On Reliability of Remote Sensing Data and Classification Methods for estimating transition rules of the land-use Cellular Automata

Yulia Grinblat*[1, 2], Michael Gilichinsky[3] and Itzhak Benenson*[1]

[1]Department of Geography and Human Environment, Tel Aviv University, Tel-Aviv,
[2]The Porter School of Environmental Studies, Tel Aviv University, Tel-Aviv
[3]Elbit Systems, Rehovot

*Corresponding author: juliagri@post.tau.ac.il, bennya@post.tau.ac.il

Abstract

Typically, the Cellular Automata (CA) models of Land-Use/Land-Cover (LULC) changes focus on estimating the rules of the LULC changes and analysis of the simulation results. However, the models put aside the uncertainty of the LULC maps that are used for establishing the transition rules. Our study questions the reliability of Remote Sensing (RS) data sources and classification methods applied for constructing these maps.

Based on four time intervals within a 36-year period, we construct LULC maps and estimate the transition probabilities between six LULC states. The LULC maps and transition probabilities matrices (TPM) are built based on the manual interpretation of high-resolution aerial photos and classification of multispectral Landsat images for the same years.

We consider the maps and TPM derived from the aerial photos as a reference data, and compare them to those constructed from the Landsat images classified with several methods: mean-shift segmentation algorithm followed by Random Forest classification method, and three pixel-based methods of classification: K-means, ISODATA, and maximum likelihood. Then, for each classification the TPM were compared to the referenced TPM.

The accuracy assessment of all maps obtained with the pixel-based methods is insufficient for estimating the LULC TPM. The LULC map obtained with the object-based classification method fit well to that based on the aerial photos, but the estimates of TMP are qualitatively different from those constructed from the aerial photos.

This article raises doubts regarding the adequacy of Landsat data and standard classification methods for establishing LULC CA model rules, and calls for the careful reexamination of the entire land-use CA framework.

Keywords

Cellular automata, Landsat images, land-use/land-cover changes, Markov transition probabilities matrices, validation of RS classification methods

I INTRODUCTION

Conceptual simplicity and the ability of explicit representation of landscapes and their changes make Cellular Automata (CA) a standard tool for simulating urban and regional land-use dynamics (Clarke et al. 1996; White and Engelen 1997), which potential for modeling Land-Use/Land-Cover (LULC) dynamics is widely recognized (Wu and Webster 1998; Pijanowski et al. 2002; Verburg et al. 2002). The major source of data for the CA modeling is Remote Sensing (RS) multispectral imagery classified for establishing land uses and covers.

It is often reported that the CA models are quite successful in predicting LULC changes, with the high overall fit (80-90%) between the real LULC data and model outputs. This is indeed true when the validation is based on comparing the *entire modeled area*. However, the period of time covered by the CA model is, usually, between one and few decades and the fraction of the modeled area that has been changed during such a period is, typically, few percent of the entire city area. As far as initial area is excluded from the comparison, the spatial fit between the predicted and real *changes* drops down (Hagen-Zanker et al. 2005; Pontius and Petrova 2010).

A hierarchy of reasons of limited capacity of the CA models for predicting LULC changes can be proposed: (1) CA framework as a whole is insufficient for predict LULC dynamics, due to, say, essential part of human bounded rational decisions in land planning and management; (2) The CA framework works, but wrong CA rules are chosen; (3) The CA framework works, the rules are properly established, but the data chosen for estimating parameters of the rules do not represented the real of the LULC changes. In this paper we deal with the latter and investigate the adequacy of the RS data for calibration and validation of the CA models.

II Testing the adequacy of classifications methods

Strangely enough, the adequacy of the RS classification for representing LULC *changes* remains on the margin of the CA modeling studies. Despite a series of publications that regard the erroneous consequences of misclassification (van de Voorde et al. 2009; van der Kwast et al. 2009) and sensitivity of the CA dynamics to the parameters of the CA rules (Liu and Andersson 2004; Jantz and Goetz 2005; Dietzel and Clarke 2006), the majority of modeling studies carelessly exploit the simplest methods of the RS images classification, take their outputs for granted, and focus on model calibration. This may evidently result in inadequate transition rules regardless of the calibration methods. The standard source of data for the CA model calibration and validation is 30m resolution LANDSAT multispectral imagery. We thus investigated the adequacy of different methods of LANDSAT images classification for establishing CA model rules.

The background of the CA model is Transition Probability Matrix (TPM) $\{p_{ij}\}$ - a set of probabilities, per time unit, of transition $S_i \rightarrow S_j$ between the states S_i and S_j of the LULC CA. Our study compares TPMs estimated based on the LANDSAT maps obtained by the different classification methods to the ground truth – the TPM that is estimated based on the manual interpretation of high-resolution aerial photos of the same area.

III Study area and Data preparation

The experimental area is the 15x6 km transect that starts in the center of the city of Netanya, Israel, and extends to surrounding agriculture areas. The period of comparison 1972 – 2008 (36 years) is divided into 4 intervals of 6 - 11 years, depending on availability of the LANDSAT images and aerial photos. Based on the manual interpretation of the high-resolution aerial photos, we have constructed the maps of Netanya LULC dynamics of LULC states. Six LULC states are considered: built-up areas (BU), roads (RD), agricultural (AG) and vegetation (VG) areas, open spaces (OS) and water surfaces (WA). In this short paper we present the results aggregated into three states only - – built-up (BU), agriculture (AG) and the rest states (RE) that aggregate the rest four LULC states.

Four popular in the CA studies pixel-based methods and one object-based method were applied for classifying the LANDSAT images. To remind, pixel-based classification methods consider pixels individually, while object-based methods recognize spatially continuous homogeneous domains of pixels, and then assign a land use to these segments (Lu and Weng 2007). All exploited pixel-based methods are traditional first choice of a CA modeler: K-means, ISODATA, Maximum Likelihood (ML) and hybrid classification. The object-based method we apply is two-staged: mean-shift clustering segmentation (Comaniciu and Meer 2002) is followed by a Random Forest classification (Breiman 2001).

IV The results

As can be seen in Figure 1 the fit between the LANDSAT-based maps and the map that is based on manual classification varies depending on the method. We do not present here the results of numerical analysis of this fit and focus on comparison of transition probabilities matrices. For this comparison, the TPMs all estimated for the time periods of different length, were normalized to the 10-year period.

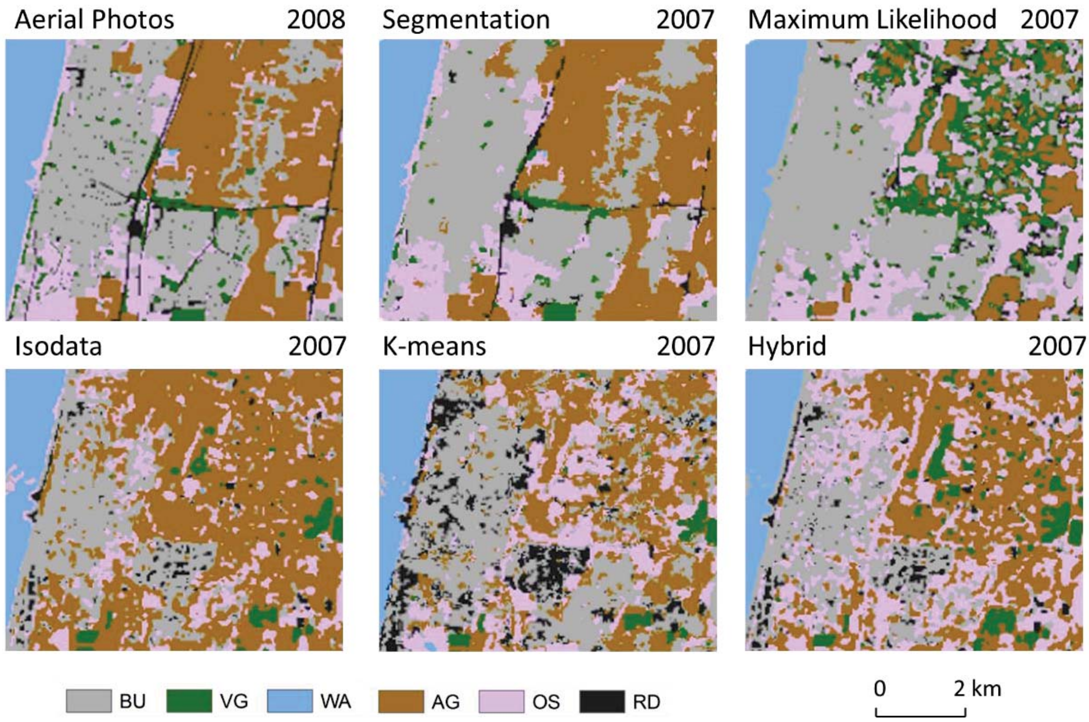

Figure 1. Land-use maps for the part of the study area obtained with the manual classification of the aerial photos and with several classification methods applied to the LANDSAT images

Table 1 presents the TPMs obtained with two of the applied methods, Maximum Likelihood and Segmentation, together with the TPM obtained for the manually classified map. Due to limited space, the TPMs are presented for the LULC uses aggregated into three classes and for the last time interval of the investigated period - 2000-2007, only. The aggregated land uses classes are as follows: BU - built-up areas and roads; AG - agricultural areas; RE –the rest of land uses - open spaces, vegetated areas and water surfaces.

		Maximal Likelihood				**Segmentation**				**Aerial photos**			
		BU	**AG**	**RE**	**Tot**	**BU**	**AG**	**RE**	**Tot**	**BU**	**AG**	**RE**	**Tot**
Transition probability	**BU**	**0.75**	0.05	0.20		**0.78**	0.13	0.09		**0.99**	0.00	0.01	
	AG	0.14	**0.64**	0.22		0.08	**0.86**	0.06		0.02	**0.95**	0.03	
	RE	0.32	0.09	**0.58**		0.13	0.13	**0.74**		0.07	0.10	**0.83**	
Area of transition, km^2	**BU**	**7.10**	0.46	1.90	9.50	**16.9**	2.80	1.8	21.5	**19.2**	0.02	0.11	19.3
	AG	5.50	**8.90**	25.9	40.3	3.7	**40.6**	3.0	47.3	0.87	**41.2**	1.30	43.4
	RE	11.9	3.30	**21.4**	36.6	2.3	2.20	**13.2**	17.7	1.70	2.30	**19.9**	23.9

Table 1. TPM for transitions between three land uses based of the 2000-2007(LANDSAT) / 1999-2008(Aerial photos) data. Probabilities are normalized to the 10 year period

As can be seen from Table 1, for the presented period, the TPMs obtained with the ML and Segmentation methods are qualitatively and quantitatively different from the TPM estimated based on the aerial photos. Most important, in reality, LULC states are changing in time essentially less frequently than it is obtained based on the RS images classified with the ML method; for example, in reality, the probability of the AG→AG transition per 10 years is 0.95, while according to the ML map this probability is 0.64 only. The same is true for the rest of the pixel-based methods (not presented here). The TPM obtained with the Segmentation method is much closer to the reference TPM that the TPM of the pixel-based methods, but yet essentially biased towards more changes than in reality. Similar differences are characteristic of all other periods, as well as for the TPMs constructed for the basic, non-aggregated, set of the LULC states.

V Conclusions

We thus conclude that none of the maps obtained, based on the LANDSAT images, with the help of the popular pixel-based classification methods can be exploited for establishing CA transition rules. Object-based method provided better, but yet insufficiently precise estimates. We call for the revision of approach to the CA calibration and validation. An open depository of high-resolution, carefully validated, long-term series of the land-use/cover maps that reflect different types of LULC dynamics, and represent different types of land planning systems for different periods of population growth and economic development should be established. Instead of establishing a new database for every new CA model, one has to use these data series for calibration and validation of her/his new model. Only then, the model can be applied to the new dataset which, as we have demonstrated, must be constructed with the great care.

References

Breiman, L. 2001. Random Forests. *Machine Learning* 45 (1):5-32.

Clarke, K. C., S. Hoppen, and L. Gaydos. 1996. A self-modifying cellular automata model of historical urbanization in the San Franciso Bay area. *Environment and Planning B: Planning and Design* 24:247 - 261.

Comaniciu, D., and P. Meer. 2002. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (5):603-619.

Dietzel, C., and K. Clarke. 2006. The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Computers, Environment and Urban Systems* 30 (1):78-101.

Hagen-Zanker, A., J. van Loon, A. Maas, B. Straatman, T. de Nijs, and G. Engelen. 2005. Measuring performance of land use models: An evaluation framework for the calibration and validation of integrated land use models featuring Cellular Automata. Paper read at 14th European Colloquium on Quantitative Geography, September 9-13, at Tomar, Portugal.

Jantz, C. A., and S. J. Goetz. 2005. Analysis of scale dependencies in an urban land-use-change model. *International Journal of Geographical Information Science* 9 (2):217-241.

Liu, X., and C. Andersson. 2004. Assessing the impact of temporal dynamics on land-use change modeling. *Computers, Environment and Urban Systems* 28 (1–2):107-124.

Lu, D., and Q. Weng. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28 (5):823-870.

Pijanowski, B. C., D. G. Brown, B. A. Shellito, and G. A. Manik. 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems* 26 (6):553-575.

Pontius Jr, R. G., and S. H. Petrova. 2010. Assessing a predictive model of land change using uncertain data. *Environmental Modelling & Software* 25 (3):299-309.

[ISBN: 978-2-9105-4510-5]

van de Voorde, T., F. Canters, J. van der Kwast, G. Engelen, M. Binard and Y. Cornet. 2009. Quantifying intra-urban morphology of the Greater Dublin area with spatial metrics derived from medium resolution remote sensing data. *2009 Joint Urban Remote Sensing Event*, Shanghai: 1-7.

van der Kwast, J., I. Uljee, G. Enegelen, T. Van de Voorde, F. Canters, and C. Lavalle. 2009. Using Remote Sensing Derived Spatial Metrics for the Calibration of Land-Use Change Models. Paper read at 7th International Urban Remote Sensing Conference.

Verburg, P. H., W. Soepboer, A. Veldkamp, R. Limpiada, and V. Espaldon. 2002. Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. *Environmental Management* 30 (3):391 - 405.

White, R., G. Engelen, and I. Uljee. 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design* 24 (3):323-343.

Wu, F., and C. J. Webster. 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B: Planning and Design* 25 (1):103-126.