is covered by that category. For each stratum, we compute this membership to each category as the weighted proportion of the cells that belong to that category. Similarly, the entire landscape has membership to each particular category, where the membership is the proportion of the landscape that is covered by that category. We compute the landscape-level membership by taking the weighted proportion over all grid cells. Equation 17.5 through Equation 17.9 show how to compute these levels of membership for every category at both the stratum scale and the landscape scale. These equations utilize standard dot notation to denote summations, where N_d denotes the number of cells that have some positive membership in stratum d of the map and D denotes the number of strata. Equation 17.5 shows that $W_{d\cdot}$ denotes the sum of the cell weights for stratum d . Equation 17.6 shows that $R_{d\cdot j}$ denotes the proportion of category j in stratum d of the reference map. Equation 17.7 shows that $R_{\cdot j}$ denotes the proportion of category j in the entire reference map. Equation 17.8 shows that $S_{d\cdot j}$ denotes the proportion of category j in stratum d of the comparison map. Equation 17.9 shows that $S_{\cdot j}$ denotes the proportion of category j in the entire comparison map:

$$W_{d\cdot} = \sum_{n=1}^{N_d} W_{dn} \quad (17.5)$$

$$R_{d\cdot j} = \frac{\sum_{n=1}^{N_d} (W_{dn} * R_{dnj})}{W_{d\cdot}} \quad (17.6)$$

$$R_{\cdot j} = \frac{\sum_{d=1}^D \sum_{n=1}^{N_d} (W_{dn} * R_{dnj})}{\sum_{d=1}^D W_{d\cdot}} \quad (17.7)$$

$$S_{d \cdot j} = \frac{\sum_{n=1}^{N_d} (W_{dn} * S_{dnj})}{W_d} \quad (17.8)$$

$$S \cdot j = \frac{\sum_{d=1}^D \sum_{n=1}^{N_d} (W_{dn} * S_{dnj})}{\sum_{d=1}^D W_d} \quad (17.9)$$

17.2.3 Minimum Function

The Minimum function gives the agreement between a cell of the reference map and a cell of the comparison map. Specifically, Equation 17.10 gives the agreement in terms of proportion correct between the reference map and the comparison map for cell n of stratum d . Equation 17.11 gives the landscape-scale agreement weighted appropriately with grid cell weights, where $M(\mathbf{m})$ denotes the proportion correct between the reference map and the comparison map:

$$\text{agreement in cell } n \text{ of stratum } d = \sum_{j=1}^J \text{MIN}(R_{dnj}, S_{dnj}) \quad (17.10)$$

$$M(\mathbf{m}) = \frac{\sum_{d=1}^D \sum_{n=1}^{N_d} W_{dn} \left[\sum_{j=1}^J \text{MIN}(R_{dnj}, S_{dnj}) \right]}{\sum_{d=1}^D \sum_{n=1}^{N_d} W_{dn}} \quad (17.11)$$

The Minimum function expresses agreement between two cells in a generalized way because it works for both hard and soft classifications. In the case of hard classification, the agreement is either 0 or 1, which is consistent with the conventional definition of agreement for hard classification. In the case of soft classification, the agreement is the sum over all categories of the minimum membership in each category. The minimum operator makes sense because the agreement for each category is the smaller of the membership in the reference map and the membership in the comparison map for the given category. If the two cells are identical, then the agreement is 1.

17.2.4 Agreement Expressions and Information Components

Figure 17.3 gives the 15 mathematical expressions that lay the foundation of our philosophy of map comparison. The central expression, denoted $M(\mathbf{m})$, is the agreement between the reference map and the comparison map, given by Equation 17.11. The other 14 mathematical expressions show the agreement between the reference map and an "other" map that has a specific combination of information. The first argument in each Minimum function (e.g., R_{dnj}) denotes the cells of the reference map and the second argument in each Minimum function (e.g., S_{dnj}) denotes the cells of the other map. The components of information in the other maps are grouped into two orthogonal concepts: (1) information of quantity and (2) information of location.

There are three levels of information of quantity no, medium, and perfect, denoted, respectively, as \mathbf{n} , \mathbf{m} , and \mathbf{p} . For the five mathematical expressions in the "no information of quantity" column,

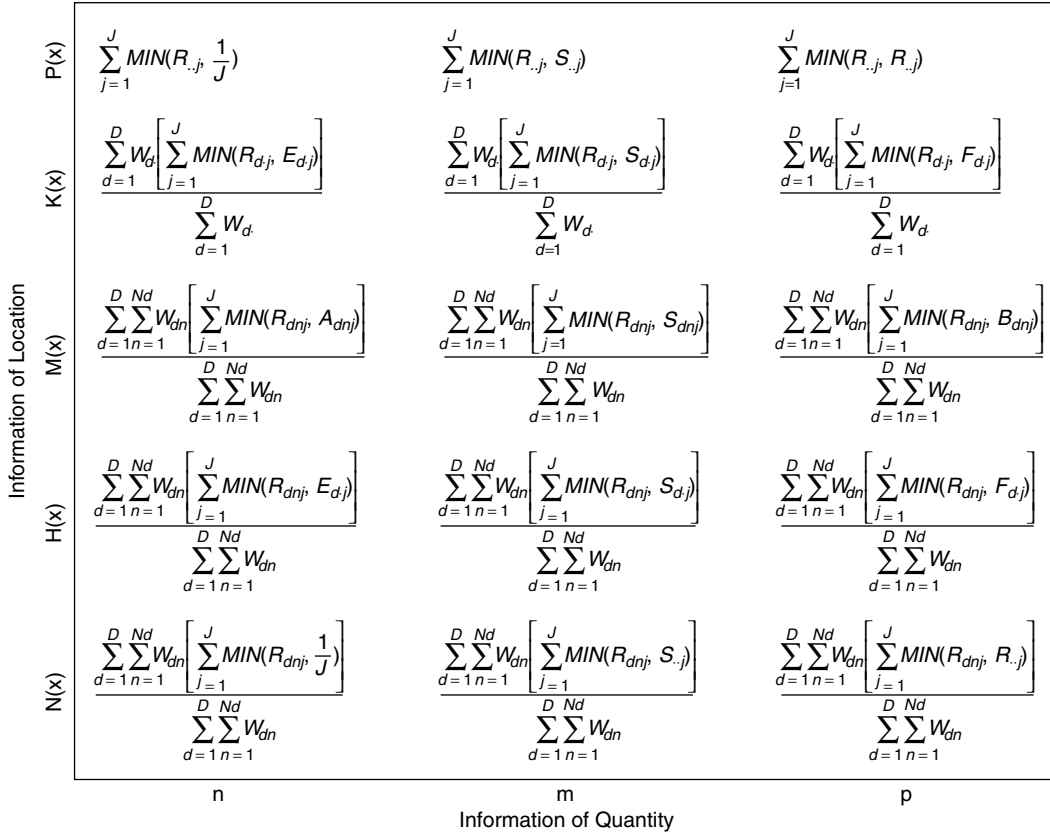


Figure 17.3 Expressions for 15 points defined by a combination of the information of quantity and the information of location. The vertical axis shows information of location and the horizontal axis shows information of quantity. The text defines the variables.

the other maps are derived from an adjustment to the comparison map, such that the proportion of membership for each of the J categories is $1/J$ in the other maps (Foody, 1992). This adjustment is necessary to answer the question, What would be the agreement between the reference map and the comparison map, if the scientist who created the comparison map would have had no information of quantity during its production? The adjustment holds the level of information of location constant while adjusting each grid cell such that the quantity of each of the J categories in the landscape is $1/J$.

Equations 17.12 and 17.13 give the necessary adjustment to each grid cell in order to scale the comparison map to express no information of quantity:

$$A_{dnj} = S_{dnj} \left(\frac{1/J}{S_{\cdot j}} \right), \quad \text{if } 1/J \leq S_{\cdot j} \tag{17.12}$$

$$= 1 - (1 - S_{dnj}) \left(\frac{S_{\cdot j}}{1/J} \right), \quad \text{else}$$

$$E_{d \cdot j} = S_{d \cdot j} \left(\frac{1/J}{S_{\cdot j}} \right), \quad \text{if } 1/J \leq S_{\cdot j} \tag{17.13}$$

$$= 1 - (1 - S_{d \cdot j}) \left(\frac{S_{\cdot \cdot j}}{1/J} \right), \text{ else}$$

Equation 17.12 performs the scaling at the grid cell level, and hence creates an “other” map, denoted A_{dnj} . Equation 17.13 performs the scaling at the stratum level, and hence creates an “other” map, denoted $E_{d \cdot j}$.

The logic of the scaling is as follows, where the word “paint” can be substituted for the word “category” to continue the painting analogy. If the quantity of category j in the comparison map is less than $1/J$, then more of category j must be added to the comparison map. In this case, category j is increased in cells that are not already 100% members of category j . If the quantity of category j in the comparison map is more than $1/J$, then some of category j must be removed from the comparison map. In that case, category j is decreased in cells that have some of category j .

For expressions in the “medium information” column of Figure 17.3, the other maps have the same quantities as the comparison map. For the expressions in the “perfect information” column, the other maps are derived such that the proportion of membership for each of the J categories matches perfectly with the proportions in the reference map. This adjustment is necessary to answer the question, What would be the agreement between the reference map and the comparison map, if the scientist would have had perfect information of quantity during the production of the comparison map? The adjustment holds the level of information of location constant while adjusting each grid cell such that the quantity of each of the J categories in the landscape matches the quantities in the reference map. The logic of the adjustment is similar to the scaling procedure described for the other maps in the “no information of quantity” column of Figure 17.3.

Equation 17.14 and Equation 17.15 give the necessary mathematical adjustments to scale the comparison map to express perfect information of quantity:

$$B_{dnj} = S_{dnj} \left(\frac{R_{\cdot \cdot j}}{S_{\cdot \cdot j}} \right), \quad \text{if } R_{\cdot \cdot j} \leq S_{\cdot \cdot j} \quad (17.14)$$

$$= 1 - (1 - S_{dnj}) \left(\frac{S_{\cdot \cdot j}}{R_{\cdot \cdot j}} \right), \text{ else}$$

$$F_{d \cdot j} = S_{d \cdot j} \left(\frac{R_{\cdot j}}{S_{\cdot j}} \right), \quad \text{if } R_{\cdot j} \leq S_{\cdot j} \quad (17.15)$$

$$= 1 - (1 - S_{d \cdot j}) \left(\frac{S_{\cdot j}}{R_{\cdot j}} \right), \text{ else}$$

Equation 17.14 performs this scaling at the grid cell level, and hence creates an “other” map, denoted B_{dnj} . Equation 17.15 performs this scaling at the stratum level, and hence creates an “other” map, denoted $F_{d \cdot j}$.

There are five levels of information of location: no, stratum, medium, perfect within stratum, and perfect, denoted, respectively, as $N(\mathbf{x})$, $H(\mathbf{x})$, $M(\mathbf{x})$, $K(\mathbf{x})$ and $P(\mathbf{x})$. Figure 17.3 shows the differences in the 15 mathematical expressions among these various levels of information of location. In $N(\mathbf{x})$, $H(\mathbf{x})$, and $M(\mathbf{x})$ rows, the mathematical expressions of Figure 17.3 consider the reference map at the grid cell level, as indicated by the use of all three subscripts: d , n , and j . In the $K(\mathbf{x})$ row, the mathematical expressions consider the reference map at the stratum level, as indicated by the use of two subscripts: d and j . In the $P(\mathbf{x})$ row, the expressions consider the reference map at the study area level, as indicated by the use of one subscript: j . In the $M(\mathbf{x})$ row, the

expressions consider the other maps at the grid cell level, as indicated by the use of all three subscripts: d , n , and j . In the $H(\mathbf{x})$ and $K(\mathbf{x})$ rows, the expressions consider the other maps at the stratum level, as indicated by the use of two subscripts: d and j . In the $N(\mathbf{x})$ and $P(\mathbf{x})$ rows, the expressions consider the other maps at the study area level, as indicated by the use of one subscript: j .

The concepts behind these combinations of components of information of location are as follows. In row $N(\mathbf{x})$, the categories of the other maps are spread evenly across the landscape, such that every grid cell has an identical multinomial distribution of categories. In row $H(\mathbf{x})$, the categories of the other maps are spread evenly within each stratum, such that every grid cell in each stratum has an identical multinomial distribution of categories. In row $M(\mathbf{x})$, the grid cell level information of location in the other maps is the same as in the comparison map. In row $K(\mathbf{x})$, the other maps derive from the comparison map, whereby the locations of the categories in the comparison map are swapped within each stratum in order to match as best as possible the reference map; however, this swapping of grid cell locations does not occur across stratum boundaries. In row $P(\mathbf{x})$, the other maps derive from the comparison map, whereby the locations of the categories in the comparison map are swapped in order to match as best as possible the reference map, and this swapping of grid cell locations can occur across stratum boundaries.

Each of the 15 mathematical expressions of Figure 17.3 is denoted by its location in the table. The \mathbf{x} denotes the level of information of quantity. For example, the overall agreement between the reference map and the comparison map is denoted $M(\mathbf{m})$, since the comparison map has a medium level of information of quantity and a medium level of information of location, by definition. The expression $P(\mathbf{p})$ is in the upper right of Figure 17.3 and is always equal to 1, because $P(\mathbf{p})$ is the agreement between the reference map and the other map that has perfect information of quantity and perfect information of location.

There are seven mathematical expressions that are especially interesting and helpful. They are $N(\mathbf{n})$, $N(\mathbf{m})$, $H(\mathbf{m})$, $M(\mathbf{m})$, $K(\mathbf{m})$, $P(\mathbf{m})$, and $P(\mathbf{p})$. For $N(\mathbf{n})$, each cell of the other map is the same and has a membership in each category equal to $1/J$. For $N(\mathbf{m})$, each cell of the other map is the same and has a membership in each category equal to the proportion of that category in the comparison map. For $H(\mathbf{m})$, each cell within each stratum of the other map is the same and has a membership in each category equal to the proportion of that category in each stratum of the comparison map. For $M(\mathbf{m})$, the other map is the comparison map. For $K(\mathbf{m})$, the other map is the comparison map with the locations of the grid cells swapped within each stratum, so as to have the maximum possible agreement with the reference map within each stratum. For $P(\mathbf{m})$, the other map is the comparison map with the locations of the grid cells swapped anywhere within the map, so as to have the maximum possible agreement with the reference map. For $P(\mathbf{p})$, the other map is the reference map, and therefore the agreement is perfect.

17.2.5 Agreement and Disagreement

The seven mathematical expressions $N(\mathbf{n})$, $N(\mathbf{m})$, $H(\mathbf{m})$, $M(\mathbf{m})$, $K(\mathbf{m})$, $P(\mathbf{m})$, and $P(\mathbf{p})$ constitute a sequence of measures of agreement between the reference map and other maps that have increasingly accurate information. Therefore, usually $0 < N(\mathbf{n}) < N(\mathbf{m}) < H(\mathbf{m}) < M(\mathbf{m}) < K(\mathbf{m}) < P(\mathbf{m}) < P(\mathbf{p}) = 1$. This sequence partitions the interval $[0,1]$ into components of the agreement between the reference map and the comparison map. $M(\mathbf{m})$ is the total proportion correct, and $1 - M(\mathbf{m})$ is the total proportion error between the reference map and the comparison map. Hence, the sequence of $N(\mathbf{n})$, $N(\mathbf{m})$, $H(\mathbf{m})$, and $M(\mathbf{m})$ defines components of agreement, and the sequence of $M(\mathbf{m})$, $K(\mathbf{m})$, $P(\mathbf{m})$, and $P(\mathbf{p})$ defines components of disagreement.

Table 17.2 defines these components mathematically. Beginning at the bottom of the table and working up, the first component is agreement due to chance, which is usually $N(\mathbf{n})$. However, if the agreement between the reference map and the comparison map is less than would be expected by chance, then the component of agreement due to chance may be less than $N(\mathbf{n})$. Therefore, Table 17.2 defines the component of agreement due to chance as the minimum of $N(\mathbf{n})$, $N(\mathbf{m})$, $H(\mathbf{m})$,

Table 17.2 Definition and Values of Seven Components of Agreement for COM1 vs. Reference Derived from the Mathematical Expressions of Figure 17.3

Name of Component	Definition	Percentage of Each Component	
		Stratum	Substratum
Disagreement due to quantity	$P(\mathbf{p}) - P(\mathbf{m})$	2.0	2.0
Disagreement at stratum level	$P(\mathbf{m}) - K(\mathbf{m})$	8.0	8.0
Disagreement at grid cell level	$K(\mathbf{m}) - M(\mathbf{m})$	20.0	20.0
Agreement at grid cell level	$\text{MAX} [M(\mathbf{m}) - H(\mathbf{m}), 0]$	12.2	11.5
Agreement at stratum level	If $\text{MIN} [N(\mathbf{m}), H(\mathbf{m}), M(\mathbf{m})] = N(\mathbf{m})$, then $\text{MIN} [H(\mathbf{m}) - N(\mathbf{m}), M(\mathbf{m}) - N(\mathbf{m})]$, else 0	7.5	8.2
Agreement due to quantity	If $\text{MIN} [N(\mathbf{n}), N(\mathbf{m}), H(\mathbf{m}), M(\mathbf{m})] = N(\mathbf{n})$, then $\text{MIN} [N(\mathbf{m}) - N(\mathbf{n}), H(\mathbf{m}) - N(\mathbf{n}), M(\mathbf{m}) - N(\mathbf{n})]$, else 0	0.3	0.3
Agreement due to chance	$\text{MIN} [N(\mathbf{n}), N(\mathbf{m}), H(\mathbf{m}), M(\mathbf{m})]$	50.0	50.0

and $M(\mathbf{m})$. The component of agreement due to quantity is usually $N(\mathbf{m}) - N(\mathbf{n})$; Table 17.2 gives a more general definition to account for the possibility that the comparison map's information of quantity can be worse than no information of quantity. The component of agreement at the stratum level is usually $H(\mathbf{m}) - N(\mathbf{m})$; Table 17.2 gives a more general definition to restrict this component of agreement to be nonnegative. Similarly, the component of agreement at the grid cell level is usually $M(\mathbf{m}) - H(\mathbf{m})$; Table 17.2 restricts this component of agreement to be nonnegative. Table 17.2 also defines the components of disagreement. It is a mathematical fact that $M(\mathbf{m}) \leq K(\mathbf{m}) \leq P(\mathbf{m}) \leq P(\mathbf{p})$; therefore, the components of disagreement are the simple definitions of Table 17.2.

The partition of the components of agreement can be performed for any stratification structure. Table 17.2 shows the results for the comparison of REF and COM1 at both the stratum level and the substratum level. Figure 17.4 shows this information in graphical form. The stratum bar shows the components at the stratum level and the substratum bar shows the components at the substratum level. Since the substrata are nested within the strata, it makes sense to overlay the stratum bar on

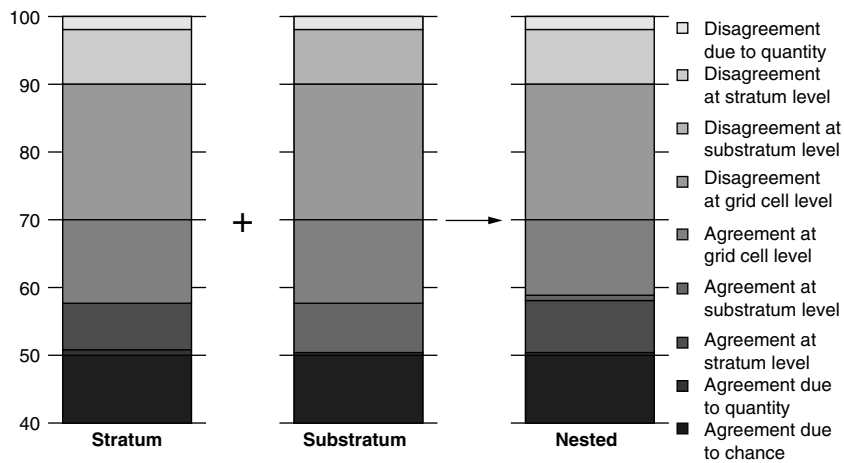


Figure 17.4 Stacked bars showing components of agreement between COM1 and REF. The vertical axis shows the cumulative percentage of cells in the study area. The nested bar is the stratum bar overlaid on top of the substratum bar to show agreement at both the stratum and substratum levels. Table 17.2 gives the numerical values for the components in the stratum and substratum bars.

top of the substratum bar to produce the nested bar. Depending on the nature of the maps, the nested bar could show nine possible components listed in the legend. In the comparison of REF and COM1, the bar shows eight nested components.

17.2.6 Multiple Resolutions

Up to this point, our analysis of the maps of Figure 17.2 has been based on a cell-by-cell analysis with hard classification. The advantage of cell-by-cell analysis with hard classification is its simplicity. The disadvantage of cell-by-cell analysis with hard classification is that if a specific cell fails to have the correct category, then it is counted as complete error, even when the correct category is found in a neighboring cell. Therefore, cell-by-cell analysis can fail to indicate general agreement of pattern because it fails to consider spatial proximity to agreement. In order to remedy this problem we perform multiple resolution analysis.

The multiple resolution analysis requires a new set of maps for each new resolution. Figure 17.2 shows maps that are hard classified, whereas Figure 17.5 shows the COM1 map at four coarser resolutions. Each cell of each map of Figure 17.5 is an average of neighboring cells of the original COM1 map of Figure 17.2. For example, for resolution 2, four neighboring cells become a single coarse cell; therefore, the 12×12 map of original cells yields a 6×6 map. At resolution three, we obtain a 4×4 map of coarse cells, in which the length of the side of each coarse cell is three times the length of the side of each original fine-resolution cell. At resolution four, we obtain a 2×2 map of coarse cells, where each coarse cell is its own substratum. At resolution 12, the entire map is in one cell. For each coarse cell, the membership in each category is the average of the memberships of the contributing cells. When using this aggregation technique, the lack of a square study area can result in an unequal number of fine-resolution cells in each of the coarser cells. This

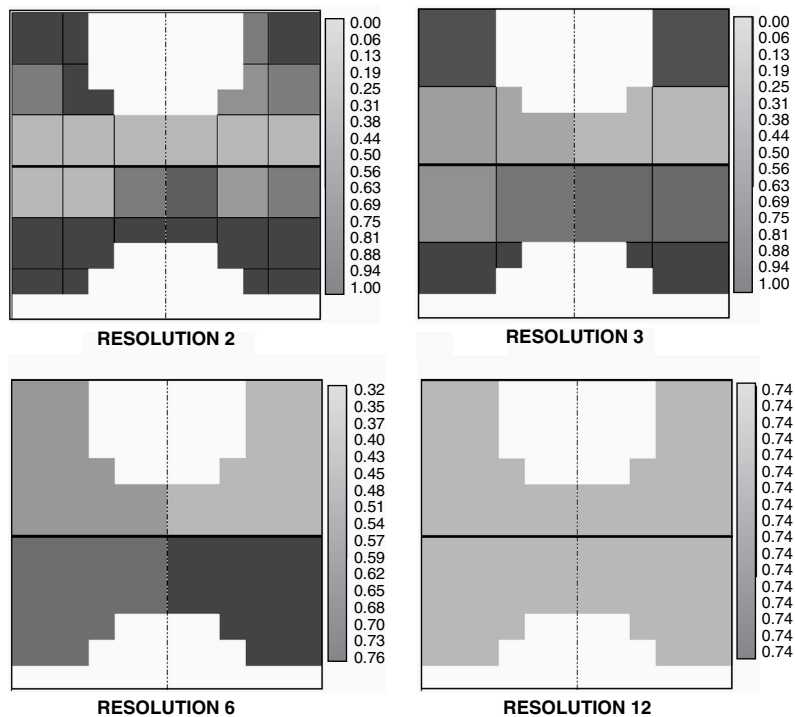


Figure 17.5 Map COM1 at four different resolutions. On the legend, 0 means completely forest, 1 means completely nonforest, and white is outside of the study area.

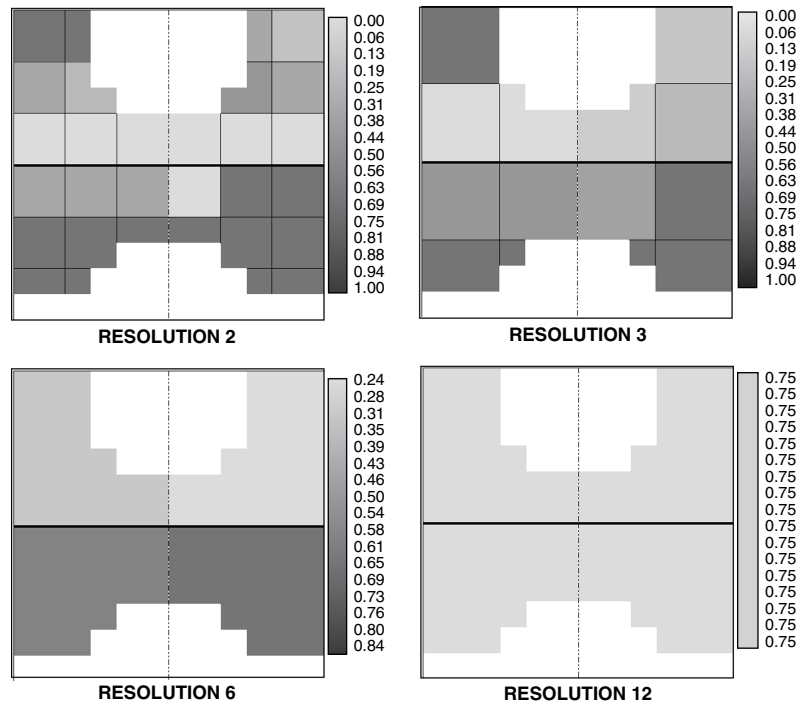


Figure 17.6 Map REF at four different resolutions. On the legend, 0 means completely forest, 1 means completely nonforest, and white is outside of the study area.

is taken into consideration by the weights that give each cell's membership in the study area. This characteristic of the technique allows the method to apply to accuracy assessment where the grid cells of interest are not contiguous.

Figure 17.5 shows the cell configuration. The darker shading shows stronger membership in the nonforested category. Figure 17.6 shows this same type of aggregation for the REF map. For each resolution, we are able to generate a bar similar to the nested bar of Figure 17.4, because the equations of Figure 17.3 allow for any cell to have partial membership in any category.

17.3 RESULTS

Figure 17.7 shows the components of agreement and disagreement between REF and COM1 at all resolutions. Figure 17.8 shows analogous results for the comparison between REF and COM2. The overall proportion correct is the top of the component of agreement at the grid cell level and the overall proportion correct at the coarser resolutions is the top of the component of agreement due to quantity. Proportion correct tends to rise as resolution becomes coarser; however, the rise is not monotonic. Proportion correct rises for each resolution that is nested within finer resolutions. That is, the proportion correct for resolution 1 < proportion correct for resolution 2 < proportion correct for resolution 6 < proportion correct for resolution 12. In addition, the proportion correct for resolution 1 < proportion correct for resolution 3 < proportion correct for resolution 6 < proportion correct for resolution 12. However, the proportion correct for resolution 2 > proportion correct for resolution 3. Note that resolution 2 is not nested within resolution 3.

The largest component is agreement due to chance, which is 50% at the finest resolution since there are two categories. Agreement due to chance rises as resolution becomes coarser. Besides the component due to chance, the largest components at the finest resolution are agreement at the grid

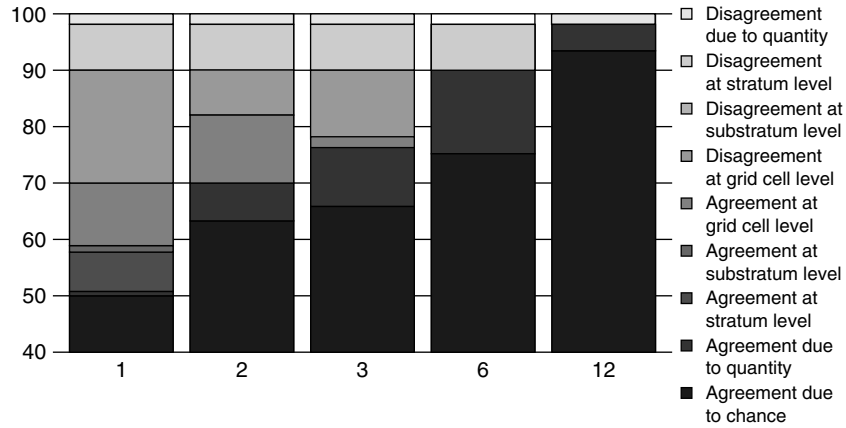


Figure 17.7 Stacked bars showing agreement between COM1 and REF. The vertical axis shows the cumulative percentage of the total study area. The numbers on the horizontal axis give the resolutions.

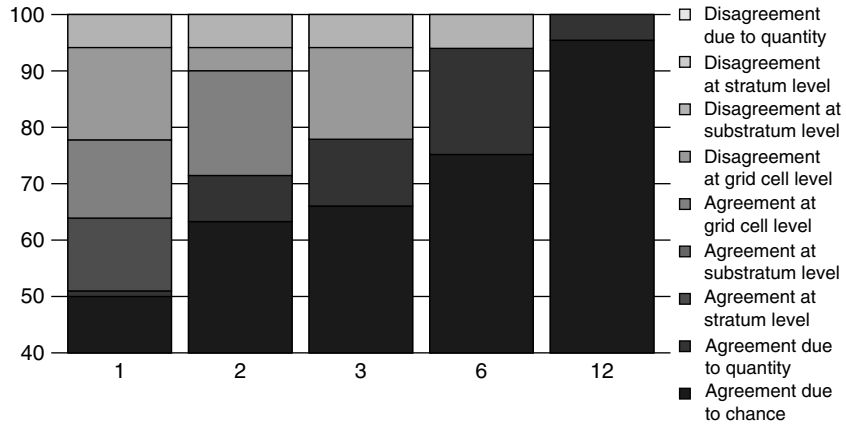


Figure 17.8 Stacked bars showing agreement between COM2 and REF. The vertical axis shows the cumulative percentage of the total study area. The numbers on the horizontal axis give the resolutions.

cell level and disagreement at the grid cell level. As resolution becomes coarser, the grid cell level information becomes less important, relative to information of quantity. At the coarsest resolution, where the entire study area is in one cell, the concept of location has no meaning; hence, the only components are agreement due to chance, agreement due to quantity, and disagreement due to quantity. COM1 has a component of disagreement due to quantity, which does not change as resolution changes, since quantity is a concept independent of resolution. COM2 has no disagreement in quantity.

Figure 17.9 shows that at a fine resolution the agreement between COM2 and REF is greater than the agreement between COM1 and REF. The components that account for the greater agreement are the agreement at the stratum level and at the grid cell level.

Table 17.3 and Table 17.4 display contingency tables that show the nested stratification structure of strata and substrata. These tables are another helpful way to present results. The information on the diagonal indicates the number of cells for each substratum that are in agreement. Therefore, the number of correct cells may be calculated for each substratum by summing the diagonal for each subset of the table. Furthermore, the row and column totals indicate stratum-level agreement. For example, Table 17.3 shows disagreement at the stratum level, since there are 31 forested cells in COM1 vs. 35 in REF for the north stratum and there are 16 forested cells in COM1 vs. 10 in

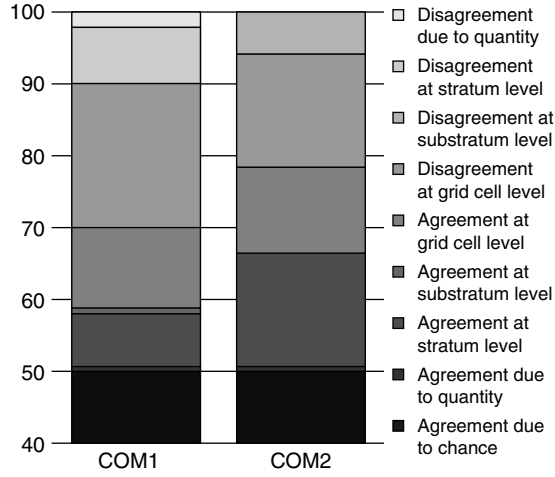


Figure 17.9 Stacked bars showing comparison of COM1 and COM2 with the REF map at the finest resolution. The vertical axis shows the cumulative percentage of the total study area.

Table 17.3 Confusion Matrix for COM1 vs. REF by Strata and Substrata; F Denotes Forest Cells and N Denotes Nonforest Cells

		REF																	
		Substratum																	
		NW		NE		SW		SE		Total									
		F	N	F	N	F	N	F	N										
COM1	Substratum	NW	F	13	1							14	31	F	North	Stratum			
			N	3	8							11							
	NE	F			12	5					17	19	N						
		N			7	1					8								
	SW	F					5	5			10	16	F	South					
		N					1	14			15								
	SE	F							1	5	6	34	N						
		N							3	16	19								
		Total		16	9	19	6	6	19	4	21	100							
				35		15		10		40									
			F		N		F		N										
			North				South												
			Stratum																

Table 17.4 Confusion Matrix for COM2 vs. REF by Strata and Substrata; F Denotes Forest Cells and N Denotes Nonforest Cells

				REF													
				Substratum													
				NW		NE		SW		SE		Total					
				F	N	F	N	F	N	F	N						
COM2	Substratum	NW	F	12	7								19	35	F	North	Stratum
			N	4	2							6					
		NE	F			12	4						16	15	N		
			N			7	2					9					
		SW	F					6	0			6	10	F	South		
			N					0	19			19					
		SE	F							4	0	4	40	N			
			N							0	21	21					
	Total			16	9	19	6	6	19	4	21	100					
				35		15		10		40							
				F		N		F		N							
				North				South									
				Stratum													

REF for the south stratum. This disagreement at the stratum level is reflected in the component of disagreement at the stratum level in Figure 17.7. In contrast, Table 17.4 shows perfect agreement at the stratum level; hence, Figure 17.8 shows no component of disagreement between COM2 and REF at the stratum level.

17.4 DISCUSSION

17.4.1 Common Applications

The three maps in Figure 17.2 represent a common situation in map comparison analysis. There are many applications where a scientist wants to know which of two maps is more similar to a reference map. Three likely applications are in remote sensing, simulation modeling, and land-change analysis.

In remote sensing, when a scientist develops a new classification rule, the scientist needs to compare the map generated by the new rule to the map generated by a standard rule. Two fundamental questions are (1) Did the new method perform better than the standard method concerning its estimate of the quantity of each category? and (2) Did the new method perform better than the standard method concerning its specification of the location of each category? The format of Figure 17.9 is an effective way to display the results, because it conveys the answer to both of these questions quickly. Specifically, COM1 makes some error of quantity while COM2

does not. COM2 shows a better specification of location than does COM1. When the analysis is stratified, the scientist can see whether errors of location exist at the stratum level, substratum level, or grid cell level. For example, Figure 17.9 shows that COM1 has errors at the stratum level and not the substratum level, while COM2 has errors at the substratum level and not at the stratum level. In another application from remote sensing, we could examine the influence of a hardening rule, since the techniques work for both hard and soft classification. For example, COM1 could show the soft category membership before a hardening rule is applied, and COM2 could show the hard category membership after the hardening rule is applied. The format of Figure 17.9 would then summarize the influence of the hardening rule.

In simulation modeling, a scientist commonly builds a model to predict how land changes over time (Veldkamp and Lambin, 2001). The scientist performs validation to see how the model performs and to obtain ideas for how to improve the model. When the model is run from T_1 to T_2 , the scientist validates the model by comparing the simulated landscape of T_2 with a reference map T_2 . A null model would predict no change between T_1 and T_2 . In other words, if the scientist had no simulation model, then the best guess at the T_2 map would be the T_1 map. Therefore, to see whether the simulation model is performing better than a null model, the scientist needs to compare (1) the agreement between the T_2 simulation map and the T_2 reference map vs. (2) the agreement between the T_1 map and the T_2 reference map. In this situation, the format of Figure 17.9 is perfectly suited to address this question because the analogy is that COM1 is the T_1 map, COM2 is the T_2 simulation map, and REF is the T_2 reference map.

The methods described here are particularly helpful in this case since land-cover and land-use (LCLU) change models are typically stratified according to political units because data are typically available by political unit and because the process of land change often happens by political unit. For example, land-use activities in Brazil are planned at the regional and household scales, where the household stratification is nested within the regional stratification. Researchers are dedicating substantial effort to collecting data at a relevant scale in order to calibrate and to improve change models. Therefore, it is essential that statistical methods budget the components of agreement and disagreement at relevant scales, because researchers want to collect new data at the scale at which the most uncertainty exists.

In land-change analysis, the scientist wants to know the manner in which land categories change and persist over time. For this application, the methods of this chapter would use COM1 as the T_1 map and REF as the T_2 map. Figure 17.7 would supply a multiple-resolution analysis of LCLU change, where agreement means persistence and disagreement means change. A disagreement in quantity indicates that a category has experienced either a net gain or a net loss. Disagreement at the stratum level means that a loss of a category in one stratum is accompanied by gain in that category in another stratum. Disagreement at the grid cell level means that a loss of a category at one location is accompanied by a gain of that category at another location within the same stratum. Therefore, Figure 17.7 would show at what scales LCLU change occurs.

17.4.2 Quantity Information

We focus primarily on the center column of mathematical expressions of Figure 17.3, because those expressions give the components of agreement. However, the other two columns can be particularly helpful depending on the purpose of map comparison. In the case of remote sensing, guidance is needed to improve the classification rules. For simulation modeling, guidance is required to improve the simulation model's rules. It would be helpful to know the expected improvement if the rule's specification of quantity changes, given a specific level of information of location. The mathematical expressions in the rightmost column of Figure 17.3 show the expected results when the rule specifies the quantity of each category perfectly with respect to the reference quantities. At the other end of the spectrum, the mathematical equations of the leftmost column show the expected results when the rule uses random chance to specify the quantities of each category.

For example, $M(\mathbf{n})$ expresses the agreement that a scientist would expect between the reference map and the other map when the other map is the adjusted comparison map that is scaled to show the quantity in each category as $1/J$. $M(\mathbf{p})$ expresses the agreement that a scientist would expect between the reference map and the other map when the other map is the adjusted comparison map that is scaled to show the quantity in each category as matching perfectly the reference map.

The definitions of $M(\mathbf{n})$ and $M(\mathbf{p})$ in Table 17.2 are slightly different from the definitions of $M(\mathbf{n})$ and $M(\mathbf{p})$ in Pontius (2000, 2002). Table 17.2 gives expressions for $M(\mathbf{n})$ and $M(\mathbf{p})$ that depend on the scaling given by Equation 17.12 through Equation 17.15. The method of scaling simulates the change in quantity spread evenly across the grid cells as one moves from $M(\mathbf{m})$ to $M(\mathbf{p})$ or from $M(\mathbf{m})$ to $M(\mathbf{n})$. In contrast, Pontius (2000, 2002) does not scale the comparison map and does not represent an even spread of the change in quantity across the cells. The methods of Pontius (2000, 2002) define $M(\mathbf{n})$ and $M(\mathbf{p})$ in a manner that makes sense for applications of land-cover change simulation modeling and slightly confounds information of quantity with information of location. Table 17.2 defines $M(\mathbf{n})$ and $M(\mathbf{p})$ in a manner that is appropriate for a wider variety of applications, since it maintains complete separation of information of quantity from information of location.

17.4.3 Stratification and Multiple Resolutions

If we think of grid cells as tiny strata, then the maps of Figure 17.2 show a three-tiered, nested stratification structure. The cells are 100 tiny strata that are nested within the four substrata that are nested within the two broadest strata. The multiple-resolution procedure grows the cells such that at the resolution of 6 the four coarse grid cells constitute the four quadrants of the substrata. Another similarity between strata and cells is that they both can indicate information of location; hence, they both appear on the vertical axis of Figure 17.3.

However, there are three major conceptual differences between grid cells and strata. First, the concept of location within a grid cell does not exist because category membership within a grid cell is completely homogenous. By definition, we cannot say that a particular category is concentrated at a particular location within a cell. In contrast, the concept of location within a stratum does exist because we can say that a particular category is concentrated at a particular location within a stratum, since strata usually contain numerous cells. Second, the multiple-resolution procedure increases the lengths of the sides of the grid cells and thus reduces the number of coarse grid cells within each stratum, but the multiple-resolution analysis does not change the number of strata. Third, each cell is a square patch, whereas a stratum can be nonsquare and noncontiguous. As a consequence of these differences, analysis of multiple resolutions of cells shows how the landscape is organized in geographic space, whereas analysis of multiple strata shows how the landscape is organized with respect to the strata definitions.

17.5 CONCLUSIONS

The profession of accuracy assessment is advancing past the point where assessment consists of only a calculation of percentage correct or Kappa index of agreement (Foody, 2002). Now, measures of agreement are needed that indicate how to create more-accurate maps. Here we presented novel methods of accuracy assessment to budget the components of agreement and disagreement between any two maps that show a categorical variable. The techniques incorporate stratification, examine multiple resolutions, apply to both hard and soft classifications, and compare maps in terms of quantity and location. Perhaps most importantly, this chapter shows how to present the results of a complex analysis in a simple graphical form. We hope that this technique of accuracy assessment will soon become as common as today's use of percentage correct.

17.6 SUMMARY

This chapter presented novel methods of accuracy assessment to budget the components of agreement and disagreement between a reference map and a comparison map, where each map shows a categorical variable. The measurements of agreement can take into consideration soft classification and can analyze multiple resolutions. Ultimately, the techniques express the agreement between any two maps in terms of various components that sum to 1. The components may be agreement due to chance, agreement due to quantity, agreement due to location at one of the stratified levels, agreement due to location at the grid cell level, disagreement due to location at the grid cell level, disagreement due to location at one of the stratified levels, and/or disagreement due to quantity. These techniques can be used to compute components of agreement at all resolutions and to present the results of a complex analysis in a simple graphical form.

ACKNOWLEDGMENTS

We thank Clark University's Master of Arts program in Geographic Information Science for Development and Environment and the people of Clark Labs, especially Hao Chen, who has programmed these methods into the GIS software Idrisi32®.* We also thank the Center for Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University with which this work is increasingly tied intellectually and programmatically through the George Perkins Marsh Institute of Clark University.

REFERENCES

- Congalton, R., A review of assessing the accuracy of classification of remotely sensed data, *Remote Sens. Environ.*, 37, 35–46, 1991.
- Congalton, R. and K. Green, *Assessing the Accuracy of Classification of Remotely Sensed Data: Principles and Practices*, Lewis, Boca Raton, FL, 1999.
- Foody, G., On the comparison of chance agreement in image classification accuracy assessment, *Photogram. Eng. Remote Sens.*, 58, 1459–1460, 1992.
- Foody, G., Status of land cover classification accuracy assessment, *Remote Sens. Environ.*, 80, 185–201, 2002.
- Lewis, H. and M. Brown, A generalized confusion matrix for assessing area estimates from remotely sensed data, *Int. J. Remote Sens.*, 22, 3223–3235, 2001.
- Pontius, R., Quantification error versus location error in comparison of categorical maps, *Photogram. Eng. Remote Sens.*, 66, 1011–1016, 2000.
- Pontius, R., Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions, *Photogram. Eng. Remote Sens.*, 68, 1041–1049, 2002.
- Veldkamp, A. and E. Lambin, Predicting land-use change, *Agric. Ecosyst. Environ.*, 85, 1–6, 2001.

* Registered Trademark of Clark Labs, Worcester, Massachusetts.

