Title
What are the Probabilities of Land-Use Transitions? The Answer Depends on the Classification Method

Journal
International Conference on GIScience Short Paper Proceedings, 1(1)

Authors
Grinblat, Yulia
Gilichinsky, Michael
Itzhak, Itzhak

Publication Date
2016

DOI
10.21433/B3111b9432ds

Peer reviewed

What are the Probabilities of Land-Use Transitions? The Answer Depends on the Classification Method

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

[1]Department of Geography and Human Environment, Tel Aviv University, Tel-Aviv
Email: juliagri@post.tau.ac.il, bennya@post.tau.ac.il

[2]The Porter School of Environmental Studies, Tel Aviv University, Tel-Aviv

[3]Elbit Systems, Rehovot
Email: gilichinsky@gmail.com

Abstract

Based on four time intervals within a 36-year period, we construct Land-Use/Land-Cover (LULC) maps and estimate the transition probabilities between six LULC states: built-up, agriculture, green and open spaces, transportation, and water surfaces. The LULC maps and transition probabilities matrices (TPM) were built based on the manual classification of high-resolution aerial photos and multispectral Landsat images for the same years.

We considered the maps and TPM constructed from the aerial photos as a control, and compared them to those constructed from the Landsat images classified with several methods: mean-shift segmentation followed by Random Forest classification methods, and three pixel-based methods of classification: K-means, ISODATA, and maximum likelihood. For each classification the TPM were compared to the TPM constructed from the aerial photos.

The goodness of fit of all maps obtained with the pixel-based methods was 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 were 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.

1. 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. Typically, the CA models 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.

The major source of data for the CA modeling is Remote Sensing (RS) multispectral imagery classified for establishing LULC dynamics. It is often reported that the CA models are quite successful in predicting LULC, with the high overall fit (80-90%) between the real LULC and model outputs. This is indeed true when the validation is based on comparing the *entire modeled area*. However, 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; (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.

2. Testing the adequacy of classifications methods

The adequacy of the RS classification for representing LULC *changes* remains on the margin of the CA modeling studies. The modeling studies carelessly exploit 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.

LANDSAT imagery is a common choice of RS data and we investigated the adequacy of different methods of their classification for establishing CA model rules.

1.1 LULC Transition Probability Matrix

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 TPM that is estimated based on the manual interpretation of high-resolution aerial photos of the same area.

1.2 Experimental area and Remote Sensing Data

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 six LULC states: 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 two states only – Built-up (BU) and the NB-state that aggregates the rest five LULC states.

1.3 Classification Methods

In parallel to the manual interpretation, four pixel-based methods and one object-based method were applied for classifying these LULC on the LANDSAT imagery. 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 is followed by a Random Forest classification.

2. The results

The fit between the LANDSAT-based maps and the map that is based on manual classification varies depending on the method. Segmentation and maximum likelihood methods represent better results of LANDSAT classification (Figure 1). All the rest methods showed low values of accuracy.

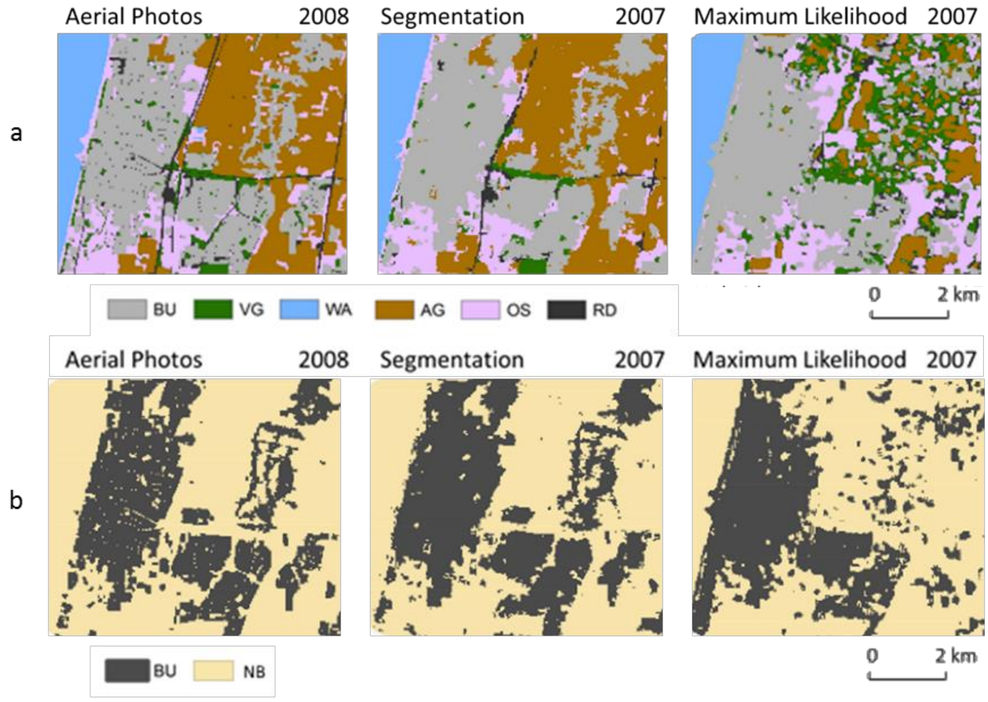

Figure 1. Land-use maps of 2007/2008 for three of six investigated methods, part of the study area: (a) all six LULC states and (b) aggregated BU and NB states

For each observation period we constructed TPM normalized to the 10-year period and compared them to the TPM constructed based on the manual interpretation (Table 1). Due to limited space, the TPMs are presented for the LULC uses aggregated into two classes: BU - built-up areas; NB – non-built-up areas, which include agricultural and vegetated areas, open spaces, water surfaces and roads.

Table 1. Average TPM for the probabilities normalized by the 10 year period

		LULC at year t + 10											
		Aerial Photos		Segmentation		ML		Hybrid		K-means		Isodata	
		BU	NB	BU	NB	BU	NB	BU	NB	BU	NB	BU	NB
LULC at year t	BU	**0.98**	0.02	**0.81**	0.20	**0.58**	0.42	**0.42**	0.58	**0.45**	0.55	**0.42**	0.58
	NB	0.05	**0.95**	0.08	**0.92**	0.11	**0.89**	0.08	**0.92**	0.16	**0.84**	0.19	**0.81**

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 BU→BU and NB→NB transitions per 10 years are close to 1, while according to the ML map these probabilities are 0.58 and 0.89. The fit is even worse for the rest of the pixel-based methods.

The TPM obtained with the Segmentation method fits to the TPM for the aerial photos better than the TPM of the pixel-based methods, but is yet essentially biased.

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

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.

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.