Computers, Environment and Urban Systems 64 (2017) 204–214

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Spatial and temporal modeling of parcel-level land dynamics

Emre Tepe [a], Jean-Michel Guldmann [b,*]

[a] Department of City and Regional Planning, Cankaya University, Cankaya, 06530 Ankara, Turkey
[b] Department of City and Regional Planning, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 14 August 2016
Received in revised form 25 February 2017
Accepted 25 February 2017
Available online xxxx

Keywords:
Urban dynamics
Spatio-temporal modeling
Land use change
Urban growth
Autologistic regression

ABSTRACT

Neighborhood and historical conditions are important factors in land dynamics. However, models that explicitly incorporate spatial and temporal dependencies face challenges in data availability, methodology and computation. In this research, parcel-level dynamics are investigated using the geocoded Auditor's tax database for Delaware County, Ohio, including 73,560 parcels over the period 1990–2012. A binary spatio-temporal autologistic model (STARM), incorporating space and time and their interactions, is used to investigate parcel-level dynamics. The results show that the model is able capture the impacts of contemporaneous and historical neighborhood conditions around parcels, as well as the effects of other variables such as distances to various facilities and infrastructures, agricultural and residential land-use shares within a half mile radius circle, and population density and growth expectation at the census tract level.

1. Introduction

The land development process alters land uses over space and time, depending on parcel characteristics, neighborhood structures, historical trends, available services and amenities, infrastructures, socio-economic factors, zoning, and other policies. The goal of this research is to improve our understanding of the determinants of this process, using a dynamic model that incorporates both spatial and temporal dependencies.

In theory, land-use change takes place when the expected benefits from an alternative land use are larger than the benefits from the current one after accounting for one-time conversion costs. While conversion costs are instantly incurred, benefits are collected over the long term, thus inhibiting changes once land is developed. Generally, these costs and benefits are not directly observable. However, other observable data may be used as proxies, including site, locational and socio-economic characteristics, which have been incorporated in past land-use change models. There is a growing body of research on land development at a disaggregated level, as a result of increasing availability of spatially-referenced parcel-level data, with extensive use of statistical and simulation models. These studies highlight the importance of both spatial and temporal dynamics, as current and historical conditions of parcel neighborhoods may be influential. However, modeling these effects increases the size of the necessary datasets and the complexity of computations.

This research investigates parcel-level dynamics, using the geocoded Auditors' tax database for Delaware County, Ohio. In contrast to earlier research using time series of remote-sensing and land-cover data to derive measures of urban dynamics, information on the year when construction took place on each parcel is used to derive these dynamics. A binary spatio-temporal autologistic regression model (STARM), incorporating space and time and their interactions, is used to capture the impacts of contemporaneous and historical neighborhood conditions around parcels. It is a modified version of the autologistic model introduced by Zhu, Zheng, Carroll, and Aukema (2008). Because methods for estimating the parameters of binary STARM models are not available in commercial or open-source software, a dedicated program has been written in Python to estimate Monte Carlo Maximum Likelihood parameters, and parallel processing techniques used to process a very large dataset. The model successfully quantifies the impacts of current and historical neighborhood conditions on land dynamics. It is validated with a high level of accuracy and used to generate a robust forecast of land-use change.

The remainder of the paper is organized as follows. Section 2 critically reviews the relevant literature. The modeling methodology is presented in Section 3. The data are described in Section 4. Model results are presented in Section 5. Model validation and forecasts are discussed in Section 6. Section 7 concludes and presents areas for further research.

2. Review of the literature

While there are alternative approaches to modeling land-use change, such as cellular automata and agent-based simulation models, the focus of this review is on statistical modeling approaches. Irwin and Geoghegan (2001) suggest that the maximization of the expected

* Corresponding author at: Department of City and Regional Planning, The Ohio State University, 275 West Woodruff Avenue, Columbus, OH 43210, USA.
E-mail addresses: emretepe@cankaya.edu.tr (E. Tepe), guldmann.1@osu.edu (J.-M. Guldmann).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.02.005
0198-9715/

utility from land is the theoretical basis of most econometric models of urban land-use change.

Discrete response models have been commonly used, where the probability of land development is conditional on a given set of factors. Wear, Turner, and Flamm (1996) estimate the probability of land being classified as potential timberland at the county level. Carrion-Flores and Irwin (2004) operationalize the utility-maximization behavior of land-owners in land-use change decisions by using a binary response model at the parcel level. Carrion-Flores, Flores-Lagunes, and Guci (2009) modify this model to incorporate multiple land-use choices, using a multinomial discrete choice model. Chomitz and Gray (1996) estimate a multinomial logistic regression at 1 km rectangular grid to predict the occurrence of each of the following states: natural vegetation, semi-subsistence agriculture, and commercial farming. Hardie and Parks (1997) introduce a multinomial model to estimate land-use proportions at the county level. The UrbanSim model (Waddell, 2002), which generates short-term predictions of land development, involves a multinomial logistic model to predict land-use transformations at grid cell of 150 by 150 m.

Recent advances in GIS technology have helped incorporate spatial features into models. Spatial regression models have been used to deal with spatial autocorrelation and avoid inconsistent estimates with misleading statistical inferences. A neighborhood assumption is the most common approach in these models, using a spatial weight matrix. Chomitz and Gray (1996) investigate the conversion of forest land to agricultural land, based on remote-sensing data and other spatial data. They account for the endogeneity of accessibility and population, because road construction is influenced by agricultural production, and government policies that control population distribution may have an impact on forest conversion to agriculture. While accounting for spatial dependence is relatively straightforward in a standard regression model, it is much less so for binary response models (Bell & Irwin, 2002).

Due to the difficulties in solving N-dimensional integrals, new computational approaches have been introduced, such as the EM algorithm, Gibbs sampler, and simulation methods (Fleming, 2004). Carrion-Flores and Irwin (2004) use an iterative method of N integrals to correct for spatial dependence. In the case of the spatial probit model, there are two approaches. The first is introduced by McMillen (1992), where the Expectation Maximization (EM) algorithm is applied to replace the latent unobserved variable with an estimated value. The second is, the Bayesian Gibbs sampler approach, proposed by LeSage (2000). It is an extension of the Gibbs sampling method proposed by Geman and Geman (1984). Autologistic models are conceptually similar to the Bayesian approach to estimating spatial discrete choice models. In contrast, the Spatial Probit EM algorithm is inefficient because it requires the inversion of a $(N \times N)$ matrix, which may be problematic for very large samples.

When spatial regression models are used, the definition of the neighborhood structure is a critical issue. Changes in neighborhood settings result in different estimates. In practice, various neighborhood structures are considered, because there are no firm rules for neighborhood definition in spatial econometrics (Bell & Irwin, 2002). Neighborhood structures are generally defined based on proximity, contiguity, or both. Although the temporal dynamics of land development have been investigated at the disaggregate parcel level (Irwin & Geoghegan, 2001), to the best of our knowledge there are no land-use change models explicitly accounting for spatial and temporal dynamics at this level. Irwin, Bell, and Geoghegan (2003) employ a duration model for land-use conversion, that accounts for both the spatial and temporal aspects of land conversions, using parcel-level data. However this model does not account for temporal lags in land development. Huang, Zhang, and Wu (2009) also investigate spatio-temporal dynamics in land-use conversions, but with no temporal lagged variable. They use a common smoothing technique in time-series data to uncover trends. Wang, Kockelmanb, and Lemp (2012) discuss the absence of space-time data that would allow researchers to develop a dynamic spatial model. Finally, as an alternative spatial discrete model of land-use change, Bhat, Dubey, Alam, and Khushefati (2015) propose a multiple discrete-continuous probit model, with the maximum approximate composite marginal likelihood method.

The availability of spatial panel data provides an opportunity to deal with time-invariant variables in modeling land development. In fixed-effect models, time-invariant variables are generally dropped from the model, since their effects are assumed constant over time. Many spatial features of land can be considered as time-invariant variables, such as location, land size, surface slope, and orientation. However, this modeling approach does not provide any information about the impacts of these time-invariant variables. Chakir and Gallo (2013) introduce a more sophisticated approach to investigate land development dynamics, aggregated at the French department (province) level over 1992–2003, using a spatial Seemingly Unrelated Regression (SUR) model, to control for spatial autocorrelation and unobserved individual heterogeneity. However, their spatial SUR model does not provide an explicit explanation of the impacts of historical conditions.

While standard spatial econometric models cannot deal with spatial and temporal factors when datasets are large, the spatio-temporal binary autologistic model (STARM), introduced by Zhu et al. (2008) and used to investigate Mountain pine beetle outbreaks in Western Canada, is considered feasible even with large datasets. The probability of outbreak is assumed to be conditional on current and past observations. A spatial weight matrix is used to control for spatial dependencies. A Markov chain assumption is made, where the distribution of the endogenous variable is dependent on the most recent time points, because historical tree conditions are considered to have an effect on current three conditions. STARM is the methodological basis of the modeling approach used in this research, as detailed in the next section.

3. Methodology

In the binary spatio-temporal autologistic model (STARM) used to investigate parcel-level dynamics, the following two outcomes are considered: change in land status or continuation of the existing status. These outcomes are captured by a discrete (1, 0) variable. The Bernoulli distribution can be used to represent the process of land development as a trail of decisions resulting in either success (change) or failure (no change). The conditional probability of a response variable y with two potential outcomes (1, 0), given input variables, represents the likelihood p of development of a parcel. In a sequence of independent Bernoulli trials $y_1, \dots, y_n$ with a constant probability of success p, the joint probability of the distribution is:

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} y_i = (0,1) \tag{1}$$

In the binary STARM specification, $Y_{i,t}$ $(=0, 1)$ denotes a status change in parcel i at time t, i is the index of parcels $(1, \dots, N)$, and t is the index of discrete time points $(1, \dots, T)$. The conditional distribution of the response variable $Y_{i,t}$ is assumed to depend on land development decisions in the neighborhood of parcel i in both current and recent times. Because of the neighborhood feedback mechanism, $Y_{i,t}$ is endogenous, as land development in the neighborhood of parcel i has an impact on the development potential of parcel i, and a change in the status of parcel i simultaneously affects the development potential of the parcels in its neighborhood. In contrast, $Y_{i,t-s}$ is exogenous because current land development cannot change past development, while historical conditions have an effect on current land development.

The probability of land development is estimated using logistic regression. The model estimates the probability of change in the state of parcel i at time t as conditional on covariates, spatial lag, temporal lag and spatio-temporal lag variables. The logit function is used to insure that $Y_{i,t} \in [0-1]$. Eq. (2) represents the systematic component of the model. In the case of $\rho_l = 0$ for $l = 1, \dots, L$ and $\gamma_{ls} = 0$ for $ls = (1, \dots,$