

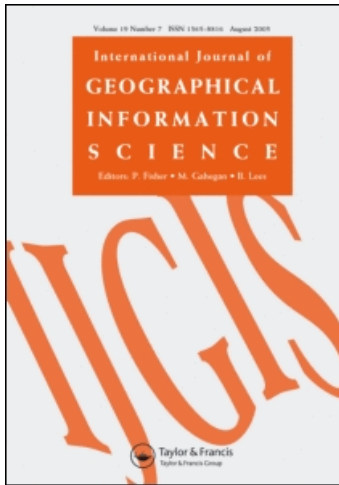
This article was downloaded by: [University of Colorado, Boulder campus]

On: 18 July 2009

Access details: Access Details: [subscription number 785022308]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t71359799>

Fuzzy set approach to assessing similarity of categorical maps

Alex Hagen^a

^a Research Institute for Knowledge Systems, P.O. Box 463, 6200 AL Maastricht, The Netherlands; e-mail: ahagen@riks.nl.

Online Publication Date: 01 April 2003

To cite this Article Hagen, Alex(2003)'Fuzzy set approach to assessing similarity of categorical maps',International Journal of Geographical Information Science,17:3,235 — 249

To link to this Article: DOI: 10.1080/13658810210157822

URL: <http://dx.doi.org/10.1080/13658810210157822>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Research Article

Fuzzy set approach to assessing similarity of categorical maps

ALEX HAGEN

Research Institute for Knowledge Systems, P.O. Box 463, 6200 AL Maastricht,
The Netherlands; e-mail: ahagen@riks.nl

(Received 18 October 2001; accepted 13 May 2002)

Abstract. For the evaluation of results from remote sensing and high-resolution spatial models it is often necessary to assess the similarity of sets of maps. This paper describes a method to compare raster maps of categorical data. The method applies fuzzy set theory and involves both fuzziness of location and fuzziness of category. The fuzzy comparison yields a map, which specifies for each cell the degree of similarity on a scale of 0 to 1. Besides this spatial assessment of similarity also an overall value for similarity is derived. This statistic corrects the cell-average similarity value for the expected similarity. It can be considered the fuzzy equivalent of the Kappa statistic and is therefore called K_{Fuzzy} . A hypothetical case demonstrates how the comparison method distinguishes minor changes and fluctuations within patterns from major changes. Finally, a practical case illustrates how the method can be useful in a validation process.

1. Introduction

With the growth of high-resolution spatial modelling, geographical information systems and remote sensing the need for map comparison methods increases. Good comparison methods are needed to perform calibration and validation of spatial results in a structured and controllable manner. The importance of map comparison methods is recognized and has growing interest among researchers (Monserud and Leemans 1992, Metternicht 1999, Winter 2000, Pontius 2000, Pontius and Schneider 2001, Power *et al.* 2001).

For most purposes visual, human comparison still outperforms automated procedures. When comparing maps the human observer takes many aspects into consideration without deliberately trying. Local similarities, but also global similarities, logical coherence, patterns etc. are recognized. Map comparison methods performed by software usually capture one of these aspects, but overlook the others. Furthermore, they generally lack the flexibility to switch from one aspect to the other when the data requires it. The best example of this rigidity is the cell-by-cell comparison of two checkerboards; the first board has a white field in the upper left corner, the second a black field. The average observer would immediately recognize the two boards as being highly similar in quality, however a cell-by-cell comparison method would find a black cell where a white one is expected and vice versa. Hence total disagreement would be concluded.

Despite these clear disadvantages, there are situations where automated map comparison is preferred above visual comparison. One reason is that an automated procedure can save time and human effort. More important is that automated procedures are explicitly defined and therefore repeatable. Thus, the method can be analysed and evaluated and the results can be verified. A visual comparison will always be subjective and often intuitive. The outcome of a visual comparison will therefore depend on the person performing the comparison.

The comparison method presented here, was primarily developed to be of use in the calibration and validation process of cellular models for land-use dynamics. The method is based on fuzzy set theory (Bandemer and Gottwald 1995, Zadeh 1965). Several authors addressed the potential of fuzzy set theory for geographical applications (Cheng *et al.* 2001, Fisher 2000) and fuzzy set theory has been used before to assess the accuracy of map representations and for map comparisons (Metternicht 1999, Lewis and Brown 2001, Power *et al.* 2001).

The subject of map comparison is closely related to accuracy assessment of maps, in the sense that accuracy assessment is one of its applications. Foody (2002) presents an overview of the status of land cover classification accuracy assessment. Several issues that are brought to attention in that overview are, at least partly addressed in this paper. Foody (2002) asks: 'Why cannot some level of positional tolerance be more generally incorporated into thematic map accuracy assessment'. Also, it is stressed that 'spatial variability of error can be a major concern'. Finally Foody (2002) states that there is 'scope for considerable research' on the topic of fuzzy classifications in accuracy assessment.

The objective is to find a method that to some extent mimics the human comparison and gives a detailed assessment of similarity. The method is aimed at comparing categorical raster maps. The assessment results are spatial and gradual; additionally an overall figure for similarity is aggregated from the detailed spatial results.

2. Methods

For the comparison of maps, two sources of fuzziness are considered: fuzziness of location and fuzziness of category. A similar distinction is found in (Cheng *et al.* 2000), where thematic and geometric aspects of uncertainty are treated separately. In this paper, fuzziness means a level of uncertainty and vagueness of a map. This fuzziness is not inherently present in the map, but follows from an observer's interpretation. Fuzziness of category means the observation that some categories in the legend of a map are more similar to each other than others. With fuzziness of location is meant that the spatial specification found in a categorical map is not always as precise as appears; a category that in the map is positioned at a specific location may be interpreted as being present somewhere in the proximity of that location.

In the original map every cell is represented by a single category. In the fuzzy representation a cell will partially belong to multiple categories. To allow cells to belong to multiple categories simultaneously they are assigned a membership vector. The elements of the vector give the degree of belonging to each category. In this paper three types of membership vectors will be distinguished the Crisp Vector (V_{crisp}) the Fuzzy Category Vector (V_{cat}) and the Fuzzy Neighbourhood Vector (V_{nbh}). The Crisp Vector does not involve fuzziness at all. The Fuzzy Category Vector represents a cell when only fuzziness of category is considered. Finally, the

Fuzzy Neighbourhood Vector represents a cell considering fuzziness of both category and location.

Equation (1) gives the general form of the Crisp Vector, its membership values are set according to equation (2). It signifies that in the Crisp Vector representation of a cell has a degree of membership of 1 for its original category and 0 for all other categories. Table 1 gives examples of crispvectors at four different locations each in different categories.

$$\mathbf{V}_{crisp} = \begin{pmatrix} \mu_{crisp,1} \\ \mu_{crisp,2} \\ \vdots \\ \mu_{crisp,C} \end{pmatrix} \quad (1)$$

$$\text{Original category } i \rightarrow \mu_{crisp,i} = 1, \mu_{crisp,j} = 0, (i \neq j) \quad (2)$$

2.1. Representation of fuzziness of categories

Vagueness may exist in the definition of categories. This is especially true if some or all categories on the map have in fact an ordinal definition, such as for instance the categories 'high-', 'medium-' and 'low-density residential area' on a land use map.

Similarity between categories is expressed in the Fuzzy Category Vector (equation 3), by assigning a higher degree of membership for categories that are more similar to the original category. That means that for the original category it will have a full membership degree of 1. For the other categories the membership will be between 0 and 1 according to level of similarity, as expressed in equation (4).

$$\mathbf{V}_{cat} = \begin{pmatrix} \mu_{cat,1} \\ \mu_{cat,2} \\ \vdots \\ \mu_{cat,C} \end{pmatrix} \quad (3)$$

$$\text{Original category } i \rightarrow \mu_{cat,i} = 1, 0 \leq \mu_{cat,j} \leq 1, (i \neq j) \quad (4)$$

Table 2 demonstrates, for example how the fuzziness of the categories can be expressed in the Fuzzy Category Vector. The meaning of this particular fuzzy representation of categories is that, for instance, 'low density residential' is considered more similar to 'high density residential' than 'industry'. On the other hand 'low density residential' is less similar to 'high density residential' than 'medium density residential'.

Table 1. Crisp Vector representation of four categories.

Category	Original representation	Crisp Vector
Urban area	1	(1, 0, 0, 0)
Undeveloped	2	(0, 1, 0, 0)
Agriculture	3	(0, 0, 1, 0)
Water	4	(0, 0, 0, 1)

Table 2. Fuzzy representation of ordinal data.

Category	Number	Fuzzy Category Vector
High density residential	1	(1 0.4 0.2 0 0 0)
Medium density residential	2	(0.4 1 0.4 0 0 0)
Low density residential	3	(0.2 0.4 1 0 0 0)
Agriculture	4	(0 0 0 1 0 0)
Industry	5	(0 0 0 0 1 0)
Water	6	(0 0 0 0 0 1)

In the previous example it is clear that ‘high-’, ‘medium-’ and ‘low-density residential’ are sub-categories of ‘residential’. Maps will more often contain a mixture of categories and sub-categories. The sub-categories are not always ordinal; they can also be nominal. The difference between categories in the legend that are sub-categories of the same main category is often less distinct than between categories that do not belong to a common group of categories. This can also be expressed in the Fuzzy Category Vector, as is illustrated by an example in table 3.

In the example of table 3 the sub-categories ‘citrus-’, ‘sugarcane-’ and ‘banana agriculture’ are considered more similar to each other than to the other categories, ‘residential’, ‘industry’ and ‘water’.

It should be kept in mind that the fuzzy representation is in reality an interpretation of the original crisp data. There are no straightforward rules for assigning membership values. The definition of the appropriate set depends, for instance, on the nature of the map, the aim of the comparison and the number of categories present.

2.2. Representation of fuzziness of location

Besides fuzziness of category also fuzziness of location is considered. The calculation of fuzziness of location is based upon the notion that the fuzzy representation of a cell depends on the cell itself and, to a lesser extent, also the cells in its neighbourhood. The extent to which the neighbouring cells influence the fuzzy representation is expressed by a distance decay function. For instance a cone (defined by radius), an exponential decay (defined by halving distance) or a 3-D Gausse curve (defined by variance), see figure 1 (Bandemer and Gottwald 1995).

Which function is most appropriate and also the size of the neighbourhood depends on the nature of the uncertainty, vagueness of the data and the observer’s tolerance for spatial error. From a theoretical point of view, there is not a best alternative, hence it is worthwhile to experiment with size and form of the function.

Table 3. Fuzzy representation of hierarchical data.

Category	Number	Fuzzy Category Vector
Residential	1	(1 0 0 0 0 0)
Citrus agriculture	2	(0 1 0.3 0.3 0 0)
Sugarcane agriculture	3	(0 0.3 1 0.3 0 0)
Banana agriculture	4	(0 0.3 0.3 1 0 0)
Industry	5	(0 0 0 0 1 0)
Water	6	(0 0 0 0 0 1)

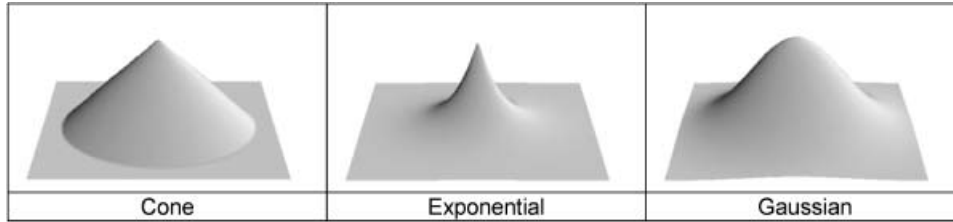


Figure 1. Some 3D memberships.

The different membership contributions of the neighbouring cells are combined by calculating the fuzzy union of all neighbouring cells multiplied by their respective distance based membership. The vector that results from this operation is the Fuzzy Neighbourhood Vector. This is expressed in equations (5) and (6) for a map with C categories and N cells in the neighbourhood. Equation (6) shows how cells in the neighbourhood contribute to the fuzzy representation of the central cell. With increasing distance from the central cell, the contribution decreases, as expressed by the distance based membership m_j . The highest contribution of each category sets the membership value of that category.

$$\mathbf{V}_{nbh} = \begin{pmatrix} \mu_{nbh,1} \\ \mu_{nbh,2} \\ \vdots \\ \mu_{nbh,C} \end{pmatrix} \quad (5)$$

$$\mu_{nbh,i} = |\mu_{nbh,i,1} * m_1, \mu_{cat,i,2} * m_2, \dots, \mu_{cat,i,N} * m_N|_{Max} \quad (6)$$

where F_i = the degree of membership for category i , $\mu_{nbh,i,j}$ = membership of category i for neighbouring cell j in \mathbf{V}_{nbh} , $\mu_{cat,i,j}$ = membership of category i for neighbouring cell j in \mathbf{V}_{cat} , m_j = distance based membership of neighbouring cell j .

Figure 2 and equation (7) illustrate this for a cell in a neighbourhood with a radius of 2 cells. Figure 2 describes the situation. Equation (7) applies equations (5) and (6) for the central cell of the particular situation.

$$\mathbf{V}_{nbh} = \begin{pmatrix} \mu_{nbh,1} = \begin{vmatrix} 1 \times 0.2 & 1 \times 0.5 & 0 \times 0.2 \\ 0 \times 0.5 & 1 \times 1 & 0 \times 0.5 \\ 0 \times 0.2 & 1 \times 0.5 & 0 \times 0.2 \end{vmatrix}_{Max} \\ \mu_{nbh,2} = \begin{vmatrix} 0 \times 0.2 & 0 \times 0.5 & 1 \times 0.2 \\ 0 \times 0.5 & 0 \times 1 & 0 \times 0.5 \\ 0 \times 0.2 & 0 \times 0.5 & 1 \times 0.2 \end{vmatrix}_{Max} \\ \mu_{nbh,3} = \begin{vmatrix} 0 \times 0.2 & 0 \times 0.5 & 0 \times 0.2 \\ 1 \times 0.5 & 0 \times 1 & 1 \times 0.5 \\ 1 \times 0.2 & 0 \times 0.5 & 0 \times 0.2 \end{vmatrix}_{Max} \end{pmatrix} = \begin{pmatrix} 1 \\ 0.2 \\ 0.5 \end{pmatrix} \quad (7)$$

In the example of figure 2, the Fuzzy Category Vector is equal to the Crisp Vector, indicating that similarity between categories has not been considered. The procedure is identical if the Fuzzy Category Vector does express similarity between categories.


Neighbourhood	Legend		Membership definition	
	Categories	Fuzzy Category Vector	Distance	M
	Black	(1, 0, 0)	0	1
	White	(0, 1, 0)	1	0.5
	Grey	(0, 0, 1)	$\sqrt{2}$	0.2

Figure 2. Neighbourhood, legend and membership definition.

2.3. The comparison

2.3.1. Comparison of two fuzzy cells

The similarity of two maps can be assessed by cell-by-cell comparison of the fuzzy vectors assigned to all cells. The expression for similarity at each location is based upon the fuzzy set intersection of the two fuzzy vectors, and is given in equation (8).

$$S(\mathbf{V}_A, \mathbf{V}_B) = [|\mu_{A,1}, \mu_{B,1}|_{Min}, |\mu_{A,2}, \mu_{B,2}|_{Min}, \dots, |\mu_{A,C}, \mu_{B,C}|_{Min}]_{Max} \quad (8)$$

In equation (8), $S(\mathbf{V}_A, \mathbf{V}_B)$ stands for the similarity between a cell in map A and one at the same location in map B . Zadeh (1965) indicates the same expression by the letter M and refers to it as the ‘maximal grade for the intersection $A \cap B$ ’. This similarity index is chosen because it is functional, relatively simple and intuitive. Many other fuzzy similarity measures have been researched and proposed, however, and a better alternative may be found (Zwick *et al.* 1987, Shyi-Ming 1995, Xuzhu 1995, Tolia, Panas and Tsoukalas 2001).

Equation (8) calculates the similarities if the Fuzzy Neighbourhood Vectors of the two central cells found in figure 3. The membership settings and notations are those used before in figure 2.

$$S(\mathbf{V}_{nbh,A}, \mathbf{V}_{nbh,B}) = [|1, 0.5|_{Min}, |0.2, 1|_{Min}, \dots, |0.5, 0.5|_{Min}]_{Max} = 0.5 \quad (9)$$

The value for similarity ranges from 0 to 1. $S(\mathbf{V}_A, \mathbf{V}_B)$ will equal 0 for two completely dissimilar neighbourhoods and 1 for neighbourhoods with matching central cells. The value of 0.5 resulting from the operation is to be interpreted as ‘considerably similar’. It is noted, however, that this similarity value is due to the fact that both central cells neighbour a grey cell. Thus the calculated similarity is based on the neighbours rather than the cells themselves.

2.3.2. Two-way comparison

By directly comparing the fuzzy representations of two cells a part of the comparison result depends on the comparison of the two neighbourhoods, excluding the central cell. The consequence is that even if two cells at the same location in two maps belong to different categories and these two categories are not similar to any



		Central cell A	Central cell B
Map A	Map B	Crisp vector	(1, 0, 0) (0, 1, 0)
		Fuzzy Nbh Vector	(1, 0.2, 0.5) (0.5, 1, 0.5)

Figure 3. Two neighbourhoods and their central cells.

of the categories in the neighbourhood, there is a possibility that the cells are considered similar because their neighbourhoods are similar. This is not intended for the map comparison.

To avoid an overpowering influence of the similarities between the neighbourhoods, the so-called *two-way comparison* is introduced. It proceeds as follows: in first instance the Fuzzy Neighbourhood Vector of cell A is compared to the Crisp Vector of cell B . Next the Crisp Vector of cell A is compared to the Fuzzy Neighbourhood Vector of cell B . Finally, the lower of the two comparison results establishes the similarity at that location (equation 10).

$$S_{TwoWay}(A, B) = |S(\mathbf{V}_{nbh,A}, \mathbf{V}_{Crisp,B}), S(\mathbf{V}_{Crisp,A}, \mathbf{V}_{nbh,B})|_{Min} \quad (10)$$

The calculation of the two-way similarity value of the central cells in figure 3 is calculated according to equations (11–13). A lower similarity of 0.2 is found.

$$S(\mathbf{V}_{nbh,A}, \mathbf{V}_{Crisp,B}) = [|0.5, 1|_{Min}, |1, 0|_{Min}, |0.5, 0|_{Min}]_{Max} = 0.5 \quad (11)$$

$$S(\mathbf{V}_{nbh,A}, \mathbf{V}_{Crisp,B}) = [|1, 0|_{Min}, |0.2, 1|_{Min}, |0.5, 0|_{Min}]_{Max} = 0.2 \quad (12)$$

$$S_{TwoWay}(A, B) = |0.5, 0.2|_{Min} = 0.2 \quad (13)$$

Figure 4 shows six situations to illustrate the preference for the two-way comparison over the direct comparison of Fuzzy Neighbourhood Vectors. For each situation both the similarity according to the direct comparison of the Fuzzy Neighbourhood Vectors and the two-way comparison are given. It demonstrates that only the two-way comparison yields the intended similarity results.

2.4. K_{Fuzzy} statistic for overall map similarity

The previous paragraphs specify how for each cell a local measure of similarity can be calculated. In addition to this, it is for some applications useful to obtain an overall value of similarity. An overall value can be obtained by integrating the similarity values over the whole map. Division by the total area yields a result between 1 (for identical maps) and 0 (for total disagreement). Since regular grid maps are considered, this is equivalent to calculating the average similarity of all cells.

The average similarity, however, is not necessarily a good measure for overall similarity, because the expected value for similarity would be strongly influenced by the number of categories in the map and also by the numerical distribution of cells over those categories. In order to make the results of maps with different numerical distributions more comparable a statistic is introduced that corrects the percentage of agreement for the expected percentage of agreement, based upon the number of cells taken in by each category on each map (i.e. based upon the histograms of the two maps).

The statistic is similar to the Kappa statistic and is therefore called K_{Fuzzy} . The formula for K_{Fuzzy} (equation 14) is identical in form to that of the Kappa statistic (Carletta 1996, Monserud and Leemans 1992). The difference lies in the calculation of the expected similarity.

$$K_{Fuzzy} = \frac{P_o - P_e}{1 - P_e} \quad (14)$$

where P_o = observed percentage of agreement (i.e. average similarity); P_e = expected similarity, based upon given histograms.

In the following paragraphs P_e is derived for two-way comparisons in which

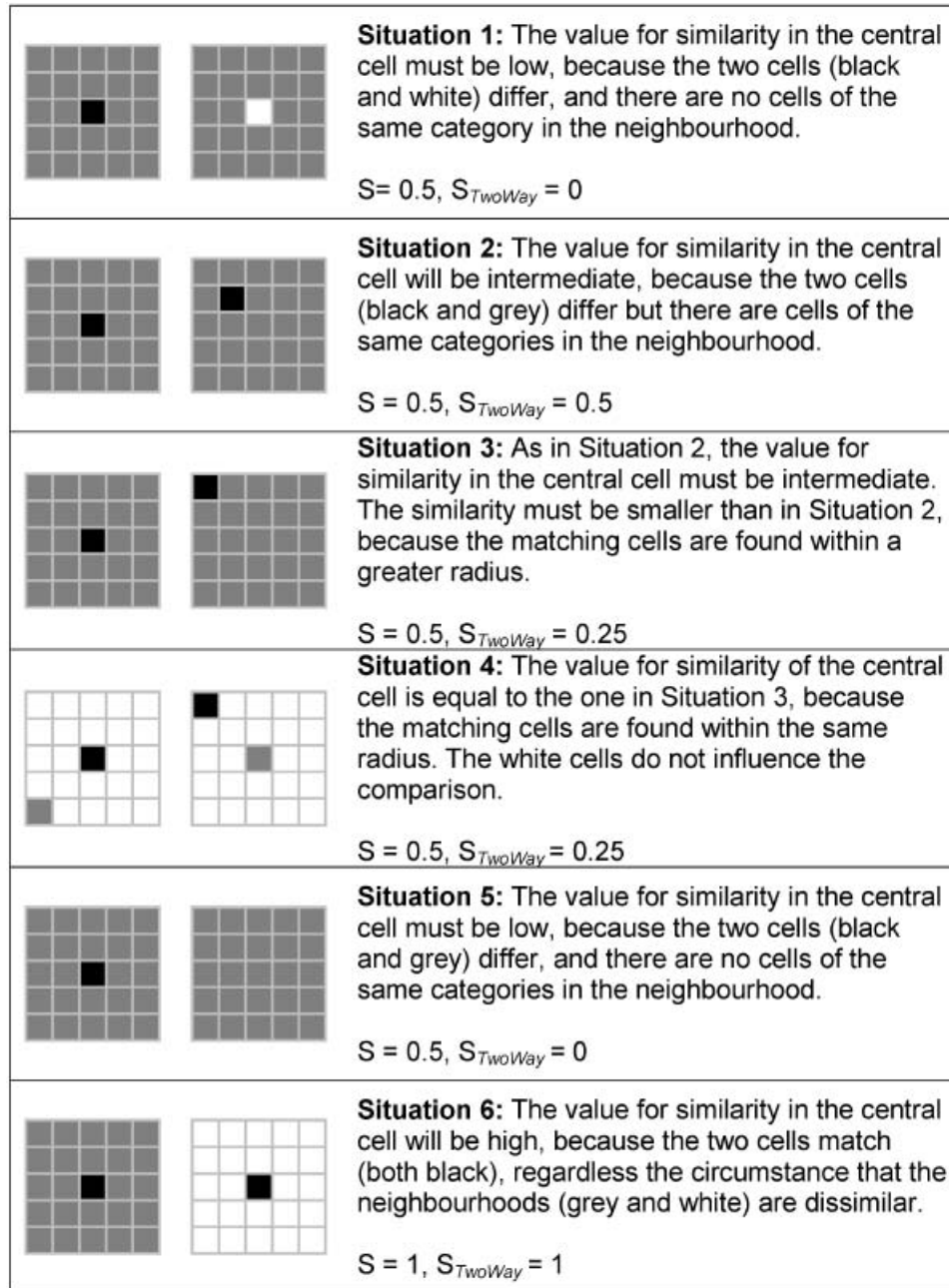


Figure 4. Six situations in which the middle cells of the left and right map are compared, with consideration of fuzziness of location. Weights according exponential decay function with halving distance of 2.

fuzziness of categories is not considered. The concept of neighbourhood ring needs to be introduced. In a raster map cells that are at the same distance from a central cell are said to form a neighbourhood ring. In figure 5 the first nine rings are

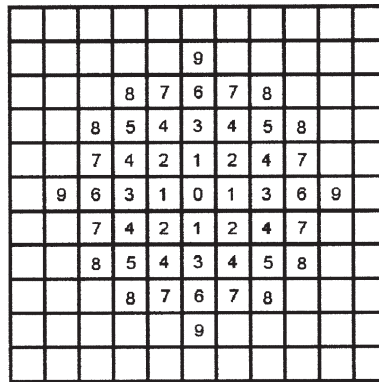


Figure 5. Numbered rings within a four cell radius.

numbered 1 to 9. The central cell is numbered 0. In table 4 their relevant characteristics are presented.

The calculation of K_{Fuzzy} as described below applies for fuzziness of location with a distance decay membership function. The membership values depend on the membership function. In this case (equation 15) it is an exponential decay function, with a halving distance of two cells.

$$M(d) = e^{\ln(1/2) \times d/2} = 2^{-d/2} \tag{15}$$

Consider the generic contingency table comparing maps X and Y (table 5), where p_{ij} = fraction of cells which are of category i in map X and category j in map Y ; and X_i = total fraction of category i in map X .

In case the two central cells, category a in map Y and category b in map X , do not match, then the probability that both the central cells have their counterpart on

Table 4. Ring characteristics.

	Ring									
	0	1	2	3	4	5	6	7	8	9
Number of cells	1	4	4	4	8	4	4	8	8	4
Cumulative number of cells excluding central	0	4	8	12	20	24	28	36	44	48
Distance (cells)	0	1	$\sqrt{2}$	2	$\sqrt{5}$	$\sqrt{8}$	3	$\sqrt{10}$	$\sqrt{13}$	4
Membership value	1	0.71	0.61	0.5	0.46	0.38	0.35	0.33	0.30	0.25

Table 5. Generic contingency table.

Map X categories	Map Y categories				Total
	1	2	...	c	
1	p_{11}	p_{12}	...	p_{1c}	X_1
2	p_{21}	p_{22}	...	p_{2c}	X_2
...
C	p_{c1}	p_{c2}	...	p_{cc}	X_c
Total	Y_1	Y_2	...	Y_c	1

a cell within a certain distance is calculated as $P(n)$ (equation 16). There, n is the number of cells present within that distance

$$P(n) = (1 - (1 - X_a)^n) \times (1 - (1 - Y_b)^n) \quad (16)$$

The smallest distance within which the central cells of both cells are matched on the other map determines the similarity in a two-way fuzzy comparison. The probability that this is the i -th neighbourhood ring is the probability that both cells match within the cumulative number of cells of the i -th ring, $P(n_i)$, minus the probability that both cells already match within the previous ring, $P(n_{i-1})$.

$$E(i|_{i \geq 1}) = \sum_{a=1}^c \sum_{b=1}^c [(1 - \delta_{ab}) \times Y_a \times X_b \times (P(n_i) - P(n_{i-1}))] \quad (17)$$

Equation (17) calculates for each combination of categories, a and b , the probability that their determining ring is the i -th. δ_{ab} stands for the Kronecker-delta of a and b , which has the value 1 if a and b are equal, and 0 if they are not.

The probability of matching central cells is calculated separately and according to the Kappa statistic (Monserud and Leemans 1992) (equation 18).

$$E(i|_{i=0}) = \sum_{a=1}^c Y_a \times X_a \quad (18)$$

The total statistic for the expected percentage of agreement is the weighted summation of all rings, according to equation (19).

$$P_e = \sum_{i=0}^R E(i) \times M(d_i) \quad (19)$$

In equation (19), R is the number of the furthest ring, M is the fuzzy membership function and d_i is the radius of the i -th ring.

The derivation of \mathbf{K}_{Fuzzy} as presented here does not consider the size of the map. The size of the maps is relevant however, because the neighbourhoods are different at the edges of maps. This should be considered in case small or irregularly shaped maps are compared. In these cases \mathbf{K}_{Fuzzy} is underestimated because P_e is overestimated. A solution to this problem is to find the cumulative number of cells in each neighbourhood ring for every cell, calculate the expected similarity for each cell and derive the average per cell. An alternative for the analytical calculation of P_e is to find an estimate by Monte Carlo analysis.

3. Results

3.1. Hypothetical case

The two maps in figure 6 were created in order to demonstrate the features of the map comparison method. Several types of differences occur: minor shifts, major shifts, growth/decline, introduction/removal, and differences of cell categories within clusters of similar content. The method is symmetrical; this means that there is no difference between comparing map 1 with map 2 or vice versa. Therefore, growth is equivalent to decline, as is introduction to removal. A large part of the map is coloured white, this does not indicate a so-called no-data value, but rather the white cells represent a category, just like the coloured cells.

Figure 7 gives the results of the direct cell-by-cell method (a) and the proposed fuzzy cell-by-cell method (b). The fuzzy membership function is that of exponential

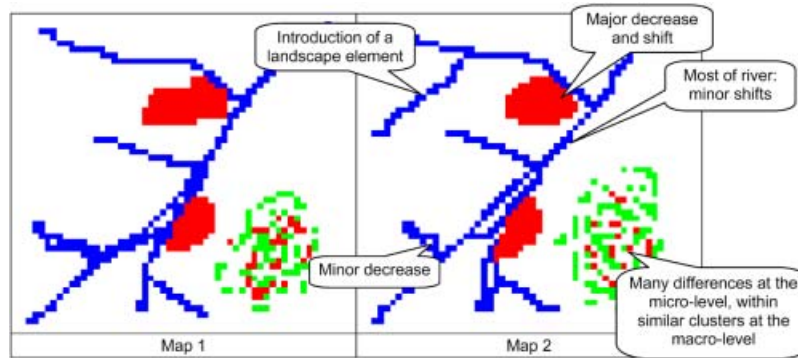


Figure 6. The two maps to compare.

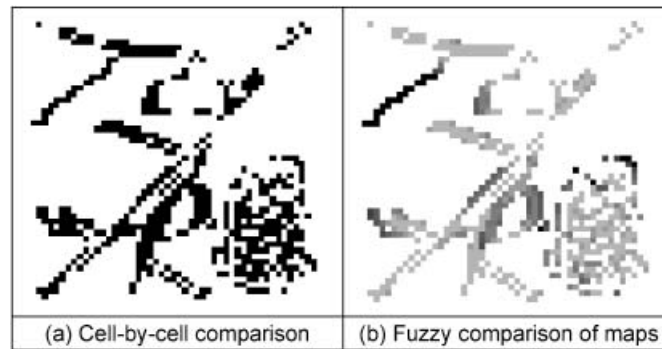


Figure 7. Comparison results (a) Cell-by-cell comparison. (b) Fuzzy comparison of maps.

decay with a halving distance of two cells and a neighbourhood with a four-cell radius. The direct cell-by-cell method consists of the pair-wise comparison of the categories in each cell of the two maps; cells where the maps are identical in both maps are in white, cells where the categories differ are in black. In the fuzzy comparison map lighter cells are more similar than darker cells.

The comparison map that results from the procedure contains values between 0 and 1. This can be more detailed than required. Based on the objective of the map comparison it can be worthwhile to include a classifying step. For instance it is possible to distinguish between total agreement, medium similarity and low similarity. Figure 8 gives the map resulting from classification with the use of a threshold level at 0.65. The areas containing new introductions (e.g. the added linear element in the upper-left corner) or major shifts (e.g. the shifts of two larger oval shapes) are distinguished from the areas of minor shifts (e.g. the other linear elements) and fluctuations within patterns (e.g. the pattern of coloured cells at the lower-right side of the map).

K_{Fuzzy} is calculated to be 0.49. This means that the maps are significantly more similar than would be expected solely from the number of cells of each category, because that level of similarity has the K_{Fuzzy} value of 0. The maps are, however, also clearly distinct, because highly similar maps will have a K_{Fuzzy} value close to 1, which stands for completely identical. As a bare figure the K_{Fuzzy} statistic is not



Figure 8. Three levels of agreement by the proposed fuzzy comparison method.

highly informative. It is more informative if there is reference material available as in the practical case presented in §3.2.

3.2. Practical case

The case presented here applies the two-way fuzzy comparison method for validation. It compares results generated by a model with real data. The particular model is a constrained cellular automaton (White *et al.* 1997) applied for the study of the urban development of Dublin, as part of the Murbandy project (White *et al.* 2000).

Three maps are compared with the observed 1998 data (figure 9). The first map is the 1988 base map (figure 10(a)), which was the starting situation for the model. Next is the 1998 map generated by the original model (figure 10(b)). Finally, the 1998 map generated by an improved version of the model (figure 10(c)) is used. The land-use maps are found in the left column, the comparison maps in the right (figure 10). Lighter cells in the comparison maps indicate larger similarity.

The comparison with the base data (figure 10(a)) yields a relatively high K_{Fuzzy} (0.90), even though the modelling effort is zero. This means that between 1988 and 1998 a small number of cells change land-use, however the changes are severe (not many cells are coloured grey; they are mostly dark grey).

The K_{Fuzzy} of the base map can be used as a reference level. Models scoring lower

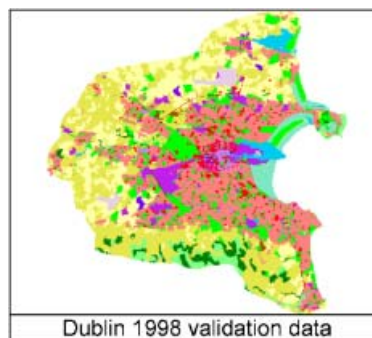


Figure 9. Dublin 1998 validation data.

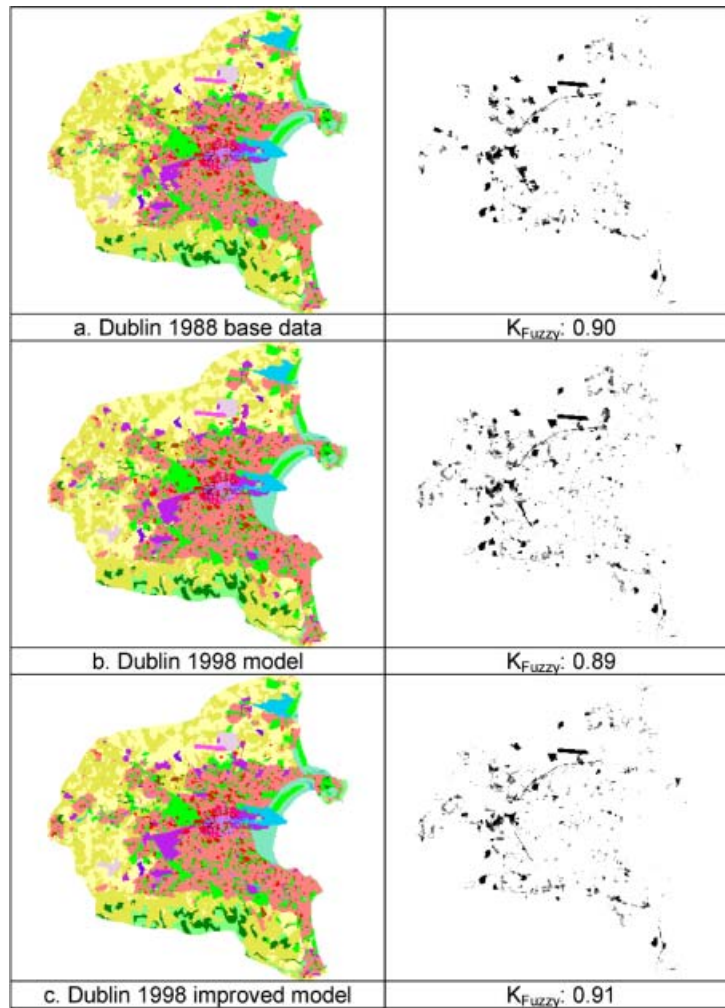


Figure 10. Three comparison results from validation process.

than 0.90 do ‘more damage than good’, while models scoring higher achieve ‘better than minimally required’.

The results from the original model (figure 10(b)) contain a relatively large number of cells that are not identical (they are grey) and their similarity is relatively low (they are mostly dark grey). As a result K_{Fuzzy} is smaller than that of the 1988 base data.

Finally, the result map of the improved model still contains a large number of non-identical cells, however the similarity of these cells is relatively high (they are lighter grey). The resulting K_{Fuzzy} is higher than that of the base data and therefore yields a positive validation of this model.

4. Discussion

By applying fuzzy set theory for the comparison of categorical maps it is possible to obtain a spatial and gradual analysis of the similarity of two maps. The results

from the comparison are basically in accordance with those of a visual inspection: it distinguishes minor deviations and fluctuations within similar areas from major deviations. The comparison method considers uncertainty and vagueness in the specification of the location of categories (fuzziness of location) as well as in the definition of the categories (fuzziness of category).

The values for similarity will range from 0 to 1. The average of all cells can be used as a measure of overall similarity of the two maps and also lies between 0 and 1. The comparison method yields results that are more gradual than those from other methods (kappa statistic or cell-by-cell comparison); hence it is more likely to give an adequate indication of small differences.

The introduction of the K_{Fuzzy} statistic makes it possible to compare individual comparison results, and therefore makes it possible to rank a collection of maps according to similarity to a reference map. In the calculation of K_{Fuzzy} the observed level of similarity is corrected for the statistically expected level of similarity. The derivation of expected similarity presented in this paper is valid for comparisons considering only fuzziness of location. Furthermore, the derivation assumes infinitely large maps. For small or irregularly shaped maps and for comparisons that also involve fuzziness of category, K_{Fuzzy} has not been derived yet. Instead of formally deriving the expected level of similarity it is also an option to apply Monte Carlo analysis of randomly generated maps. A general expression or procedure for calculation of K_{Fuzzy} will be subject of further research.

The selection of the appropriate shape and size of the membership function deserves further research as well. These settings determine the tolerance of the comparison. It is expected that the appropriate tolerance is related to the uncertainty contained in the map. There are many sources of uncertainty for instance data quality, model complexity, spatial scale and definition of map categories. Once more is known about the relationship between uncertainty and fuzzy representation of maps, it will be worthwhile to further explore the possibilities of differentiation of fuzzy representation; the two maps that are compared can be subject to different membership functions, the neighbourhood radius may vary per category, for model results that look further in the future a larger tolerance may be used, and many other refinements can be considered.

The comparison methods can be of practical use in calibration procedures. The overall figure for similarity can be used directly to qualify model results. It is potentially more effective to incorporate the spatial results in the procedure and focus the model improvements on those areas or categories with the most severe disagreement.

The results of remote sensing and high-resolution spatial models can be assessed in more detail than before. Based upon the spatial comparison results it is possible to specify the discrepancies between observed data and model results. Furthermore the comparison map can be used to find correlations between similarity and other spatial occurrences (e.g. certain categories, distances from landmarks, geographical and political boundaries etc.).

The applicability of the method is not restricted to geographical problems; other fields of potential use are image analysis, pattern recognition and video image analysis.

References

- BANDEMER, H., and GOTTWALD, S., 1995, *Fuzzy sets, fuzzy logic, fuzzy methods with applications* (Chichester, New York: J. Wiley).

- CARLETTA, J., 1996, Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, **22**, 249–254.
- CHENG, T., MOLENAAR, M., and LIN, H., 2001, Formalizing fuzzy objects from uncertain classification results. *International Journal of Geographical Information Science*, **15**, 27–42.
- FISHER, P., 2000, Sorites paradox and vague geographies. *Fuzzy Sets and Systems*, **113**, 7–18.
- FOODY, G. M., 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80**, 185–201.
- LEWIS, H. G., and BROWN, M., 2001, A generalized confusion matrix for assessing area estimates from remotely sensed data. *International Journal of Remote Sensing*, **22**, 3223–3235.
- METTERNICHT, G., 1999, Change detection assessment using fuzzy sets and remotely sensed data: an application of topographic map revision. *ISPRS Journal of Photogrammetry & Remote Sensing*, **54**, 221–233.
- MONSERUD, R. A., and LEEMANS, R., 1992, Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275–293.
- PONTIUS, JR. R. G., 2000, Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering & Remote Sensing*, **66**, 1011–1016.
- PONTIUS JR, R. G., and SCHNEIDER, L. C., 2001, Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, **85**, 239–248.
- POWER, C., SIMMS, A., and WHITE, R., 2001, Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, **15**, 77–100.
- SHYI-MING, C., 1995, Measures of similarity between vague sets. *Fuzzy Sets and Systems*, **74**, 217–223.
- TOLIAS, Y. A., PANAS, S. M., and TSOUKALAS, L. H., 2001, Generalized fuzzy indices for similarity matching. *Fuzzy Sets and Systems*, **120**, 255–270.
- WHITE, R., ENGELEN, G., and ULJEE, I., 1997, The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design*, **24**, 323–343.
- WHITE, R., ENGELEN, G., ULJEE, I., LAVALLE, C., and EHRLICH, D., 2000, Developing an Urban Land use Simulator for European Cities. In *Proceedings of the 5th EC-GIS Workshop held in Stresa, Italy 28–30 June 1999*, edited by E. Fullerton (Ispra, Italy: European Commission, Joint Research Centre), pp. 179–190.
- WINTER, S., 2000, Location similarity of regions. *ISPRS Journal of Photogrammetry & Remote Sensing*, **55**, 189–200.
- XUZHU, W., DE BAETS, B., and KERRE, E., 1995, A comparative study of similarity measures. *Fuzzy Sets and Systems*, **73**, 259–268.
- ZADEH, L. A., 1965, Fuzzy sets. *Information and Control*, **8**, 338–353.
- ZWICK, R., CARLSTEIN, E., and BUDESCU, D. V., 1987, Measures of similarity among fuzzy concepts: a comparative analysis. *International Journal of Approximate Reasoning*, **1**, 221–242.