

Assessing the Accuracy of Changes in Spatial Explicit Land Use Change Models

Jasper van Vliet ¹

¹ Research Institute for Knowledge Systems b.v. P.O. box 463, 6200 AL Maastricht, The Netherlands
(jvanvliet@riks.nl)

ABSTRACT

Results of land use change models are often assessed by comparing simulation results with actual land use data. For this the Kappa coefficient of agreement is a common algorithm for map comparison; it expresses the agreement between two categorical datasets, corrected for the distribution of class sizes. However, most land use has a high inertia over the length of a typical simulation period, and therefore the changing land use comprises only a small part of the map. Since methods for accuracy assessment do not acknowledge this inertia, measured accuracies are generally very high, which suggests that models are very accurate. However the similarity between simulation results and actual land use data depends on the amount of land use change at least as much as the accuracy of simulated land use changes. To gain more insight in the performance of a land use change model, it should be acknowledged that most land use persists rather than changes. Therefore the expected agreement is much higher than what is computed from the distribution of class sizes only. This paper describes $K_{\text{simulation}}$, a method that does consider the amount of change in the accuracy assessment of results of land use change models. This method is illustrated with a simple example and then applied to the results of a case study.

Keywords: Maps comparison, Accuracy assessment, Land use change, Kappa statistics

1. CALIBRATION AND ASSESSMENT OF LAND USE CHANGE MODELS

Over the last decade land use change modeling experienced a gradual shift from conceptual and mostly synthetic studies to real world applications. This development is fuelled by an increase in the availability of spatial data and tools for modeling. Applications are used to study land use change processes, but also for scenario studies and policy analysis (de Nijs et al., 2004). For scientific use as well as scenario studies and policy analysis it is important to assess the accuracy of results of land use change models and to be aware of its implication.

The application of any land use change model to a specific case region requires model calibration. Calibration is the process of adjusting parameter values to improve a model's goodness of fit. Validation is the assessment of this goodness of fit with an independent dataset. Essentially, calibration and validation are an iterative process that stops when a calibration is assessed as good enough. For a more elaborate discussion on this terminology see Rykiel (1996) or Refsgaard and Henriksen (2004).

Land use models are typically calibrated to simulate known historical land use changes. For this procedure two sets of land use data are required: one for the start of the simulation period (T0) and one for the end of the simulation period (T1). You can assess a calibration by simulating land use changes from T0 to T1 and comparing the result with actual land use data. In this paper, land use data are raster maps that represent the predominant land use as a categorical value for each grid cell.

Visual interpretation by experts is arguably the best way to assess these simulation results. Unfortunately this is also highly subjective and not reproducible. Consequently, there is a need for objective and reproducible methods to assess simulation results (Power et al. 2001; Hagen 2002). Although it is acknowledged that there are several aspects that can be assessed from land use maps (Hagen-Zanker and Martens, 2008) results are predominantly assessed on a pixel level. Pixel-by-pixel similarity can be computed from a contingency table, from which you can compute statistics, such as

the Kappa coefficient of agreement. Kappa is commonly used for accuracy assessment of remote sensing image classifications (Foody 2002) and results of spatial models (Monserud and Leemans 1992).

1.1 Inertia in land use changes

Land use change models typically simulate periods of years to decades. Over these periods most land use does not change and accordingly models will also leave most cells in the same land use category. As a consequence the agreement between the simulated land use and the actual land use data will be high regardless of the accuracy of simulated changes. A high agreement therefore does not necessarily indicate an accurate model.

In this respect the interpretation of simulation results is fundamentally different from the assessment of remote sensing classification. Although both assess the accuracy of a land use map, remote sensing classification does not start from an initial land use map. For this reason you can use the absolute value of Kappa to assess the accuracy of remote sensing image classifications (Thomlinson et al. 1999), but not for the evaluation of model results.

The notion that the total amount of land use change is of importance for the interpretation of results of land use change models is discussed before. Pontius and Malanson (2005) use the initial land use map as a no change model to compare with simulation results. At the original resolution the no change model is typically more accurate than the simulation result, even though it doesn't simulate any change at all. They then aggregate all three land use maps until the resolution at which a simulation result is more accurate than the null model.

Pontius et al. (2008) elaborate further on the use of the initial land use map in the assessment of simulation results by introducing the "figure of merit". This statistic assesses the accuracy of land use changes rather than the all land use. The accuracy of land use changes is compared with the sum of land use changes that are only simulated, only in the actual data, or simulated and in the actual data. Still, the predictive value of a model (Pontius and Malanson 2005) cannot be derived from the absolute value of the figure of merit. Moreover this method naturally favors results that have few changes in the simulation, since incorrectly simulated land use changes account for two errors, one on the location of the actual land use change and one on the location of the incorrectly simulated land use change.

Alternatively Hagen-Zanker and Lajoie (2008) propose the use of neutral models in the evaluation of simulation results. Particularly they propose a "random constraint match" to create land use maps that simulates the correct amount of land use change, allocated randomly over the map with minimal adjustments to the original land use map. This method creates reference maps that can be used as a benchmark for the evaluation of simulation results. Simulation results should at least be more accurate than the reference map (see for example van Vliet et al. 2009). Still, this does not allow comparing results from different applications directly nor does it give a meaning to the absolute value of the accuracy.

This paper introduces $K_{Simulation}$, which expresses the accuracy of a simulation result relative to the accuracy that can be expected given the amount of land use change in the simulation and in the data. The next section first describes the Kappa coefficient of agreement in more detail. From there $K_{Simulation}$ is derived from the actual land use changes and the simulated land use changes. Section three illustrates the implication of $K_{Simulation}$ with a simple case first and then with calibration results from an actual land use change model. This calibration result is compared with the no change model and a neutral reference model, and the results are discussed. Section four finally draws conclusions and gives some directions for further research.

2. KAPPA COEFFICIENT OF AGREEMENT

The kappa coefficient of agreement is originally a statistic for discrete multivariate analysis. It expresses the agreement between two categorical datasets, corrected for the agreement as can be expected by chance, which depends on the distribution of class sizes in both datasets (Bishop et al.

1975). Kappa values range from 1, indicating a perfect agreement, to -1 indicating no agreement at all. The value 0 represents the special case where the agreement is exactly equal to the agreement as can be expected by chance. Since land use maps are essentially categorical data sets, Kappa is well suited to compare a pair of land use maps. Therefore it is commonly used for the accuracy assessment of results of land use change models. See for examples Barredo et al. (2004), or Kocabas and Dragicevic (2007).

Kappa statistics are referred to as contingency based metrics, since they can be computed from a contingency table between two maps.

Table 1 gives a generic form of a contingency table from the comparison of map A with map B. Land uses i are indicated as $L = \{1, 2, \dots, n\}$ and fields in the table indicate the fraction of cells that has a specific land use in map A and a specific land use in map B. For example the value for p_{12} indicates the fraction of cells that have land use 1 in map A and land use 2 in map B.

		Map B categories				total map A
		1	2	...	n	
Map A categories	1	p_{11}	p_{12}	...	p_{1n}	p_{1+}
	2	p_{21}	p_{22}	...	p_{2n}	p_{2+}

	n	p_{n1}	p_{n2}	...	p_{nn}	p_{n+}
total map B		p_{+1}	p_{+2}	...	p_{+n}	1

Table 1: generic form of a contingency table.

From the contingency table you can derive the observed fraction of agreement $P(O)$, the expected fraction of agreement subject to the distribution of class sizes $P(E)$, and the maximum fraction of agreement also subject to the distribution of class sizes $P(max)$:

$$P(O) = \sum_{i=1}^n p_{ii} \quad \text{Equation 1}$$

$$P(E) = \sum_{i=1}^n (p_{i+} * p_{+i}) \quad \text{Equation 2}$$

$$P(max) = \sum_{i=1}^n \min(p_{i+}, p_{+i}) \quad \text{Equation 3}$$

With these results you can compute the Kappa coefficient of agreement. Pontius (2000) indicates that Kappa values lower than 1 can be caused by two types of dissimilarity, the dissimilarity in sizes of land use classes and the dissimilarity in allocation of land uses on a map. Hagen (2002) defines these two components of the Kappa as $K_{Histogram}$ and $K_{Location}$. They are computed according to equations 4 to 6:

$$Kappa = \frac{P(O) - P(E)}{1 - P(E)} \quad \text{Equation 4}$$

$$K_{Histogram} = \frac{P(max) - P(E)}{1 - P(E)} \quad \text{Equation 5}$$

$$K_{Location} = \frac{P(O) - P(E)}{P(\max) - P(E)} \quad \text{Equation 6}$$

2.1 Accounting for land use persistence

As argued above, the absolute value of Kappa has no intrinsic meaning in assessment of results of land use change models, since the amount of land use change is not considered. For $K_{Simulation}$ the expected accuracy depends on the amount of land use change in reality and in the simulation model, hence more land use changes decreases the expected accuracy accordingly.

Therefore, this method requires information from the initial land use map as well as the simulated and actual land use map at T1. For the simulation result (A) as well as for the actual land use (B) the fraction of cells that has land use i , give that it has land use j in the initial map (O) can be expressed as $P(i^A | j^O)$ and $P(i^B | j^O)$ respectively. Since the original land use map (O) is the same for both maps, the expected similarity between the simulation result (map A) and the actual land use data (mapB) is a function of both:

$$P(E)_{simulation} = \sum_{j=1}^n P_j^O * \sum_{i=1}^n P(i^A | j^O) * P(i^B | j^O) \quad \text{Equation 7}$$

Where P_j^O represents the fraction of the initial map (O) that has land use j . The observed fraction of agreement as well as the maximum fraction of agreement can be computed according to equations 1 and 3. $K_{Simulation}$ as well as $K_{histogram}$ and $K_{location}$ can now be computed with equations 4 to 6, using $P(E)_{Simulation}$ instead of $P(E)$.

3. APPLICATION OF $K_{Simulation}$

A very simple example illustrates the implication of $K_{Simulation}$. Consider the land use maps A and B as shown in Figure 1. They represent the actual land use data at T0 and T1. The only land use change that appears is an expansion of the urban area on the right side of the existing urban area and the small patch of agriculture that turns in forest and vice versa.

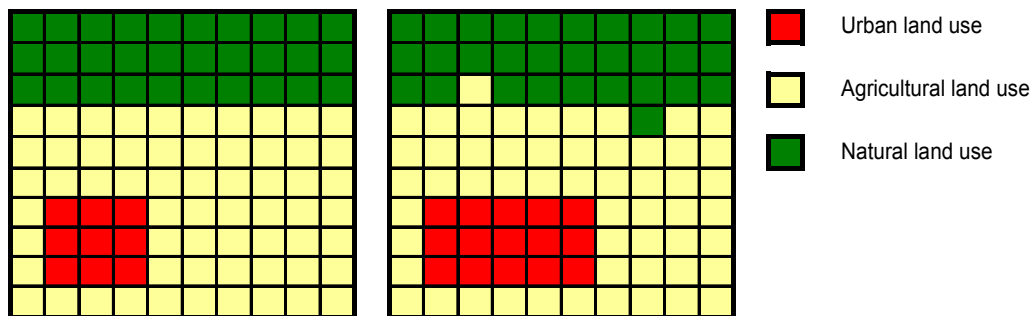


Figure 1: Map A, representing the actual land use at T0 (left map) and Map B representing the actual land use at T1 (right map).

Figure 2 below represents two model results for the same area. Map C shows the result of the no change model, and the result is therefore similar to the initial land use map. Map D shows the result of a hypothetical simulation model, which does include land use changes. The result of the simulation

model shows an expansion of the urban area, just as in the actual data, except that a large part is not allocated correctly.

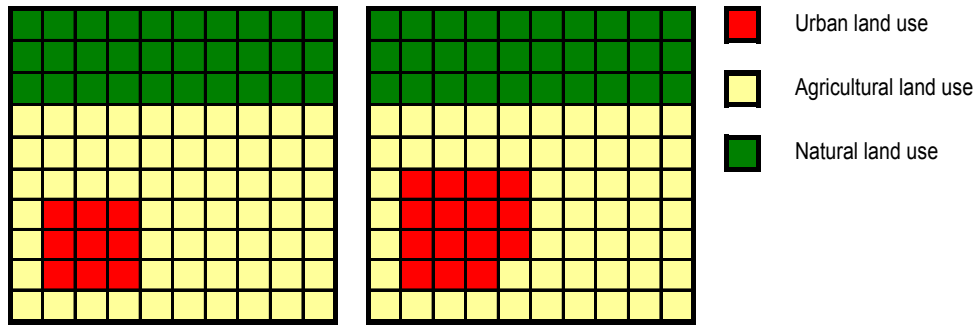


Figure 2: Map C, representing the result from the no change model T1 (left map) and Map D representing the result of a simulation model at T1 (right map).

Comparing the result maps with the actual land use data at T1 yields the results presented in Table 2. Kappa scores indicate that simulation result 1 is more similar to the actual land use at T1 than simulation result 2. Still, both scores are close to 1, which suggest that both models are very accurate. Moreover the $K_{location}$ and $K_{histogram}$ show more clearly what types of errors appear. Simulation result 1 deviates most in $K_{histogram}$, hence with respect to the size of the land use classes, and hardly with respect to $K_{location}$. This is in according to what can be observed on the maps; the largest deviation is that the urban expansion was not simulated at all, not so much the location of the land use changes. Simulation 2 on the other hand has a value of 1 for $K_{histogram}$, indicating exact similarity in terms of class sizes; hence all dissimilarity is caused by the incorrect allocation of land use changes as expressed with $K_{location}$.

comparison	Kappa	$K_{Location}$	$K_{Histogram}$
Model result 1 (Map C) - Actual land use (Map B)	0.857	0.960	0.893
Model result 2 (Map D) - Actual land use (Map B)	0.829	0.829	1.000

Table 2: Results for the similarity assessment, using Kappa.

Values for $K_{Simulation}$ are presented in Table 3. Note that $K_{Location}$ and $K_{Histogram}$ have different values as well, since they are recomputed with $P(E)_{Simulation}$. As can be seen from this table, it is not only the absolute value of Kappa that changed considerably, but also the ranking of the results. Using Kappa the score for the no change model was higher than for the simulation model while $K_{Simulation}$ shows the opposite. This indicates that the simulation model is a more accurate simulation of these land use changes.

comparison (given initial Map A)	$K_{Simulation}$	$K_{Location}$	$K_{Histogram}$
Model result 1 (Map C) - Actual land use (Map B)	0.000	0.000	0.25
Model result 2 (Map D) - Actual land use (Map B)	0.214	0.214	1

Table 3: Results for the similarity assessment, using $K_{Simulation}$.

Interpretation of $K_{Simulation}$ as the agreement, relative to the agreement as can be expected given the simulated and expected land use changes, Table 3 shows that the no change model does indeed exactly as good as can be expected by chance. This is in accordance with the reality: since the no change model does not simulate any changes, it cannot be accurate either. The positive value of $K_{Histogram}$ is caused by the small change in forest in the actual land use. The size of this change is correct, but the allocation is not.

For the second simulation however, $K_{Simulation}$ results are higher than zero, and therefore the model does indeed simulate land use changes better than what can be expected by chance. Further analysis shows that $K_{Histogram}$ is equal to 1 indicating that the class sizes correspond exactly between the simulation result and the actual land use data. Hence all discrepancies are due to the allocation of these changed land use cells.

3.1 Results from a simulation model

To further illustrate $K_{Simulation}$, it was used to assess the results of a land use change model for Western Europe. The modeled area comprises Ireland, Denmark, Germany, the Netherlands, Belgium, Austria and France. Data was taken from the Corine Land Cover Database (EEA, 2004), for which land use maps were available for the years 1990 and 2000 (CLC90 and CLC2k). Original CLC land uses were first reclassified into 17 classes: Natural vegetation, Agriculture, Residential, Industry and commerce, Tourism and recreation, Forest, Open spaces, Infrastructure, Port areas, Airports, Mineral extraction sites, Dump sites, Inland wetlands, Marine wetlands, Inland water, Marine water, and Beaches and dunes. Then reclassified land use maps were aggregated to cells of 1 km² each, using a majority aggregation.

For this case study land use change was simulated using three different models: a no change model, a random constraint model and a simulation model. The result of the no change model is similar to the 1990 land use map. The random constraint model simulates the correct amount of land use change but allocated randomly. Hence the size of the land use classes is by definition similar in the model result and the 2000 land use data, while the location of land uses is not (Hagen-Zanker and Lajoie 2008). The simulation model is an application of the Metronamica land use change model, which is a constrained cellular automata model (White and Engelen, 1997). In this model the total amount of cells per land use is defined exogenously, while their allocation is computed using the cellular automata algorithm. Therefore the size of the land use classes in the simulation result is also exactly matching the 2000 land use data.

Table 4 presents the Kappa scores for all three comparisons. Values are very close to 1 for all three comparisons, but highest for the no change model result. It may come as a small surprise that actually there were only a few land use changes recorded over the simulation period. Only 1.8 % of the cells change land uses over this period. Moreover, Table 4 shows that dissimilarities in the simulation model and the random constraint model are entirely due to the allocation of land uses, while the no change model has errors in both location and quantity.

comparison	Kappa	$K_{Location}$	$K_{Histogram}$
Null model result - CLC2k	0.969	0.980	0.989
Random constraint result - CLC2k	0.959	0.959	1
Simulation model result - CLC2k	0.960	0.960	1

Table 4: Accuracy of case study results assessed with Kappa statistics.

Table 5 presents the results as assessed with $K_{Simulation}$. Similar to the example above the ranking of the results differs from the assessment with Kappa in favor of the simulation result. Moreover, since scores are relative to 0, they allow for an interpretation. Both the random constraint match and the no change model have negative values; hence they do not have any predictive value. The result of the simulation model has a $K_{Simulation}$ value higher than 0, indicating that the model indeed has some

predictive power. Since sizes of land use classes are correctly defined in the constraint cellular automata model, all errors are due to allocation, as expressed by $K_{Location} < 1$.

comparison (given CLC90)	$K_{Simulation}$	$K_{Location}$	$K_{Histogram}$
Null model result - CLC2k	-0.278	-0.506	0.549
Random constraint result - CLC2k	-0.192	-0.192	1
Simulation model result - CLC2k	0.159	0.159	1

Table 5: Accuracy of case study results assessed with $K_{Simulation}$.

4. DISCUSSION AND CONCLUSIONS

Results of land use change models are often assessed from the similarity between actual land use data and simulation results, using statistics such as Kappa. However, this similarity depends heavily on the amount of land use changes over the simulation period as simulations with less change generally yield a higher accuracy. However, this should not be confused for a more accurate simulation of land use changes.

This paper presents $K_{Simulation}$, a statistic that expresses the agreement between a pair of maps, corrected for the agreement that can be expected, given the initial land use map. Results are therefore normalized so that a value of 0 indicates the similarity as can be expected by chance. Therefore values computed with $K_{Simulation}$ are comparable among different models, and indicate whether a model has any predictive capacities, either in the quantity of the land use changes, or in the allocation of these changes or in both. Moreover, by accounting for land use inertia this method avoids a false impression of accuracy that high Kappa scores suggest.

However, there are certainly cases where the proposed adjustment actually underestimates simulation accuracy, for example when land uses changes and then changes back as can be the case with crop shifting or forest harvest and reforestation cycles. These processes cannot be deduced from the comparison of the initial map with the actual land use data or the simulation result and hence those cases would indeed be accurate simulations not accounted for in this method.

Moreover, map comparison methods are very useful to assess the similarity of a set of land use maps on a pixel level. However, Hagen-Zanker and Martens (2008) argue that results from land use simulations can be compared on different scales: local, focal and global and with respect to both presence and structure. Practice shows that pixel based assessment is still the most commonly applied way to assess land use change models. Still patterns that emerge on the global scale are equally important characteristics of land use maps. Therefore $K_{Simulation}$ is a useful way to assess results of land use change models, but it should be complemented by methods that measure other map properties.

BIBLIOGRAPHY

- Barredo, J.I., Demichelli, L., Lavalle, C., Kasanko, M, and McCormick, N. (2004) Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria. *Environment and Planning B* 32: 65 – 84.
- Bishop Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) Agreement as a special case of association. *Discrete Multivariate analysis*. Cambridge MA, MIT press, 393 – 400.
- De Nijs, A.C.M., de Niet, R., and Crommentuijn, L. (2004) Constructing land use maps of the Netherlands in 2030. *Journal of Environmental Management* 72: 34 – 42.

- EEA (2004) Corine land cover database. <http://www.eea.europa.eu>. EEA, Copenhagen, 2004.
- Foody, G.M. (2002) Status of land cover accuracy assessment. *Remote Sensing of Environment* 80: 185 – 201.
- Hagen, A. (2002) Multi Method assessment of map similarity. In Ruiz, M., Gould, M., and Ramon, J. (ed.) *Proceedings of the 5th AGILE Conference on Geographic Information Science*: 171 – 182. Universitat de les Illes Balears, Palma, Spain.
- Hagen-Zanker, A. and Lajoie, G. (2008) Neutral models of landscape change as benchmarks in the assessment of model performance. *Landscape and Urban planning* 86 (3 – 4): 284 – 296.
- Hagen-Zanker, A. and Martens, P. (2008) Map comparison methods for comprehensive assessment of geosimulation models, In: Gervasi, O., Murgante, B., Lagana, A., Taniar, D., Mun, Y. and Gavrilova, M. (eds.) *Lecture Notes in Computer Science, Proceedings of the 2008 International Conference on Computational Science and Its Applications ICCSA 2008*: 194 – 209. Springer, Berlin, Germany.
- Kocabas, V. and Dragicevis, S. (2007) Assessing cellular automata model behaviour using a sensitivity analysis approach. *Computers, Environment and Urban systems* 30 (6), 921 – 953.
- Monserud and Leemans (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological modeling* 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 – 116.
- Pontius Jr., R.G. and Malanson, J. (2005) Comparison of the structure and accuracy of two land change models. *International Journal of Geographical Information Science* 19 (2): 243 – 265.
- Pontius Jr., R.G., Boersma, W., Castella, J.C., Clarke, K., de Nijs, T., Dietzel, C., Zengqiang, D., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Pijanowski, B., Pithadia, S., Sood, A.M., Sweeney, S., Trung, T.N., Veldkamp, A.T., Verburg, P.H., (2008) Comparing the input, output, and validation maps for several models of land change. *Annals of Regional Science* 42: 11 – 47.
- Power, C., Simms, A. and White, R. (2001) Hierarchical fuzzy pattern matching for regional comparison of land use maps. *International Journal of Geographical Information Science* 15 (1): 77 – 100.
- Refsgaard, J.C. and Henriksen, H.J. (2004) Modelling guidelines – terminology and guiding principles. *Advances in Water Resources* 27: 71 – 82.
- Rykiel, E.J.Jr. (1996) Testing ecological models: the meaning of validation. *Ecological Modelling* 90 (3): 229 – 244.
- Thomlinson, J.R., Bolstad, P.V., and Cohen, W.B. (1999) Coordinating methodologies for scaling landcover classifications from site-specific to global: steps toward validating global map products. *Remote Sensing of Environment* 70: 16 – 28.
- Van Vliet, J., Dragicevic, S. and White, R. (2009) Modelling urban growth in a variable grid cellular automaton. *Computers, Environment and Urban systems* 33 (1): 35 – 43.

White, R., & Engelen, G. (1997). Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B* 24: 235 – 246.