J Geograph Syst (2006) 8: 119–130
DOI 10.1007/s10109-006-0023-z

ORIGINAL ARTICLE

Giles M. Foody

What is the difference between two maps? A remote senser's view

Published online: 12 May 2006

Abstract In remote sensing, thematic map comparison is often undertaken on a per-pixel basis and based upon measures of classification agreement. Here, the degree of agreement between two thematic maps, and so the difference between the pair, was evaluated through visual and quantitative analyses for two scenarios. Quantitative assessments were based on basic site-specific measures of agreement that are used widely in accuracy assessment (e.g. the overall percentage of pixels with the same class label in each of the two maps and the kappa coefficient of agreement) as well as an information theory based approach that allows the degree of mutual or shared information to be assessed even if different classification schemes have been used to produce the maps. The results indicated that in the first map comparison scenario, focused on labelling, there was a fair degree of correspondence between the maps but with an overall difference in information content of ~42%. In the second comparison scenario, focused on change in time, considerable change had occurred with a change in class label for ~42% of the pixels. It was also apparent that global assessments masked local scale changes.

Keywords Accuracy · Agreement · Mutual information · Classification · Kappa coefficient · Confusion matrix

G. M. Foody
School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ UK
E-mail: g.m.foody@soton.ac.uk

1 Introduction

Map comparison underlies many scientific studies. Determining the difference between two thematic maps is, however, a non-trivial task. As part of an inter-disciplinary project to determine the difference between two sets of thematic maps (F. Csillag and B. Boots 2006, this issue), a series of analyses were undertaken using two sets of paired maps provided by the project organisers. This article reports on these analyses which form one response to the map comparison challenge. It does not aim to provide a comprehensive review of the subject but gives one viewpoint that addresses the projects' core aims, essentially to determine what differences existed between each pair of maps and indicate how these differences could be quantified.

The maps provided in this project were assumed to show the spatial distribution of nominal level thematic classes such as land cover (e.g. forest, grass, water, *etc*). In the absence of a specific pre-stated motivation for the map comparison and without details on the nature of the classes, the approach adopted was to focus on general issues from the viewpoint of a remote senser interested in both map production and use. Many approaches for map comparison have been used in remote sensing (e.g. Congalton and Green 1999; Pontius 2002; Foody 2004; Giri et al. 2004; Wulder et al. 2004; Fritz and See 2006, in press). Given the general perspective adopted, the focus here was on the use of well-established techniques for overall map level comparison and, to a lesser extent, class-specific comparison, as is common in remote sensing. Other issues such as the spatial pattern and dynamics of land cover patches within the regions mapped are addressed in other contributions to this project. Throughout this paper, the difference between the maps is taken to relate to disagreements in the pixel class labels.

Two sets of thematic maps were provided for the study and are described in Csillag and Boots (2006, this issue). Briefly, one provided a scenario in which two labeling schemes were used to represent a region while the other represented a situation in which a single labeling approach was applied to map a region at different times. In each case, the pair of maps was compared visually and quantitatively. Emphasis was placed on site-specific approaches that are widely used in standard accuracy assessments such as the kappa coefficient of agreement as well as measures of mutual information content.

2 Comparison of labeling process

The two maps (D1a, D1b) were compared on the assumption that they had been derived from the same remotely sensed data set through the application of different classifiers. The maps may, for example, have been derived from unsupervised classifiers directed to output a different number of classes.

An initial visual comparison of the two maps indicated that there was a large degree of correspondence between them. The broad general structure of the two representations was similar visually. The maps, however, clearly differed in detail. The most obvious difference was that D1a contained more classes than D1b and that, while there appeared to be some degree of association between the classes depicted in the maps, the relationship between map labels was not a simple one. Furthermore, it was apparent that D1a was busier/more fragmented than D1b but the two maps appeared to have the same spatial resolution.

To evaluate the level of correspondence between the classes, the labels for each pixel mapped were cross-tabulated. For this, the classes in D1a were labeled A–G and those in D1b labeled I–V. The cross-tabulated data are shown in the contingency table in Table 1.

Table 1 helped in the assessment of the degree of correspondence between the class labels in the two maps and hence the degree of difference that exists between them. It shows, for example, that class B in D1a corresponded very closely to class II in D1b, with 99.4% (1,540 out of 1,549) of pixels of class II lying in class B and 85.1% (1,540 out of 1,810) of cases of class B lying in class II. It would seem that the regions represented by class II and class B were of a rare but spectrally distinctive class, which was clearly discriminated from other classes in the area mapped. Additionally, Table 1 shows that cases of class A in D1a lie typically within classes I and V on map D1b. There was also quite a large level of agreement between class C on D1a and class III in D1b. Finally, it was apparent there was not a general simple one-to-one relationship between the classes, this was especially apparent for cases associated with classes IV and V in D1b, which were associated, to varying degrees, with every class in depicted in D1a.

The use of a common classification scheme would facilitate the comparison of the maps, although it is important that the class labels are consistent in their meaning in the maps (Comber et al. 2004a, b). It is sometimes possible to define a function to convert the labels used in one map to the scheme used in the other (Pontius and Malizia 2004; Fritz and See 2006, in press). Assuming this to be appropriate for the maps provided and given that there was some correspondence between the class labels (Table 1) it may be possible to derive a set of directly comparable classes. One approach to achieve the latter might be to collapse the set of classes in each map to yield directly comparable classification schemes. For example, map D1a may be reduced from seven to four classes by combining classes D, E, F and G into one class (*D*) and map D1b reduced from five to four classes by combing IV and V into a single class (*IV*). The cross-tabulation resulting from this can be treated as a standard confusion matrix (Table 2), as the classes are now assumed to be the same (i.e. I ≡ A etc.) and the correspondence of the two

Table 1 Contingency table of the cross-tabulated labels in D1a and D1b

D1a($\rightarrow$)	D1b ($\downarrow$)					Total
	I	II	III	IV	V	
A	1,791	6	90	98	3,161	5,146
B	0	1,540	37	110	123	1,810
C	0	3	11,051	2,180	1,647	14,881
D	0	0	165	9,463	4,774	14,402
E	0	0	37	649	3,446	4,132
F	0	0	60	244	20,355	20,659
G	0	0	11	264	4,231	4,506
Total	1,791	1,549	11,451	13,008	37,737	65,536

Table 2 Cross-tabulation of the collapsed class label set

D1a(→)	D1b (↓)					
	I	II	III	*IV*	Total	Agree (%)
A	1,791	6	90	3,259	5,146	34.80
B	0	1,540	37	233	1,810	85.08
C	0	3	11,051	3,827	14,881	74.26
D	0	0	273	43,426	43,699	99.37
Total	1,791	1,549	11,451	50,745	65,536	
Agree (%)	100	99.42	96.51	85.57		

The overall level of agreement represented by p_o is the sum of the main diagonal divided by the matrix grand total. On a per-class basis, the level of agreement is determined from the relevant element of the main diagonal and marginal read along the row or column of interest (as in the calculation of the User's and Producer's accuracy)

maps can, therefore, be assessed using standard accuracy assessment measures that are commonly used in remote sensing.

Popular approaches for map comparison in remote sensing are based on the measurement of the degree of agreement between the data sets, most commonly expressed as the proportion (or percentage) of cases for which the labels in the maps agree, p_o, or as the kappa coefficient of agreement (Cohen 1960; Congalton et al. 1983). The proportion of cases for which class labels agree is easy to calculate and provides a simple guide to the overall level of difference between two maps; the larger the degree of agreement in the labelling the smaller the difference between the maps. The kappa coefficient is, essentially, a rescaled version of the proportion of cases in agreement and may be calculated from,

$$\hat{\kappa} = \frac{p_o - p_c}{1 - p_c}, \tag{1}$$

where p_c is the proportion of agreement that may be expected to arise by chance. The calculation of the kappa coefficient is based on the main diagonal of the square confusion matrix together with its row and column marginals. If there is perfect agreement between the maps compared, and so no difference between them, the kappa coefficient would equal 1.0. Lower coefficient values indicate a degree of disagreement and a value of 0.0 indicates that any agreement between the maps is due to chance. The statistical significance of derived values may also be tested (Cohen 1960; Couto 2003; Foody 2004). In addition to the overall or global assessment of agreement between the maps provided by p_o or kappa coefficient it is possible to undertake the assessment based on an individual class basis if desired (Congalton and Green 1999).

The overall level of agreement between the two 4-class maps produced by collapsing the original set of classes expressed as the percentage of cases lying in the main diagonal of the matrix was 88.2%. Moreover, the kappa coefficient, which makes an allowance for chance agreement, was 0.733. This kappa coefficient indicated a very high level of agreement above that expected by chance. These results supported the general visual evaluation that the maps show a high degree of correspondence.

Although these analyses yielded basic quantitative assessments of the degree of correspondence, and thereby the level of difference between the maps, the assessment was relatively crude. The level of agreement clearly differed between the classes, an aspect masked in the analyses above. For example, the degree of correspondence between classes B and II was high but most of the disagreement was associated with Class *IV* (Table 2). The comparison was also based upon degraded versions of the original maps, each into four classes to facilitate the evaluation, and not the actual maps of interest. A more appropriate means of comparing the original maps that accommodated for the differences in the classification schemes used was required.

A variety of approaches have been used to compare two or more maps (often in remote sensing applications the comparison is essentially between a derived map and a ground data set). These have typically been undertaken to assess the level of correspondence or agreement between the maps, sometimes with the emphasis on the dissimilarity present (e.g. Li et al. 1993; Townsend 2000; Woodcock and Gopal 2000; Couto 2003). In general, site-specific methods such as the kappa coefficient of agreement discussed above have been used in the comparison of maps (e.g. Monserud and Leemans 1992; Hayes and Sader 2001), although approaches that compare spatial patterns of class distributions have great potential (e.g. Power et al. 2001). Recently, Hagen (2003) introduced an extension to the basic method for map comparison based on the use of the kappa coefficient that also provided a degree of tolerance for mis-registration effects. There are, however, many concerns with the use of the kappa coefficient for evaluating and comparing map agreement (e.g. Stehman 1997; Turk 2002; Foody 1992, 2004) and here the maps were assumed to be perfectly co-registered. A fundamental problem in this exercise was that the standard kappa coefficient, like many other approaches, cannot be derived to compare maps with different classification schemes without making additional assumptions or refinements of the data or technique (Finn 1993). The standard approaches used in remote sensing require the use of the same classification scheme to allow the formation of a square confusion matrix from which measures of agreement may be calculated.

One means of comparing maps with different classification schemes and with different numbers of classes is to evaluate the degree of information they share (Kew 1996). Finn (1993) presents a method to asses the degree of mutual or shared information in thematic maps. Specifically, Finn (1993) introduced the concept of mutual information from information theory as a basis for the evaluation of the degree of map similarity irrespective of the classification schemes used in the maps. The assessment of similarity may be undertaken on an overall map or individual class basis.

An overall assessment of the degree of shared information in the maps may be expressed in terms of the amount of shared information in the maps, represented by the average mutual information. In this regard, the information content of a map is its uncertainty. For a probabilistic system, uncertainty may be expressed in terms of entropy. For map *X* the Shannon entropy may be calculated from:

$$H(X) = -K \sum_{i=1}^{n} p_i \log(p_i), \tag{2}$$

where p_i is the proportion of the mapped area in class i and K is a constant that is often, and here, equal to 1 (Finn 1993). The conditional entropy, which expresses the average information in X given that Y is known may be calculated from:

$$H(X|Y) = H(X,Y) - H(Y). \tag{3}$$

The average mutual information, $I(X;Y)$ between maps X and Y may be defined as the average information about X gained from observing Y and derived from:

$$I(X;Y) = H(X) - H(X|Y). \tag{4}$$

This is the Kullback–Liebler divergence that provides a measure of the distance a joint distribution is from independence. It provides a means to indicate the amount of information shared by the maps, and thereby the difference between them.

The calculation of the information theory based measures can perhaps be most clearly described in the format provided by Finn (1993). The average mutual information (AMI) is:

$$\mathrm{AMI} = K \sum_j \sum_i (p(y_i, x_j)) \log[(p(y_i|x_j))/p(y_i)], \tag{5}$$

where $p(x_j)$ is the proportion of map X in class j, $p(y_i)$ is the proportion of map Y in class i, $p(y_i,x_j)$ is the joint probability of a pixel being class i in map Y and class j in map X and $p(y_i\ x_j)$ is the conditional probability of a pixel in map Y belonging to class i given that it is in class j on map X. The latter is calculated from

$$P(y_i|x_j) = p(y_i, x_j)/p(x_j). \tag{6}$$

The derived AMI expresses the amount of information that one map predicts of the other. As Finn (1993) shows, the AMI may be expressed as a percentage of the uncertainty in a map selected for reference purposes. With map X selected as the reference the %AMI is then:

$$\%\mathrm{AMI} = 100[\mathrm{AMI}/H(X)] \tag{7}$$

which is symmetric, with map X predicting the same amount of information about map Y as map Y does about map X (Finn 1993). The calculation of this type of measure makes no reference to the main diagonal of the confusion matrix and is thereby more applicable than the standard accuracy assessment approaches used widely in remote sensing (Kew 1996). The main value of this approach to the specific map comparison here is that it provides a basis to determine the amount of information shared by the maps, and thereby the difference between them, even though the maps have different classification schemes. Although this is useful in quantifying the difference between two maps it does not identify if the differences are significant. It should also be noted that the approach is sensitive to the presence of zero values in the matrix (Kew 1996).

To indicate the degree information shared by the two maps, the %AMI was calculated. The units of entropy depend on the basis of the logarithm used in the analysis (Couto 2003). Here, the calculations were undertaken using log base 10 and hence the units of entropy are Hartleys (Finn 1993; Couto 2003). Using the data in Table 1, the %AMI for the two maps was estimated as 57.92%, indicating that just over half of the information in each map was shared with the other. Consequently, it may be inferred that the difference between the two maps was 42.08%.

Sometimes it may be useful to focus on specific class(es) in the map. This is simple with the confusion matrix approaches where attention is simply directed to the relevant row or column of the matrix (e.g. Table 2). It is, however, also possible to focus on individual classes with the information theory based approach. On an individual class basis, a posteriori entropies for one map given the class label information from the other may be used to evaluate the amount of information shared by the maps. The a posteriori entropy of X once y_i is known may be calculated from

$$H(X|y_i) = -K\sum_{j=1}^{n} p(x_j|y_i)\log[p(x_j|y_i)]. \tag{8}$$

The difference between $H(X)$ and $H(X|y_i)$ is the change in uncertainty about the class of a pixel in map X. If a decrease in uncertainty occurs it shows that some information in map X was provided through knowledge of y_i. The change in entropy that occurs through knowledge of y_i can be expressed as a percentage of the entropy of map X through

$$\%\mathrm{AE}(X|y_i) = 100[H(X) - H(X|y_i)]/H(X). \tag{9}$$

Similarly, it is possible to determine the values for these variables for map Y when x_j is known (Finn 1993). Using Eqs. 8 and 9, the a posteriori entropy and percentage change were calculated (Tables 3, 4).

For every class in each map, it was apparent that there was a decrease in uncertainty when given the information in the other map (Tables 3, 4). This result highlights that there was a degree of information shared by the two

Table 3 Entropy and change in entropy of map X (D1b) given the class label information in map Y (D1a)

y_i	$H(X\|y_i)$	$H(X) - H(X\|y_i)$	$\%\mathrm{AE}(X\|y_i)$
A	0.821	0.309	27.387
B	0.569	0.561	49.603
C	0.747	0.383	33.854
D	0.696	0.434	38.678
E	0.484	0.646	57.180
F	0.084	1.046	92.576
G	0.239	0.891	78.817

Note $H(X) = 1.1302$

Table 4 Entropy and change in entropy of map Y (D1a) given the class label information in map X (D1b)

x_j	$H(Y\|x_j)$	$H(Y) - H(Y\|x_j)$	$\%AE(Y\|x_j)$
I	0.000	1.690	100
II	0.038	1.652	97.755
III	0.205	1.485	87.895
IV	0.911	0.779	46.096
V	1.421	0.269	15.910

Note $H(Y) = 1.69045$

maps even though they clearly differ in detail. The magnitude of the difference in entropy, perhaps most clearly reflected in the percentage change, differs between classes. For example, if it was known that a pixel was class I in X (map D1b) the a posteriori entropy drops to 0, effectively informing the analyst that the corresponding label in Y (D1a) was A (Table 4). This supported the simple visual interpretation of Table 1, in which it was clear that all pixels allocated to class I in D1b lie within class A on map D1a. Similarly, very low a posteriori entropies were associated with class II (Table 4) and class F (Table 3), indicating correspondence mainly with a single class in the other map; again supported from inspection of Table 1. A very large a posteriori entropy, however, indicates that knowledge of a pixel belonging to a class in one map provides little information on its membership in the other map. This was most apparent for class V in map D1b (Table 4), which was associated, to varying degrees, with every class in map D1a (Table 1). The simple information theory based measures provided by Finn (1993), therefore, provided a means of quantifying the shared information between maps that was evident in Table 1 and confirmed aspects of the visual assessment of the difference between the maps.

3 Comparison of maps over time

A quick visual comparison of maps D2a and D2b indicated that substantial change had occurred. It was assumed that the maps had been perfectly co-registered and so none of the differences were due to spatial mis-registration of the maps. Although the broad structure of the landscape remained over time it was apparent that classes C and D had become more abundant while classes A and B had declined in extent. The degree of change in the north-eastern segment of the area represented, however, seemed much less than in the rest (main part) of the region compared (Fig. 1).

To try and reduce subjectivity in the comparison, some simple quantitative indices were derived. A major concern in remote sensing studies is often on the areal extent of the classes. Therefore, as a starting point, the areal extent of each class was estimated in each image. Rather surprisingly, it was found that in D2a each class was equally abundant, with each represented by 16,384 pixels. In the intervening time to that represented by D2b, however, considerable change had occurred. Class A had greatly reduced in extent and now was represented by 11,588 pixels. Class B had also reduced in

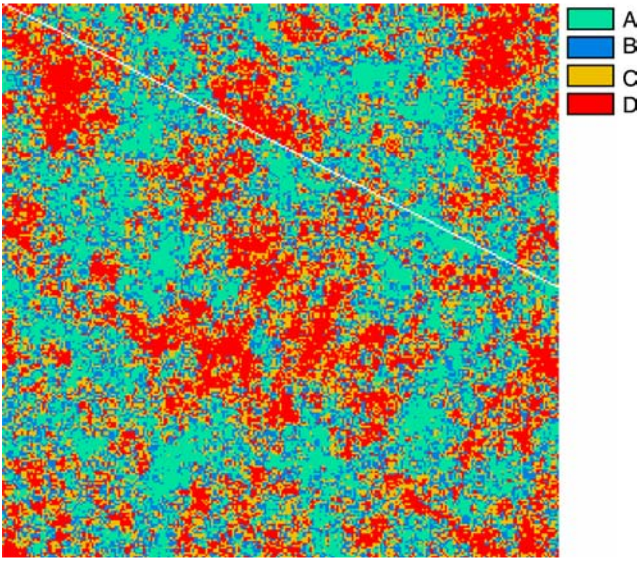

Fig. 1 Map D2a with *white line* separating the main and north eastern segment sub-areas of the region

extent in time but the decline was less severe than with class A with 14,743 pixels present. The remaining classes both expanded in extent. Class C increased slightly to 18,016 and class D substantially to 21,189 pixels, respectively. Thus, over the entire area it was apparent that the greatest changes were associated with classes A and D which, respectively, decreased and increased by ~30%. Classes B and C changed less drastically, with respective decreases and increases in extent of ~10%. Although the concern here was not with spatial pattern, visually it appeared as if the 'clumpy' class D grew outward and took land from the other classes. This was most apparent in the main region of the map where it appeared that most growth of class D was around pre-existing clumps. In the north-eastern segment, however, little expansion was evident with loss, albeit small and relatively randomly distributed, appearing to be more common. Similarly, with class A most change occurred in the main region of the mapped area and appeared as loss of cover around clumps with little change in the north-eastern segment. These interpretations indicate that a study of landscape pattern might be a useful accompaniment to the analyses reported here.

To try and gain a more objective feel for the nature of the change, the class labels observed at the two time periods were cross-tabulated (Table 5). From Table 5 it was apparent that the class label remained unchanged for 37,645 pixels or 57.44% of the mapped area. The overall difference between the maps, without regard to class-specific concerns, was, therefore, 42.56 %. As before, the kappa coefficient, despite its limitations, may be used to indicate the degree of agreement between the two maps and hence the amount of change that had occurred. In this instance the kappa coefficient was estimated to be 0.433, indicating a fair level of agreement between the two maps above that expected by chance. Given that the maps had the same

Table 5 Cross-tabulation of the labels in D2a and D2b

D2a(→)	D2b (↓)				Total	Agree (%)
	A	B	C	D		
A	8,488	3,057	2,656	2,183	16,384	51.81
B	1,582	8,228	3,605	2,969	16,384	50.22
C	967	2,355	8,977	4,085	16,384	54.79
D	551	1,103	2,778	11,952	16,384	72.95
Total	11,588	14,743	18,016	21,189	65,536	
Agree (%)	73.25	55.81	49.83	56.41		

classes and could be directly compared, these proportion of cases in agreement and kappa coefficient values describe the degree of agreement and so difference between the maps. Although the information theory based measures could be calculated for this map comparison this analysis was not undertaken as their main value is apparent for the situation where the classification schemes differ.

From Table 5, it was apparent, however, that the amount of change differed between the classes. The most stable class was class D (11,952 pixels were labelled class D in both maps), with the other classes maintaining cover from one time period to another for a broadly similar area (~8,500 pixels). It was also apparent that there was no simple one-to-one trend in time, with, for example, one class expanding into the area previously covered by another. With all of the matrix elements occupied by large numbers it was apparent that there were gains and losses of cover for each class associated with every other class.

These global summaries of the extent of class coverage above, however, ignore local variation in change over time. From the visual assessment it was apparent that the north eastern part of the region mapped was relatively more stable than the rest of the region. Focusing on these two regions (defined in Fig. 1), it was evident that the global summaries of change masked major local differences. For example, in the north eastern segment of the map, there was much less change than in the rest of the mapped area. The gross and net change associated for each class was very small in comparison to the changes observed in the remaining part of the mapped area. Moreover, the direction of change in the north eastern segment also sometimes differed from that observed in the rest of the mapped region. For example, in the north eastern segment it was evident that for classes A, B and C there was a very small net gain in extent while class D, which globally increased substantially, actually declined by a very small amount (Table 6). Conversely, in the main region, classes A and B declined while classes C and D increased in extent, the net changes also being of considerably larger magnitude than in the north eastern portion.

4 Conclusions

The difference between two maps was assessed through visual interpretation and basic quantitative analyses. Simple visual analyses revealed general

Table 6 Summary of the changes observed over the entire mapped region (global) and the two sub-regions (main and north-eastern segment)

Class	Global				Main				North-east segment			
	Ab/Ab	Loss	Gain	Pr/Pr	Ab/Ab	Loss	Gain	Pr/Pr	Ab/Ab	Loss	Gain	Pr/Pr
A	0.703	0.120	0.047	0.129	0.708	0.146	0.046	0.098	0.686	0.042	0.050	0.220
B	0.651	0.124	0.099	0.125	0.636	0.150	0.112	0.100	0.691	0.048	0.062	0.198
C	0.612	0.113	0.138	0.137	0.580	0.136	0.166	0.116	0.704	0.043	0.054	0.197
D	0.609	0.068	0.141	0.182	0.565	0.074	0.183	0.175	0.735	0.046	0.015	0.202

Data shown are the proportions of the region where the class was absent at each time period (Ab/Ab) and present at each time period (Pr/Pr) as well as that where there was a gain or loss in coverage

trends and could accommodate the difference in classification schemes that occurred in the first comparison. It was possible to quantify the differences between the maps. This quantification was based mainly on cross-tabulations of map labels from which standard measures of classification agreement such as the proportion of cases in agreement or the kappa coefficient could be used to infer the degree of difference. To use these measures, however, the maps need to show the same classes to allow the formation of a square confusion matrix. It may be possible to adjust the mapped information to meet this requirement. If this is inappropriate, however, an alternative approach such as that based on AMI content may be used. It has, however, been demonstrated that the difference between two maps may be quantified. Moreover, this assessment can be undertaken on an overall and, if desired, per-class basis although the focus here was on general map level changes. The main observations arising from the two sets of map comparisons were:

1. Difference in labeling: the maps differed, notably in number of classes, but visually represented the same general structure. Although the maps shared a large amount of information ~42% was different and the relationship between the labels depicted on the maps differed between the classes.
2. Difference over time: substantial change had occurred, with a change in class label noted for ~42% of the pixels. The global summary of change in time masked local detail as the mapped region comprised two sub-areas in which the changes differed in magnitude and sometimes direction. These general overviews of the differences could be usefully supported by further analyses, notably addressing issues of spatial pattern in the assessment of change over time.

Acknowledgements I am grateful to Ferko Csillag and Barry Boots for organising the project and the referees for their constructive comments.

References

Cohen J (1960) A coefficient of agreement for nominal scales. Edu Psychol Meas 20:37–46

Comber A, Fisher P, Wadsworth R (2004a) Integrating land-cover data with different ontologies: identifying change from inconsistency. Int J Geogr Inf Sci 18:691–708

Comber A, Fisher P, Wadsworth R (2004b) Assessment of a semantic statistical approach to detecting land cover change using inconsistent data sets. Photogramm Eng Remote Sens 70:931–938
Congalton RG, Green K (1999) Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, Boca Raton
Congalton RG, Oderwald RG, Mead RA (1983) Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogramm Eng Remote Sens 49:1671–1678
Couto P (2003) Assessing the accuracy of spatial simulation models. Ecol Model 167:181–198
Csillag F, Boots B (2006) Categorical maps, comparisons, and confidence: results from a virtual workshop. What is the difference between two maps? A remote senser's view (this issue)
Finn JT (1993) Use of average mutual information index in evaluating classification error and consistency. Int J Geogr Inf Syst 7:349–366
Foody GM (1992) On the compensation for chance agreement in image classification accuracy assessment. Photogramm Eng Remote Sens 58:1459–1460
Foody GM (2004) Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogramm Eng Remote Sens 70:627–633
Fritz S, See L (2006) Comparison of land cover maps using fuzzy agreement. Int J Geogr Inf Sci (in press)
Giri C, Zhu Z, Reed B (2004) A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. Remote Sens Environ 94:123–132
Hagen A (2003) Fuzzy set approach to assessing similarity of categorical maps. Int J Geogr Inf Sci 17:235–249
Hayes DJ, Sader SA (2001) Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. Photogramm Eng Remote Sens 67:1067–1075
Kew NR (1996) Information-theoretic measures for assessment and analysis in image classification. In: Binaghi E, Brivio PA, Rampini A (eds) Soft computing in remote sensing data analysis. World Scientific, Singapore, pp 173–180
Li B, Chu Y, Loh DK (1993) Event probability correlation analysis for comparison of two-phase ecological maps. Ecol Model 69:287–302
Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. Ecol Model 62:275–293
Pontius RG (2002) Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogramm Eng Remote Sens 68:1041–1049
Pontius RG, Malizia NR (2004) Effect of category aggregation on map comparison. Lect Notes Comput Sci 3234:251–268
Power C, Simms A, White R (2001) Hierachical fuzzy pattern matching for the regional comparison of land use maps. Int J Geogr Inf Sci 15:77–100
Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 62:77–89
Townsend PA (2000) A quantitative fuzzy approach to assess mapped vegetation classifications for ecological applications. Remote Sens Environ 72:253–267
Turk G (2002) Map evaluation and "chance correction". Photogramm Eng Remote Sens 68:123
Woodcock CE, Gopal S (2000) Fuzzy set theory and thematic maps: accuracy assessment and area estimation. Int J Geogr Inf Sci 14:153–172
Wulder MA, Boots B, Seemann D, White JC (2004) Map comparison using spatial autocorrelation: an example using AVHRR derived land cover of Canada. Can J Remote Sens 30:573–592