

An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation

ALEX HAGEN-ZANKER

1. Research Institute for Knowledge Systems, PO Box 463, Maastricht, the Netherlands
2. Urban Planning Group, Eindhoven University of Technology, PO Box 513, Eindhoven, the Netherlands

Email: a.h.hagen.zanker@tue.nl

Abstract: The Fuzzy Kappa statistic expresses the agreement between two categorical raster maps. The statistic goes beyond cell-by-cell comparison and gives partial credit to cells based on the categories found in the neighborhood. When matching categories are found at shorter distances the agreement is higher. Like the well-established Kappa statistic the Fuzzy Kappa statistic expresses the mean agreement relative to the expected agreement. The model underlying the expected agreement assumes absence of spatial autocorrelation in both compared maps. In reality however, spatial autocorrelation does lower the expected agreement as matching categories become less likely to be found close-by. Since most maps have some degree of spatial autocorrelation, the calculated expected agreement is generally higher than the true expected agreement. This leads to counterintuitive results when maps that appear to have considerable agreement obtain negative Fuzzy Kappa values. Furthermore, the Fuzzy Kappa may be biased, as it systematically attributes lower agreement to maps with stronger spatial autocorrelation. This paper proposes an improved

Fuzzy Kappa statistic that is based on the same local agreement and has the same attractive properties as the original Fuzzy Kappa. The novelty is that the new statistic accounts for spatial autocorrelation, such that the expected Fuzzy Kappa for maps that are not cross-correlated is equal to zero. The improved statistic is applied on two cases to demonstrate its properties.

Keywords: Fuzzy Kappa, Map Comparison, Accuracy, Validation

1 Introduction

Hagen (2003) and Hagen-Zanker *et al.* (2005) introduced a method for the comparison of categorical raster maps. The method differs from existing methods by not only giving credit for cell-by-cell agreement, but also, through fuzziness of location, for near cell-by-cell agreement. Furthermore, the incorporation of fuzziness of categories means that some pairs of categories are considered more similar than others. The results of the map comparison consist of a comparison map, detailing for each cell the degree of agreement on a scale of 0 to 1, and a Fuzzy Kappa statistic that summarizes the agreement in a single number. The method is primarily intended as a goodness-of-fit measure for the validation and calibration of high resolution land use models (e.g. Hagoort *et al.* 2008), but it is finding applications in other fields and beyond the tasks of calibration and validation as well (e.g. Hemer 2006, Iverson & Prasad 2007, Klingseisen *et al.* 2008, Meyfroidt & Lambin 2008).

The Fuzzy Kappa is identical in form to the Kappa statistic (Cohen 1960, Monserud & Leemans 1992); It ‘corrects’ the mean agreement of the comparison map for the agreement that can be expected given the total number of cells taken in by each category, i.e. their histograms. The effect of the correction is that a Kappa value of 1 corresponds to a perfect agreement and 0 to the expected agreement of two maps that are not cross-correlated. For the Fuzzy Kappa this is not the case however, since the expected agreement not only depends on the histograms of the compared maps, but also their spatial autocorrelation. The calculation of the expected

agreement however, does not consider the impact of spatial autocorrelation on the expected agreement; it assumes that the probability of a cell to be taken in by a category is independent of the categories found in the neighbourhood (Hagen-Zanker *et al.* 2005).

The fuzziness of location holds that the degree of belonging of a cell to a category is a function of the distance to the nearest cell of that, or a similar, category. In a map without spatial autocorrelation it is generally more likely to find different categories in the close neighbourhood than with spatial autocorrelation (depending on the precise nature of the spatial autocorrelation). Therefore the expected agreement without spatial autocorrelation is larger than with spatial autocorrelation. Consequently, two maps that are not cross-correlated will generally have a negative Fuzzy Kappa value, which makes the interpretation of the statistic less intuitive. This limitation of the Fuzzy Kappa has been recognized by several authors (Wealands *et al.* 2005, Hagen-Zanker *et al.* 2005, Tang 2008). Foody (2008) made a caution about a possible bias.

This paper introduces an Improved Fuzzy Kappa statistic. The statistic is based on the same local agreement as the original Fuzzy Kappa, but applies an expected agreement that accounts for the spatial autocorrelation.

2 Method

2.1 Local agreement

Central to the comparison method is the introduction of a degree of belonging of each cell in either map to the categories found in the legend of the other map. The degree of belonging takes into account two sources of fuzziness; fuzziness of categories which recognizes that some categories are more similar to each other than others and fuzziness of location which states that cells partially belong more strongly to categories in their proximity.

Fuzziness of categories is expressed in a categorical similarity matrix M , where $M_{i,j}$ is a value between 0 and 1 expressing the similarity between category i in the legend of map A and category j in the legend of map B .

Hagen-Zanker *et al.* (2005) focused on different applications of the matrix, including the possibility of weighting different aspects of agreement and comparing maps with unequal legends.

Fuzziness of location is introduced through a distance decay function that specifies the degree to which a cell belongs to the categories found in its proximity. At distance 0 the function returns 1, indicating that every cell fully belongs to its own category. At larger distances the function return values between 0 and 1. The contributions of all cells are combined using the fuzzy union, meaning that for each category the highest degree of belonging over all neighbours prevails. Considering that the function is distance decaying, this means that the degree of belonging is based on the nearest neighbour distance to cells of all categories.

The combined effect of fuzziness of location and category is expressed in the following equations:

$$\mu_{c,j}^A = \max_{i=0}^m (M_{i,j} * f(d_{c,i}^A)) \quad 1$$

$$\mu_{c,i}^B = \max_{j=0}^n (M_{i,j} * f(d_{c,j}^B)) \quad 2$$

where: $\mu_{c,j}^A$ is the degree of belonging of cell c in map A to the category j found in the legend of map B , $d_{c,i}^A$ is the nearest neighbour distance of cell c in map A to a cell of category i and m is the number of categories in the legend of map A (likewise for map B). $f(d)$ is some distance decay function and $M_{i,j}$ is the similarity of category i in the legend of map A to j in the legend of map B . The calculation of nearest neighbour distances is the computational bottleneck of the method and it is expedient to use an efficient Euclidean distance transform algorithm such as the fast and exact algorithm of Felzenszwalb & Huttenlocher (2004).

The matrix M and distance decay function $f(d)$ are reflecting the intention of the analyst and are not inherent to the map. It is the analyst's perspective and tolerance for different kinds of errors that determines the form of the function and the values of the matrix. Therefore there is a degree of arbitrariness and uncertainty about the meaning of the results. This uncertainty may be quantified by means of a sensitivity analysis. There are no fixed methodologies for finding an appropriate function and matrix. One approach would be to optimize the

settings to a certain target, for instance correspondence to human judgement as expressed in an Internet survey (Kuhnert et al. 2005). Another target would be to optimize the value of Fuzzy Kappa itself; the optimized settings can provide insight in the nature of the differences. Other possible avenues to determining the values of M are semantic analysis of class descriptions (Ahlqvist 2005), eliciting expert opinion with interviews (Fritz & See 2005) and incorporating the degree of mutual information (Finn 1993).

The agreement of a cell is the degree to which the cell in map A belongs to the category found in map B and *vice versa*, as follows:

$$P_c^{A,B} = \min(\mu_{c,B_c}^A, \mu_{c,A_c}^B) \quad 3$$

where: $P_c^{A,B}$ is the agreement between map A and B in cell c , A_c and B_c are the categories found at cell c of maps A and B . This definition of agreement at the cell level is identical to Hagen (2003) except that the formulation on the basis of nearest neighbour distances resolves the necessity to specify a maximum radius for the neighbourhood.

The mean agreement is calculated as follows:

$$P = \frac{\sum_{c=1}^r P_c^{A,B}}{r} \quad 4$$

where: P is the mean agreement of the two compared maps, c iterates over all cells in the map and r is the number of cells.

2.2 Expected agreement

The objective of this section is to derive the expected agreement for two maps that are not cross-correlated, given their spatial structure. One approach would be to first quantify the spatial autocorrelation, for instance by means of transiograms (Weidong 2006), and consequently derive the expected agreement. That approach would introduce considerably more complications than present in the method of Hagen-Zanker *et al.* (2005)

which is already considered overly complicated (Huang & Lees 2007). Instead an approach is followed that directly quantifies the effect of the spatial autocorrelation, namely the distribution of the degrees of belonging. For illustration, consider a situation of two pairs of maps with only the categories *black* (10% of the area in all of the maps) and *white* (90% of the area). The maps of the first pair are strongly spatially autocorrelated such that the black cells form a single cluster in both maps. In the second pair of maps the spatial autocorrelation is weak or non-existent. The maps are compared without considering fuzziness of categories, i.e. black and white are not similar. Fuzziness of location is set such that directly adjacent neighbours contribute for 50% to the degree of belonging and further neighbours not at all. Applying these settings, the mean agreement of both map pairs is identical (0.815); there are 80 cells with perfect agreement, 17 with perfect disagreement and 3 with 50% agreement. See figure 1 for the maps and their comparison. The difference in spatial autocorrelation manifests itself in particular in the occurrence of white cells that neighbour a black cell (17 and 13 in the first pair versus 24 and 25 in the second pair). Table 1 gives the full distribution of degrees of belonging.

Table 1. Distribution of categories and degrees of belonging in the maps of the example

Map	Category	Area [cells]	Belonging to white [cells]			Belonging to black [cells]		
			0	0.5	1	0	0.5	1
1-A	White	90			90	73	17	
	Black	10		10				10
1-B	White	90			90	77	13	
	Black	10	1	9				10
2-A	White	90			90	66	24	
	Black	10		10				10
2-B	White	90			90	65	25	
	Black	10		10				10

Zeroes left blank

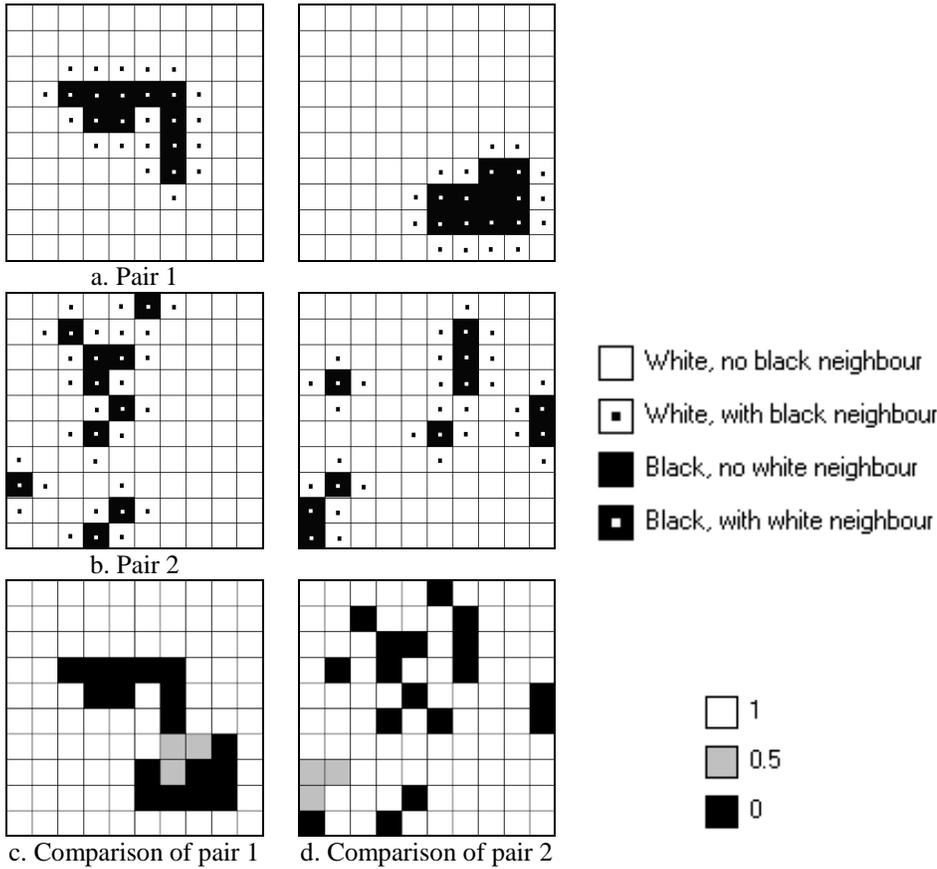


Figure 1. Two pairs of maps with different levels of spatial autocorrelation, but the same mean agreement

There are four local configurations that lead to a non-zero agreement: a white cell in both maps (WW), a black cell in both maps (BB), a white cell bordering black in map *A*, and a black cell bordering white in map *B* (WB) and a black cell bordering white in map *A*, and a white cell bordering black in map *B* (BW). The first two configurations correspond to full agreement and their probability is the same for both map pairs. The differences in spatial structure lead to distinct probabilities for the occurrence of the configurations BW and WB in both map pairs and hence distinct expected agreements. Table 2 gives the distribution of probabilities for each configuration for both pairs as well as the probability of the distribution when spatial autocorrelation is ignored. The expected agreement for the first pair is 0.834 and for the second 0.8445. When spatial

autocorrelation is ignored the expected agreement is 0.848 for both pairs; considering that there are 64 centre cells with four neighbours each (expected agreement: 0.851), 32 edge cells with three neighbours each (expected agreement: 0.844) and 4 corner cells with two neighbours each (expected agreement: 0.837).

Table 2. Probabilities of spatial configurations that result in positive agreement

Config.	Agr.	With spatial autocorrelation		Without spatial autocorrelation
		Pair 1	Pair 2	(4 neighbours) Both pairs
WW	1	0.9*0.9 = 0.81	0.9*0.9 = 0.81	0.9*0.9 = 0.81
BB	1	0.1*0.1 = 0.01	0.1*0.1 = 0.01	0.1*0.1 = 0.01
WB	0.5	0.17 * 0.09 = 0.015	0.24 * 0.1 = 0.024	0.9 * 0.1 * (1-0.9 ⁴) * (1-0.1 ⁴) = 0.031
BW	0.5	0.13 * 0.1 = 0.013	0.25 * 0.1 = 0.025	0.1 * 0.9 * (1-0.1 ⁴) * (1-0.9 ⁴) = 0.031

Config. = configuration, Agr. = agreement, WW = two white cells, BB = two black cells, WB = white cell bordering a black cell in first map and black cell bordering a white cell in the second map, BW = black cell bordering a white cell in the first map and white cell bordering a black cell in the second

The example illustrates that spatial autocorrelation has a negative effect on the expected agreement. It also shows that the expected agreement for maps with little spatial autocorrelation is very close to the expected agreement when spatial autocorrelation is ignored (0.8445 resp. 0.848).

When a larger neighbourhood is considered as well as fuzziness of categories, calculating the expected agreement becomes more cumbersome. The remainder of this section details this procedure.

The distribution of degrees of belonging is described as follows:

$$P^A(\mu_j \geq x | i) \equiv \frac{\sum_{c=1}^r [A_c = i \wedge \mu_{c,j}^A \geq x]}{\sum_{c=1}^r [A_c = i]} \quad 5$$

where: $P^A(\mu_j \geq x | i)$ is the probability that the degree of belonging of a cell in map A to the category j is at least equal to threshold value x , given that the cell is taken in by category i . The index c iterates over all cells and r is the number of cells. This conditional probability is pivotal to the improvement proposed in this paper as it characterizes the effect of spatial autocorrelation through the distance component in $\mu_{c,j}^A$ (defined in equation 1).

The probability of a cell to be taken in by one or the other category is set to its frequency of occurrence on the map, according to the following equation:

$$P^A(i) \equiv \frac{\sum_{c=1}^r [A_c = i]}{r} \quad 6$$

where: $P^A(i)$ is the probability of a cell in map A to be taken in by category i . The index c iterates over all cells and r is the number of cells. A_c is the category found at cell c in map A .

Equation 3 expresses the local agreement as the minimum of the cell in map A belonging to the category in map B and *vice versa*. This means that the probability of the local agreement to be above a threshold value is the probability that both degrees of belonging are above that threshold value.

Assuming map A and B are independent; the expected agreement can be expressed as follows:

$$P^{AB}(p \geq x|i, j) = P^A(\mu_j \geq x|i) \cdot P^B(\mu_i \geq x|j) \quad 7$$

where: $P^{AB}(p \geq x|i, j)$ is the probability that the agreement between two maps A and B at one cell is above threshold value x , given that map A is taken in by category i and map B by category j .

The discrete nature of the categorical raster implies that the range of distance values that are passed to the distance decay function is also discrete and hence that the range of values for local agreement is discrete.

Therefore the previous is equivalent to:

$$P^{AB}(p = x_k|i, j) = \begin{cases} k=1: P^{AB}(p \geq x_k|i, j) \\ k>1: P^{AB}(p \geq x_k|i, j) - P^{AB}(p \geq x_{k-1}|i, j) \end{cases} \quad 8$$

where: x_k is the k -th element in the list of all possible values for local agreement ($P_c^{A,B}$ in equation 3) sorted from high to low.

The expected agreement conditional to the presence of categories i and j is then calculated as follows:

$$E(p|i, j) = \sum_{k=1}^q x_k \cdot P^{AB}(p = x_k|i, j) \quad 9$$

where the index k iterates over all possible values of x and q is the number of possible values.

The expected agreement then follows as:

$$E = \sum_{i,j} E(p|i, j) \cdot P^A(i) \cdot P^B(j) \quad 10$$

2.3 Fuzzy Kappa

The mean and expected agreement are used to calculate the Fuzzy Kappa statistic, which is identical in form to the Kappa statistic of Cohen (1960).

$$K = \frac{P - E}{1 - E} \quad 11$$

Where K is the Fuzzy Kappa statistic, P is the mean agreement and E is the expected agreement. C++ source code for the full algorithm is available from the author upon request.

Note that if no fuzziness of location is considered, the agreement of two cells is the similarity of their categories (equation 3 simplifies to: $P_c^{A,B} = M_{A_c, B_c}$), which makes the Fuzzy Kappa an extension of the weighted Kappa (Cohen 1968). Furthermore, if fuzziness of categories is not considered either, the agreement is just their equality (equation 3 simplifies to $P_c^{A,B} = [A_c = B_c]$), which makes the Fuzzy Kappa an extension of the Kappa statistic (Cohen 1960).

3 Application

3.1 Case A: Comparing maps with spatial autocorrelation but without cross-correlation

To test whether the expected Improved Fuzzy Kappa for maps that are not cross-correlated indeed is zero, a large number of maps is required. The modified random clusters landscape pattern simulator of Saura &

Martínez-Millán (2000) offers the means to such maps. The landscape simulator randomly generates categorical raster maps that appear like realistic landscapes.

The generator creates landscapes in four steps:

1. In the first step, a binary percolation map is generated. With a given probability (parameter p) each cell in the map is marked as *on* or *off*.
2. In the second step, clusters in the percolation map are identified. Clusters are groups of adjacent cells that are marked *on* in the previous step. Adjacent cells are those that neighbour each other at the sides or diagonally (i.e. Queen contiguity is applied).
3. In the third step, each cluster is assigned a category on the basis of a probability distribution (parameter p_c for each category of the legend).
4. In the fourth step, the cells that are marked *off* in the first step are considered. These cells are assigned the category that is most frequent among the 8 neighbours. If several categories are equally frequent a random tie-breaking is applied.

Deviating from Saura & Martínez-Millán (2000), the landscape is considered a torus, which means that the first row in the raster neighbours the last row and likewise the first column neighbours the last column. This addition cures the landscape generator from the tendency to create more and smaller clusters along the edges of the map.

Three maps were generated each consisting of $256 * 256$ cells and four categories, but using different parameters. Furthermore three series of 100 maps each were generated using the same three parameter sets. The parameters are given in table 3. Figure 2 shows three realizations of the model under each parameter setting.

Table 3. Settings for random map generator

Parameter	Step 1. Percolation		Step 3. Category assignment		
	p	p_1	p_2	p_3	p_4
Map 1 / Series 1	0.50	0.4	0.2	0.3	0.1
Map 2 / Series 2	0.55	0.1	0.2	0.3	0.4
Map 3 / Series 3	0.60	0.1	0.3	0.3	0.3

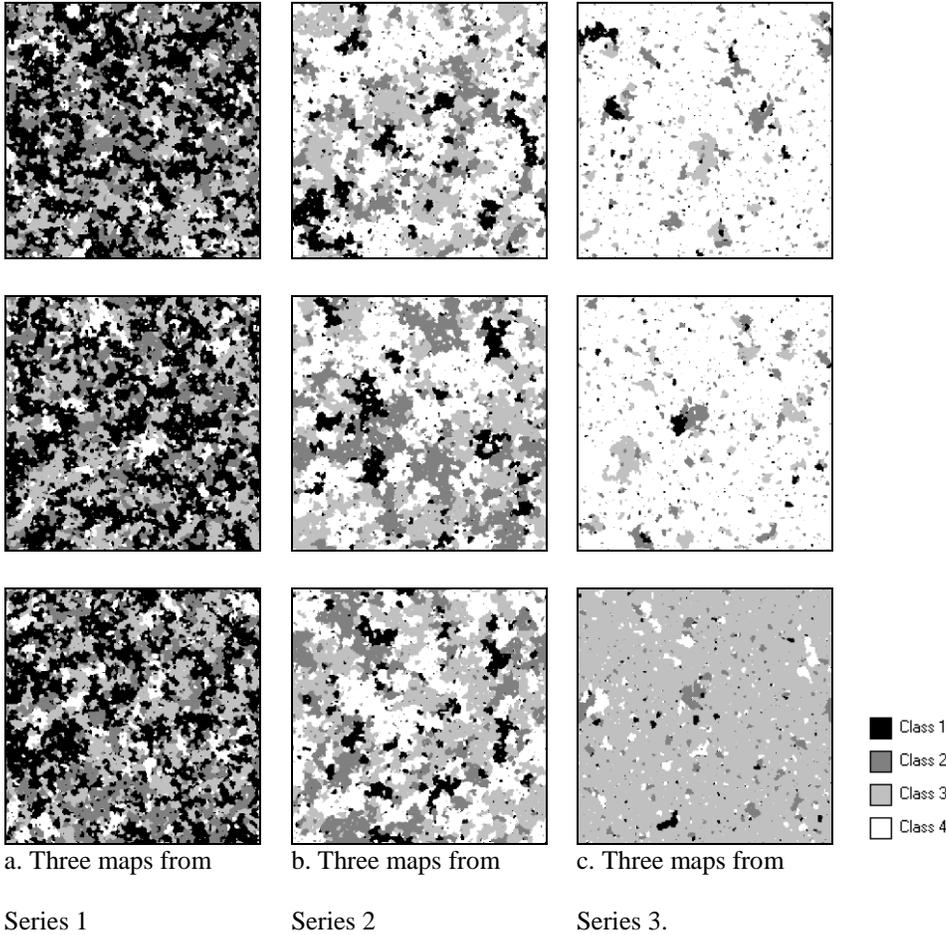


Figure 2. Three realizations of the landscape under three different settings.

The three maps were compared to the maps in the series and Kappa, Fuzzy Kappa and Improved Fuzzy Kappa were calculated. Table 4 gives both the mean and standard deviation of the results. Fuzzy Kappa was

calculated with an exponential distance decay function that is truncated at a radius of 4 cells and has a halving distance of 2 cells, i.e. the following equation was applied:

$$f(d) = \begin{cases} d \leq 4: \left(\frac{1}{2}\right)^{d/2} \\ d > 4: 0 \end{cases} \quad 12$$

. The Improved Fuzzy Kappa was also applied with an exponential distance decay function with a halving distance of 2 cells, but is not truncated, i.e. the following equation was applied:

$$f(d) = \left(\frac{1}{2}\right)^{d/2} \quad 13$$

In both cases fuzziness of categories was ignored, i.e. the following was applied:

$$M_{i,j} = \begin{cases} i = j: 1 \\ i \neq j: 0 \end{cases} \quad 14$$

Table 4. Distribution of comparison results

Map	Series	Kappa [*0.001]		Fuzzy Kappa		Improved Fuzzy Kappa [*0.001]	
		mean	std.dev.	mean	std.dev.	mean	std.dev.
1	1	0.84	9.8	-1.1	0.029	1.0	12.0
2	2	1.0	21	-1.2	0.059	1.6	24
3	3	206	351	-0.54	0.12	-0.37	9.4
1	2	1.3	12	-1	0.067	1.6	14
1	3	-0.32	6.6	-0.53	0.11	0.043	7.7
2	3	1.8	8.2	-0.7	0.06	1.9	9.0

The results confirm that the Improved Fuzzy Kappa, like the Kappa but unlike the Fuzzy Kappa has an expected value very close to zero. The 95% confidence interval for the mean Improved Fuzzy Kappa is [-0.03, 0.03] on the basis of all 600 comparisons of maps that are not cross-correlated.

3.2 Case B: Synthetic example

Hagen (2003) introduced a synthetic example to illustrate the merit of the Fuzzy Kappa statistic. This section considers the same example to illustrate the relation between Kappa, Fuzzy Kappa and Improved Fuzzy Kappa. The example consists of the two maps presented in Figure 3.

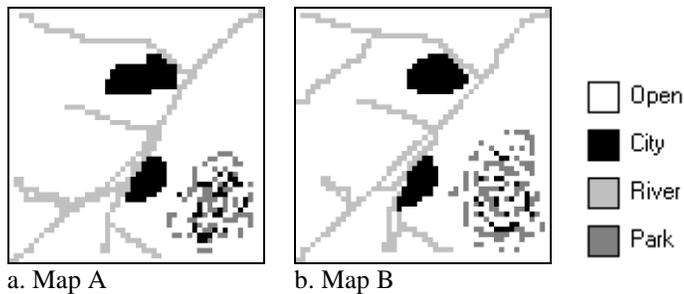


Figure 3. Two maps compared in Case B

The proposed improvement only affects the Fuzzy Kappa statistic and not the local agreement therefore the comparison maps presented in figure 4 are the same as in (Hagen-Zanker et al. 2005). The maps illustrate that the presence of a river branch in map B that is not at all present in map A is considered a strong difference (dark cells, low values). The mixed area of city and park in the lower-right quadrant is considered quite similar (light cells, high values) even though only a few cells match precisely (white cells, value 1).

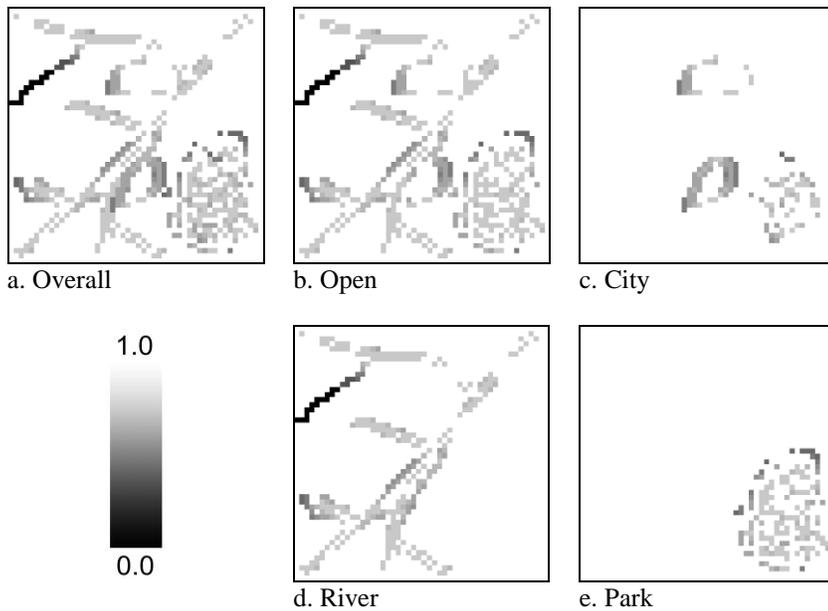


Figure 4. Fuzzy agreement, overall and per category

The statistical results (figure 5) confirm the difference between Kappa and Fuzzy Kappa. Park is the category of least agreement according to the Kappa statistic, but both Fuzzy Kappa statistics give credit to the large share of near-overlap of this category and consider the agreement on this category stronger than on the categories open and river.

Overall the values for Improved Fuzzy Kappa are higher than for Fuzzy Kappa. This is according to expectations since spatial autocorrelation affects the distribution of all categories. The increase for the category park is smaller than for river, which may be explained by the observation that park is more scattered than river, and hence gains less when spatial autocorrelation is accounted for in the expected agreement.

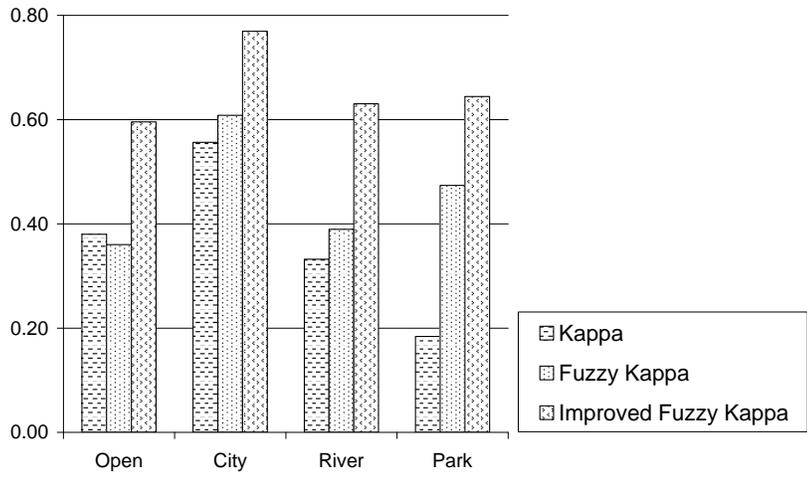


Figure 5. Per category comparison of Case B.

4 Conclusion

The Improved Fuzzy Kappa extends the model underlying expected agreement of the Fuzzy Kappa by accounting for spatial autocorrelation. It thereby resolves a problem of counterintuitive interpretation and possible bias that has been brought forward by several authors (Wealands *et al.* 2005, Hagen-Zanker *et al.* 2005, Foody 2008, Tang 2008). As a positive side effect, the method has simplified and is easier to comprehend. It can be readily implemented in any software environment. The Improved Fuzzy Kappa will make part of version 4.0 of the free Map Comparison Kit software (www.riks.nl/mck) where it replaces the Fuzzy Kappa.

Some cautions can be made about what the Improved Fuzzy Kappa statistic is not. The Improved Fuzzy Kappa statistic is not a statistical test. It expresses agreement between maps, but it does not make a statement about the likelihood that both maps are drawn from the same distribution or are realizations of the same process. A further development in this direction would require understanding of the effect of spatial autocorrelation on the variance of the agreement. The variance would be a function not only of the spatial structure but also the size of the map. Statistical tests for Kappa are available (Foody 2004, Sim & Wright

2005), but these are based on the comparison of a sample of the population (i.e. not on a full map) and assume independent observations (i.e. no spatial autocorrelation).

The Improved Fuzzy Kappa statistic does not express the similarity of spatial structure. It is possible that maps with identical structure have low Fuzzy Kappa values. A possible further development would be to find complementing statistics for the agreement in spatial structure and express Fuzzy Kappa as a function of agreement in location and spatial structure. This would be analogous to Pontius (2000) and Hagen (2002) who express the Kappa statistic as a function of agreement in location and quantity.

References

Ahlqvist, O., 2005. Using uncertain conceptual spaces to translate between land cover categories.

International Journal of Geographical Information Science, 19(7), 831-857.

Cohen, J., 1960. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Cohen, J., 1968. Weighted Kappa - Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213-220.

Felzenszwalb, P.F. and Huttenlocher, D.P., 2004, *Distance transforms of sampled functions* Technical report TR2004-1963. Faculty of Computing and Information Science, Cornell University.

Finn, J.T., 1993. Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Science*, 7(4), 349 - 366.

Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5), 627-633.

Foody, G.M., 2008. Harshness in image classification accuracy assessment. *International Journal of Remote Sensing*, 29(11), 3137-3158.

- Fritz, S. and See, L., 2005. Comparison of land cover maps using fuzzy agreement. *International Journal of Geographical Information Science*, 19(7), 787-807.
- Hagen, A., 2002. Multi-method assessment of map similarity. In: M. Ruiz, M. Gould and J. Ramon (eds.) *Proceedings of the 5th AGILE Conference on Geographic Information Science*, 25-27 April 2002 Palma. Universitat de les Illes Balears, 171-182.
- Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographical Information Science*, 17(3), 235-249.
- Hagen-Zanker, A., Straatman, B. and Uljee, I., 2005. Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, 19(7), 769-785.
- Hagoort, M., Geertman, S. and Ottens, H., 2008. Spatial externalities, neighbourhood rules and CA land-use modelling. *Annals of Regional Science*, 42(1), 39-56.
- Hemer, M.A., 2006. The magnitude and frequency of combined flow bed shear stress as a measure of exposure on the Australian continental shelf. *Continental Shelf Research*, 26(11), 1258-1280.
- Huang, Z. and Lees, B.G., 2007. Assessing a single classification accuracy measure to deal with the imprecision of location and class: Fuzzy weighted Kappa versus Kappa. *Journal of Spatial Science*, 52(1), 1-12.
- Iverson, L.R. and Prasad, A.M., 2007. Using landscape analysis to assess and model tsunami damage in Aceh province, Sumatra. *Landscape Ecology*, 22(3), 323-331.
- Klingseisen, B., Metternicht, G. and Paulus, G., 2008. Geomorphometric landscape analysis using a semi-automated GIS-approach. *Environmental Modelling and Software*, 23(1), 109-121.
- Kuhnert, M., Voinov, and Seppelt, R., 2005. Comparing raster map comparison algorithms for spatial modeling and analysis. *Photogrammetric Engineering and Remote Sensing*, 71(8), 975-984.

- Meyfroidt, P. and Lambin, E.F., 2008. Forest transition in Vietnam and its environmental impacts. *Global Change Biology*, 14(6), 1319-1336.
- 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 and Remote Sensing*, 66(8), 1011-1016.
- Saura, S. and Martínez-Millán, J., 2000. Landscape patterns simulation with a modified random clusters method. *Landscape Ecology*, 15(7), 661-678.
- Sim, J. and Wright, C.C., 2005. The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257-268.
- Tang, G., 2008. A new metric for evaluating the correspondence of spatial patterns in vegetation models. *Global Ecology and Biogeography*, 17(4), 465-478.
- Wealands, S.R., Grayson, R.B. and Walker, J.P., 2005. Quantitative comparison of spatial fields for hydrological model assessment - some promising approaches. *Advances in Water Resources*, 28(1), 15-32.
- Weidong, L., 2006. Transiogram: A spatial relationship measure for categorical data. *International Journal of Geographical Information Science*, 20(6), 693-699.

Acknowledgements

This paper addresses the most frequently asked question of Map Comparison Kit users: “Why do I get negative Fuzzy Kappa values for maps that appear quite similar?” I would like to thank the users for sending their feedback and enabling RIKS to improve on the methods and software. The elaborate and detailed feedback of three anonymous reviewers has been very helpful and is greatly appreciated.

