ACCURACY ASSESSMENT OF RETROSPECTIVE VEGETATION MAPS CONSIDERING SUCCESSION PATTERNS

CSERHALMI, D.[1*] – ERDŐS, L.[1]

[1]*Szent István University, Faculty of Veterinary Science, Department of Botany*
1078 Budapest, Rottenbiller u. 50.
(phone: +36-1-478-4235; fax: +36-1-478-4238)

[2]*Institute of Plant Sciences, University of Graz*
A-8010 Graz, Holteigasse 6, Austria.

*Corresponding author
e-mail: Cserhalmi.Daniel@aotk.szie.hu

Abstract. Accuracy assessment is an essential question in landscape analysis. Retrospective mapping is a special field of landscape ecology where ground control data is not always available thus mapping methods should use successive pairs of maps where historical maps are the reference data. However, the agreed methods of accuracy assessment, like error matrix and Kappa statistics, only measure the similarity of the maps but do not consider if land cover transitions are possible or not. This paper introduces a method that combines successional data with image change detection. Accuracy assessment had three major steps: (1) calculating error matrices and Kappa, (2) creating 'change maps', (3) identifying impossible changes using successional patterns. To prove that agreed methods can lead to false results, impaired maps were generated with increased distribution of incorrect changes. Results show that although the distribution of impossible changes was significantly higher, there was no change in observed agreement and Kappa value. The incorrectly mapped area increased during all periods. The cumulative error increased as well, suggesting that the reference map played a greater role in the incorrections of retrospective maps. Our paper suggests that considering succession patterns in retrospective mapping studies can lead to a higher accuracy.
Keywords: *aerial photo, error matrix, Kappa statistics, succession, peatland*

Introduction

It has been recognised that without adequate retrospective field data it's difficult to assess long-term change detection studies (Liira et al. 2010), but the rapid development of GIS and remote sensing techniques gives new opportunities in vegetation science. Aerial photos and satellite imagery serve retrospective mapping and are available from the early 1950s and before.

Multitemporal satellite imagery (Yuan et al. 2005, Langake et al. 2007) and satellite derived NDVI (George et al. 2006, Jamali et al. 2014) are common methods to evaluate historical land covers. Similar retrospective maps can be also generated from military survey maps (Skaloš et al. 2011) or aerial photographs (Kadmon and Harari-Kremer 1999, Hyvönen and Anttila 2006, Plieninger 2006). However, as ground control data was available the accuracy of the retrospective maps could be easily calculated and the level of accuracy was rather high.

To test the accuracy of a historical map, error matrix (Congalton and Green 1999), and Kappa statistics (Cohen 1960) are agreed methods (Stehman and Czaplewski 1998) but for long-term retrospective mapping the lack of ground reference data can lead to a limited application of these methods (Foody 2002).

Recent studies also suggest (Pontius and Milestone 2011, Foody 2011) that standard accuracy designs are not appropiate. These methods only measure the similarity of the maps, or even the changed area, but tell nothing about the quality of the changes. This means that results do not reveal whether the changes lead to natural succession, or they were the results effetcs of human activity.

Changes in land cover can be analysed with the more complex change detection error matrix, but the method cannot solve the problem of impossible changes which are the result of inaccurate georeference, or incorrect mapping (Cserhalmi and Biró 2012). Retrospective mapping disregarding impossible changes can result in a map that is incorrect from an ecological/botanical point of view, even though its overall accuracy is high. This means that the margins of vegetation patches can be correct even though the changes for the consecutive periods are impossible.

There are two explanations for impossible changes. First, changes that never occur in nature, and second, human impact (e.g. an industrial building at the site of a former grassland) but this can be regarded as a possible change when we reveal the history of the site. Thus in change detection analysis the key question is to separate possible changes from all of those that are not possible considering all circumstances.

Foody (2011) suggests that standard methods cannot be used for such change detection studies. New methods should be developed to improve historical retrospective maps by eliminating such thematic incorrections.

The main goal of this paper is to present a GIS based method that combines the agreed methods with vegetation and successional data to eliminate incorrect land cover change in retrospective studies. The study also would like to demonstrate that considering only accurate error matrices and strong agreement of Kappa statistics can still lead to a false biological result.

Materials and methods

Nyíres-tó mire is a *Sphagnum* dominated peatland, located in North-Eastern Hungary. The oxbow is approximately 750 meters long and has an average width of 80 meters. The first archive paper-based map was made by Simon (1960). There is no control map for an almost fifty-year period until Nagy (2006) published the first digital vegetation map of the area. As the land cover categories were different for the two maps, the digitalisation of the Simon map could not serve as a satisfactory control map.

For thematic harmonisation some plant associations of the Simon map (specially for grass types) were merged, like *Alopecuretum pratensis*, *Lolio-Plantaginetum* and *Agrostetum albae*, which have similar texture and cannot be separated based on aerial photos. Some plant communities are considered as sub-associations thus they were also merged, hence 8 classes were used instead of 14. However, we had to introduce two new classes, as *Rubo-Robinietum* and *Salicetum albae* were not present in the mire in the 1950s. Thus, a total of 10 categories was used.

Detailed methodology for retrospective maps was described previously (Cserhalmi et al. 2011). Six retrospective maps were available, with the following image acquisitions dates: 1956, 1966, 1975, 1988, 1997 and 2002 (Cserhalmi et al. 2010). As all images were taken in the vegetation period, the foliage was well developed and did not influence the mapping. The strongly different textures of the major land cover categories (tree dominated, shrub dominated, tall sedge, grass type, and raised bog) also helped the mapping process.

The current vegetation map served as a control for the first retrospective map (2002), which in turn was used as control for the second retrospective map (1997), and so forth. Using newer maps as controls, the following factors were considered to create thematically correct older maps: spatial distribution of associations on the newer map, climatic factors of the acquisition dates, possible succesional steps between the acquisition dates and visual interpretation, if needed.

Accuracy assessment had three major steps:

1. Calculating error matrices and Kappa statistics

2. Creating 'change maps' with successive pairs of retrospective maps

3. Indentifying impossible changes considering succession patterns, and creating 'impossible changes map'.

The first step followed the methods of Congalton and Green (1999) and Cohen (1960). To create 'change maps' a numeric thematic value was used for every single landcover category (plant association). Maps were evaluated with ArcView 3.1 using the following formula:

$C_{i,j} = (10 \times A_i) + A_j$

where $C_{i,j}$ is the "from-to" change category containing vegetation associations for both date i and date j, A_i is the thematic value of the association present in the cell at date i, and A_j is the thematic value of the association present in the cell at date j, and date A_i is older than A_j.

This formula delivered change categories whose first and second digits indicated the vegetation associations of the first (older) and second (newer) phases, respectively. This separated changed and non-changed areas without any botanical relevance. The formula works properly with 10 classes, a higher number of classes can lead to confusing results. With this method altogether five 'change maps' were evaluated.

As the area was protected for the study period, there was no direct, visible human impact on the vegation, thus using successional patterns was sufficient to evaluate changes. As they not take place in nature (e.g. acacia plantation turns to raised bog, alder carr changes to grass dominated habitats) they can be considered as the thematic incorrection of the retrospective maps.

For visualisation of impossible changes the above formula was used. 'Change maps' were reclassified with the help of succession patterns creating two new categories: (1) impossible changes and (0) correctly mapped area. This one was called 'impossible changes map'. For succession patterns we considered the works of Simon (1960), Nagy (2002) and Cserhalmi et al. (2011). The distribution of impossible changes was calculated and compared to the total area. 'Impossible changes maps' were also aggregated showing the cumulative error based on the interpreter.

To demonstrate the limitations of error matrix in retrospective land cover analysis, impaired vegetation maps were designed, where some of the land cover categories were changed to increase the distribution of impossible changes. These maps show that not taking into account successional patterns may result in maps that are incorrect from a biological perspective. Accuracy assessment followed the same steps mentioned above. Data were analysed with the statistical program R (R Development Core team, 2012) using Fisher's test and Bonferroni-Holm correction.

Results

As it was a retrospective analysis, the first period refers to the newest maps, the second for the older one, and so forth. *Table 1* shows the accuracy assessment of the original and the impaired retrospective maps supplemented with the distribution of impossible changes for each period.

Table 1. *Accuracy assesment of original and impaired (imp) retrospective maps of Nyíres-tó mire*

	Observed agreement	Chance agreement	Kappa	Impossible changes (%)	p-value	Odds ratio
2002-1997	0.70	0.21	0.62	9.04	9.9150e-05	3.47
2002-1997imp	0.71	0.19	0.63	18.25		
1997-1988	0.68	0.21	0.59	7.2	2.0750e-04	3.19
1997-1988imp	0.68	0.19	0.60	15.69		
1988-1975	0.67	0.23	0.57	5.37	9.9150e-05	4.07
1988mod-1975imp	0.61	0.21	0.51	16.23		
1975-1966	0.68	0.24	0.58	5.7	2.2785e-06	5.20
1975mod-1966imp	0.66	0.21	0.56	19.98		
1966-1956	0.78	0.30	0.68	6.34	6.6560e-06	4.58
1966mod-1956imp	0.75	0.21	0.68	16.97		

The observed agreement for the successive pairs of vegetation maps ranged between 0.67 and 0.78, which means that approximately 70% of the vegetation has not changed in each period. Kappa ranged between 0.57 and 0.68.

Considering succession patterns, the distribution of impossible changes ranged between 5.37% and 9.04% of the total area. This means that more than 90% of the area was correctly mapped. The rate of impossible changes decreased for the second (1997-1988) period, but the analysis still did not explain if the error was based on the incorrectly mapped patches of the 'control' map or if it was independent from it. However, it also did not reveal the spatial distribuiton of the incorrect patches.

Figure 1 shows the 'impossible changes map', which reveals the spatial distribution of incorrectly mapped patches. With the help of the maps, cumulative error can be calculated. Only 0.39% of the total area was mapped incorrectly for both the first and the second period (cumulative error), which means that incorrections recorded on the second retrospective map was independent from the errors of the control map, thus they are based mainly on the interpreter. This type of error was below 1% for all periods. Measuring the cumulative error for the whole study period, the total distribution where the same error was repeated once was 24.61%, where it was repeated twice was 4.52%, and the patches where the error was repeated three times covered only 0.07% of the total area.

Table 1 also shows the accuracy assessment for the impaired maps where the land cover categories were changed and successional patterns were not considered.

For the basic accuracy assessment methods, only the observed agreement was statistically tested as the other parameters were derived from this one. Observed agreement has not changed significantly during the whole study period. However, the statistical analysis revealed that the increased distribution of impossible changes was

significant for every period. The incorrectly mapped area was higher than 15%, and for one period it almost reached 20%.

The cumulative error for the first two periods increased from 0.39% to 8.01%. This suggests that the error of the reference map played a greater role in the total error of the second retrospective map. The analysis for the entire study period also confirmed this hypothesis as the total distribution where the same error was repeated once was 33.45%, wher it was repeated twice was 18.10%, and the patches where the error was repeated three times overed only 5.58% of the total area. However, there were some patches where the error was repeated four or five times respectively.
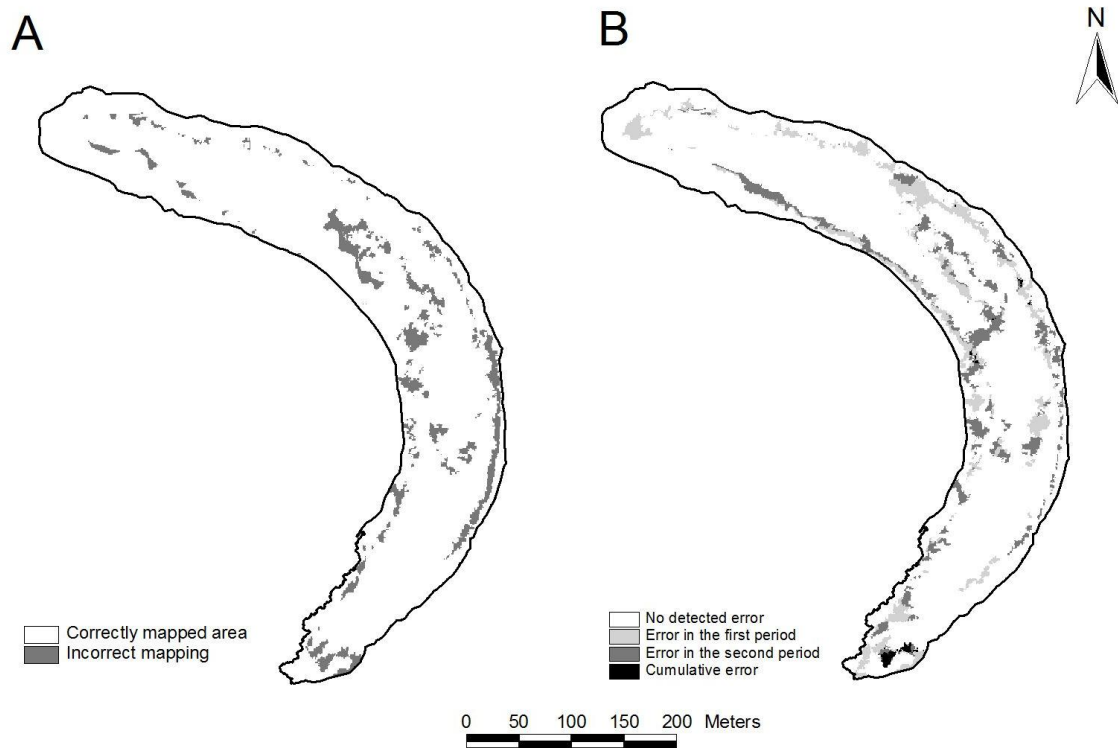

Figure 1. *'Impossible changes map' (A) and cumulative error between 1997 and 2002 (B)*

Discussion

The major accuracy assessment indicator (observed agreement) ranged between 0.6 and 0.78. However, higher agreement can only be reached with an accurate reference map (Rogan et al. 2013), which was not available for our study. On the other hand, the agreement is much lower for long-term change detection studies (Yuan et al. 2005), and it may be as low as about 70% for old photos (Skirvin et al. 2004). Even with the lack of accurate ground control data for retrospective maps the study resulted in a fair observed agreement for the whole study period. The same can be noted for Kappa statistics as well.

Our method that has been developed to reveal changes followed the idea of Biró et al. (2013). The original aim of which was to determine all of the possible landcover changes and from this database, users can separate impossible changes, which can be used in the accurcy assessment. The statistical analysis revealed that even if the distribution of impossible changes were higher, the basic metrics of accuracy

assessment were not changed significantly, and this can lead to a false botanical result. This also confirms the idea of Powell et al. (2004), who suggested that using a simple error matrix is not sufficient for accuracy assessment. Viera & Garret (2005) also state that low Kappa does not always mean low accuracy, which suggests that common methods have more limitations for retrospective analysis.

For the impaired vegetation maps, the increased rate of cumulative error showed that the incorrections of the 'control map' play greater role in accuracy assessment. Odds ratio showed that the chance to have patches that are incorrect from a botanical perspective was at least triple higher than for the original maps. This suggests that other factors should be considered besides of the classic method of accuracy assessment, particularly in retrospective studies, as error matrix and Kappa cannot bring sufficient results (Pontius and Millones, 2011). However, it seems that for long term vegetation analysis not the observed agreement plays a greater role, but the distribution and background of the changes.

Users also have a motivation not just to detect the error, but to understand its source (Bach et al. 2006). For change detection studies one can measure the difference with the error matrix but as succession can change land cover naturally, not every changed pixel can be considered an error, only those which transitions do not take place in nature, but the error matrix cannot separate them (Congalton and Green 1999). These errors can lead to misinterpretations in studies of land cover change (Foody 2011). The use of botanical and succesional data may offer a possible solution for the problem. This approach was suggested by Biró (2006) and Bergen & Dronova (2007) as a potential step towards increasing the accuracy of the maps.

One of the main limitations of the error matrix is that it cannot provide spatial distribution of the error (McGwire and Fisher 2001, Foody 2002), and there are only a few examples where spatial analysis of the measured error matrix accuracy was shown (Comber et al. 2012). *Figure 1* showed a possible way how the error can be revealed on a map. The method can help to find the 'hot spots' of incorrect mapping, thus it helps the producer to improve vegetation maps. This means that patches where cumulative (repeated) error was detected can be remapped, resulting in a more accurate retrospective map. The main principle of the 'impossible changes map' was similar to the difference image analysis introduced by Gupta el al. (2008), but our method is based more on the successional data.

Even though single-date accuracy assessment is typical, the multitemporal assessment is still not frequent in change detection studies (Olofsson et al. 2013). Just as the difference image analysis, the presented method shows a way to measure cumulative error thus this can give a chance to multi-date analysis as well. However, both methods suggest that interpretation of cumulative error can be difficult if too many acquisition dates were used.

As Shuman and Ambrose (2003) suggested, the combination of ground based techniques (biological knowledge) and digital techniques (GIS knowledge) can produce better results. The method presented here gives simple steps to separate thematical incorrections and impossible changes on a time series as well, which can be performed with ArcView, ArcGIS or QGIS. The methodology can be adapted to other sites, where the texture of the vegetation is different and the number of caterories does not exceed 10. The only essential step is the knowledge of the local land cover transitions. This method may support retrospective studies where accurate historical data are not available.

REFERENCES

[1]   Bach, M. Breuer, L., Frede, H. G., Huisman, J. A., Otte, A., Waldhardt, R. (2006): Accuracy and congruency of three different digital land-use maps. - Landscape and Urban Planning 78: 289–299.

[2]   Bergen, K. M.. & Dronova, I. (2007): Observing succession on aspen-dominated landscapes using a remote sensing-ecosystem approach. Landscape Ecology, 22(9): 1395-1410.

[3]   Biró, M. (2006): Történeti vegetációrekonstrukciók a térképek botanikai tartalmának foltonkénti gazdagításával. - Tájökológiai Lapok, 4(2): 357-384.

[4]   Biró, M., Szitár, K., Horváth, F., Bagi, I., Molnár, Z. (2013): Detection of long-term landscape changes and trajectories in a Pannonian sand region: comparing land-cover and habitat-based approaches at two spatial scales. - Community Ecology, 14(2): 219-230.

[5]   Cohen, J. (1960): A coefficient of agreement for nominal scales. - Educational and Psychological Measurement 20: 37–46.

[6]   Comber, A., Fisher, P. Brunsdon, C., Khmag, A. (2012): Spatial analysis of remote sensing image classification accuracy. - Remote Sensing of Environment 127: 237–246.

[7]   Congalton, R., Green, K. (1999): Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, CRC/Lewis Press, Boca Raton, FL

[8]   Cserhalmi, D., Biró, M. (2012): Retrospektív vegetációtérképek elemzési lehetőségei. - Kitaibelia 17(1): 16.

[9]   Cserhalmi, D., Nagy, J., Kristóf, D., Neidert, D. (2010): The recontructions of vegetation change in Nyíres-tó mire: an image segmentation study. – Acta Botanica Hungarica 50(1-2): 89-102.

[10]   Cserhalmi, D., Nagy, J., Kristóf, D., Neidert, D. (2011): Changes in a wetland ecosystem: A vegetation change reconstruction study based on historical panchromatic aerial photographs and succession patterns. - Folia Geobotanica 46: 351-371.

[11]   Foody, G. (2011): Classification accuracy assessment. - IEEE Geoscience and Remote Sensing Society Newsletter 159: 8–14.

[12]   Foody, G. M. (2002): Status of land cover classification accuracy assessment. - Remote Sensing of Environment 80: 185–201.

[13]   George, C. Rowland, C., Gerard, F., Balzter, H. (2006): Retrospective mapping of burnt areas in Central Siberia using a modification of the normalised difference water index. - Remote Sensing of Environment 104: 346–359.

[14]   Gupta, R. P., Kanungo D. P., Arora, M. K. Sarkar, S. (2008): Approaches for comparative evaluation of raster GIS-based landslide susceptibility zonation maps. - International Journal of Applied Earth Observation and Geoinformation 10: 330–341.

[15]   Hyvönen, P., Anttila, P. (2006): Change detection in boreal forests using bi-temporal aerial photographs. - Silva Fennica 40(2): 303–314.

[16]   Jamali, S., Seaquist,J., Eklundh, L., Ardö, J. (2014): Automated mapping of vegetation trends with polynomials using NDVI imagery over the Sahel. - Remote Sensing of Environment 141: 79–89.

[17]   Kadmon, R., Harari-Kremer, R. (1999): Studying Long-Term Vegetation Dynamics Using Digital Processing of Historical Aerial Photographs. - Remote Sensing of Environment 68: 164–176.

[18]   Landis RJ , Koch GG. The measurement of observer agreement for categorical data. - Biometrics 1977; 33:159-174

[19] Langanke, T., Burnett, C., Lang, S. (2007): Assessing the mire conservation status of a raised bog site in Salzburg using object-based monitoring and structural analysis. - Landscape and Urban Planning 79: 160–169.

[20] Liira, J., Feldmann, T., Mäemets, H., Peterson, U. (2010): Two decades of macrophyte expansion on the shores of a large shallow northern temperate lake - A retrospective series of satellite images. - Aquatic Botany 93: 207–215.

[21] McGwire, K. C., Fisher, P. (2001): Spatially variable thematic accuracy: Beyond the confusion matrix. In Hunsaker, C. T., Goodchild, M. F., Friedl, M. A., Case, T. J. (eds.) Spatial uncertainty in ecology: Implications for remote sensing and GIS applications, Springer–Verlag, New York.

[22] Nagy, J. (2002): Szündinamikai vizsgálatok egy tőzegmohaláp természeti értékeinek megőrzésére. Ph.D. dolgozat. Szent István Egyetem, Biológiai Doktori Iskola, Gödöllő

[23] Nagy, J. (2006): Vegetation maps of Báb-tava, Navat-patak, Nyíres-tó, Zsid-tó, Kis-tó, Nagy-tó and Hamvas-tó. The detailed description of their plant communities with the list of the plant species. In: Magura, T. (ed.) „Complex habitat rehabilitation of the Central Bereg Plain, Northeast Hungary." Archives of HNP Directoriate, Debrecen.

[24] Olofsson, P., Foody, G. M., Stehman, S. V., Woodcock, C. E. (2013): Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. - Remote Sensing of Environment 129:122–131.

[25] Plieninger, T. (2006): Habitat loss, fragmentation, and alteration – Quantifying the impact of land-use changes on a Spanish dehesa landscape by use of aerial photography and GIS. - Landscape Ecology 21: 91–105.

[26] Pontius, R. G., Millones, M. (2011): Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. - International Journal of Remote Sensing 32(15): 4407–4429.

[27] Powell, R. L., Matzke, N., de Souza Jr., C., Clark, M., Numata, I., Hess, L. L., Roberts, D. A. (2004): Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. - Remote Sensing of Environment 90: 221–234.

[28] R Development Core Team (2012): A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna. Available from: http://www.R-project.org/ (Sept 2014)

[29] Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., Fischer, L. (2013): Land-Cover Change Monitoring with Classification Trees Using Landsat TM and Ancillary Data. - Photogrammetric Engineering & Remote Sensing 69(7): 793–804.

[30] Shuman, C. S., Ambrose, R. F. (2003): A Comparison of Remote Sensing and Ground-Based Methods for Monitoring Wetland Restoration Success. Restoration Ecology 11(3): 325–333.

[31] Simon, T. (1960): Die Vegetation der Moore in den Naturschutz-gebieten des Nördlichen Alföld. - Acta Botanica Hungarica 6(1-2): 107-137.

[32] Skalos, J., Weber, M., Lipský, Z., Trpáková I., Santrucková, M., Uhlírová, L., Kukla, P. (2011): Using old military survey maps and orthophotograph maps to analyse long-term land cover changes e Case study (Czech Republic). - Applied Geography 31: 426-438.

[33] Skirvin, S. M., Kepner, W. G., Marsh, S. E., Drake, S. E., Maingi, J. K., Edmonds, C. M., Watts, C. J., Williams, D. R. (2004). Assessing the accuracy of satellite-derived land-cover classification using historical aerial photography, digital orthophoto quadrangles, and airborne video data. - Remote sensing and GIS accuracy assessment 9: 115-131.

[34] Stehman, S. V., Czaplewski, R. L. (1998): Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. - Remote Sensing of Environment 64: 331–344.

[35] Viera, A. J. & Garrett, J. M. (2005): Understanding interobserver agreement: the kappa statistic. - Family Medicine, 37(5): 360-363.

[36] Yuan, F., Sawaya, K. E., Loeffelholz, B. C., Bauer, M. E. (2005): Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. - Remote Sensing of Environment 98: 317 – 328.