Urban Growth Prediction: A Review of Computational Models and Human Perceptions

Dimitrios Triantakonstantis, Giorgos Mountrakis
State University of New York College of Environmental Science and Forestry, Syracuse, USA
Email: trdimitrios@gmail.com, gmountrakis@esf.edu

Received September 28, 2012; revised October 28, 2012; accepted November 29, 2012

ABSTRACT

Human population continues to aggregate in urban centers. This inevitably increases the urban footprint with significant consequences for biodiversity, climate, and environmental resources. Urban growth prediction models have been extensively studied with the overarching goal to assist in sustainable management of urban centers. Despite the extensive body of research, these models are not frequently included in the decision making process. This review aims at bridging this gap by analyzing results from a survey investigating developer and user perceptions from the modeling and planning communities, respectively. An overview of existing models, including advantages and limitations, is also provided. A total of 156 manuscripts is identified. Analysis of aggregated statistics indicates that cellular automata are the prevailing modeling technique, present in the majority of published works. There is also a strong preference for local or regional studies, a choice possibly related to data availability. The survey found a strong recognition of the models’ potential in decision making, but also limited agreement that these models actually reach that potential in practice. Collaboration between planning and modeling communities is deemed essential for transitioning models into practice. Data availability is considered a stronger restraining factor by respondents with limited algorithmic experience, which may indicate that model input data are becoming more specialized, thus significantly limiting wide-spread applicability. This review assesses developer and user perceptions and critically discusses existing urban growth prediction models, acting as a reference for future model development. Specific guidelines are provided to facilitate transition of this relatively mature science into decision making activities.

Keywords: Urban Models; Urbanization Prediction; Survey; GIS Modeling; Urban Planning

1. Introduction

Urbanization has significantly increased over the last two centuries. In year 1800 only 2% of people lived in cities, while in year 1900 this percent increased to 12%. Recent studies indicate that in year 2008 more than 50% of the world population lived in urban areas, with this percentage expected to reach 75% by year 2030 [1]. It is estimated that global urban land use will increase by at least 430,000 Km², about the size of Iraq, by 2030 [2]. Urban land cover occupies only 2% or 3% of the earth surface [3], yet it has been recognized that urban growth is associated with many socioeconomic and environmental problems. For example, impervious surfaces that result from urbanization dramatically increase peak discharges associated with storm and snowmelt events, which in turn makes more likely downstream flooding as storm waters exceed stream channel capacities [4]. The alteration of surface materials also changes the amount of solar radiation reflected or absorbed resulting in micro-climate changes through temperature and humidity alterations. These changes contribute to the urban heat island phenomenon, which affects human health and comfort and increases energy demands for cooling [5]. Furthermore, pollutants that are concentrated on urban surfaces degrade the biological, chemical, and physical characteristics of lakes, streams, and estuaries receiving urban runoff leading to aquatic and terrestrial habitat modifications. It is well-documented that indicators related to the biological integrity of streams and riparian habitat are inversely related to the amount of impervious surfaces adjacent to them [6].

Urban modeling studies are currently considered an essential component for numerous complex environmental approaches. For example, urban growth modeling can assist in adaptation and mitigation scenarios with respect to climate change because of the large amounts of air, soil and waste emissions that occur in large cities [7-11]. Furthermore, due to the increasing trend of urbanization along with potential environmental consequences, urban growth modeling appears to have a protagonist role in urban planning to assist in decisions related to sustainable urban development [12-17].
As a response, the scientific community has developed numerous urban growth prediction models (UGPMs) over the past decades in order to study urban land use dynamics and simulate urban growth. These models, even though they have a common goal, vary widely in underlying methodologies and theoretical assumptions, and spatial/temporal resolutions and extents. Several reviews are available on the subject [18-22]. The motivation behind our work to shine light in a well-known limitation. Currently, a significant gap exists between modeling efforts and their implementations in decision making as urban planners and decision makers have only partially incorporated these research products. To investigate further this issue an online survey was conducted acting as a reference for future UGPM applicability. Therefore the overarching goal of this paper is not only to provide the necessary framework for future development of accurate UGPMs with the modeling community but also UGPMs that are applicable to urban planning tasks.

In the next section a retrospective summary of existing works is provided acting as a reference for future UGPM development. Additional text in the appendix discusses different data sources for these models (Text S1). The survey is introduced with associated findings followed by an in depth discussion on the current state-of-the-art and potential areas of improvement in all UGPM stages, from data sources, to mathematical modeling choices to modeling characteristics facilitating effortless incorporation to decision making.

2. Urban Growth Prediction Models

Urban Growth Prediction Models (UGPMs) are tasked to capture intrinsic and complex relationships in space and time. The spatial complexity reflects the impact of numerous biophysical and socioeconomic factors and as a result heterogeneous patterns appear across location and scale thus making urban development a dynamic and non-linear process [23]. Temporal complexity presents itself through the prediction difficulty for extended temporal intervals. The urban evolution often implies irreversibility [24,25] therefore, in a changing urban environment, only short term predictions can be securely applied [23].

Furthermore, the dynamic process of urban growth is associated with decision making complexity [12,26,27]. Decisions of urban planners and policy makers are difficult to predict, especially over an extended period of time as they depend on stakeholder needs, economic pressure and relevant legislation.

A plethora of models has been applied to examine urban growth, approaching the problem from diverse views. A wide range of electronic sources was accessed leading to the eventual selection of 156 UGPM manuscripts. For manuscript selection we followed the Quality of Reporting of Systematic reviews and meta-analyses (PRISMA) guidelines [28]. Our analysis collected manuscripts until August 2012. Figure 1 describes the selection process and a detailed PRISMA statement is presented in the Appendix (Table S3). Initially records were identified through electronic searches in relevant databases (e.g. ScienceDirect) and search engines (e.g. Google Scholar). After removing duplicates unrelated records were removed from the list (e.g. records returned from real estate or urban planning manuscripts). At the next screening level manuscripts were excluded falling in three general categories: 1) did not provide a spatially explicit model output (e.g. demographic, population density, econometric modeling manuscripts); 2) did not incorporate an explicit spatially explicit prediction mechanism (e.g. mostly manuscripts detecting urban change using remotely sensed methods); and 3) did not simulate urban change but other land use types (general land use change models without urban change specialization). Manuscripts in the latter case were reviewed in [29]. At the last stage we excluded manuscripts that were deemed relevant but included either only a theoretical component or were a simple application of previously published work.

A summary of the reviewed manuscripts is presented in Figure 2. Several common characteristics are examined. Firstly, in terms of input types there is a prevalence of biophysical or biophysical/socioeconomic inputs (107 manuscripts) followed by land-use inputs. There is also a strong preference to local (64) and regional (67) studies, possibly due to data availability, development and validation costs and funding directions. The spatial resolution, defined as the cell size of model output (not model inputs), showed a preference for moderate values (<100 m). The temporal resolution, defined as the temporal length of model reference data (not to be confused with prediction temporal extent) showed a tendency for relatively short time intervals (85 manuscripts with <20 years), a constraint possibly imposed by data sources. A summary table for each of the 156 reviewed papers is provided in the Appendix (Tables S1 and S2).

From the modeling perspective two particular decisions significantly affect model design and performance. The first one is conceptual and relates to expected spatial behavior and relationships. The second decision is the underlying algorithmic type for the model. These decisions are discussed in the next two sections.

2.1. Spatial Autocorrelation and Heterogeneity

To address some of the underlying complexities, UGPMs have incorporated two major analytical characteristics of spatial analysis: spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation refers to the systematic
Figure 1. PRISMA 2009 flow diagram regarding the article selection.

Figure 2. Characteristics of reviewed UGPM manuscripts.
variation of a variable, obeying in the first law of Geography [30], in which near things are more related than distant things. According to [31] a spatial or temporal heterogeneous system is characterized by different values in specific locations or time intervals. In an urban environment spatial heterogeneity refers to the different spatial distribution of urbanization along with the underlying driving factors.

Spatial autocorrelation can be described using global and local spatial statistics. In some general spatial statistical studies, global and local spatial statistics have been used, such as: Moran’s I [32-34], Geary C [35-37], G statistic [38,39] and Local Indicators of Spatial Association (LISA) [40,41]. Spatial statistics, such as landscape metrics and texture parameters (e.g. entropy, variance, homogeneity) have been also widely used in urban growth prediction models [42-57].

The estimation of a dependent variable as a function of a matrix of independent variables can be carried out using a) a simple ordinary least-squares (OLS) regression and b) a global spatial regression [58]. The former obeys the Equation (1):

\[ y_i = b x_i + e_i \]  

where \( y_i \) is the dependent variable, \( x_i \) is the matrix of independent variables, \( b \) is a vector of coefficient and \( e_i \) is a vector of random errors. GEOMOD is an example of a land use model which uses multiple regression for determining the weight of each variable in order to specify the location of each changed cell [59]. The global spatial regression is used when there is spatial autocorrelation in the dependent variable and therefore, violation of the OLS regression assumptions is present. Thereupon, a supplementary explanatory variable is added in order to represent the spatial dependency of the dependent variable, as the following Formula (2) illustrates:

\[ y_i = b x_j + e_j + \delta \sum_{j \neq i} w_{ij} y_j \]  

where \( \delta \) is the spatial autoregressive coefficient and \( w_{ij} \) is the spatial weight of the neighbors \( i \) and \( j \) [60-63]. Spatial autocorrelation depends on spatial scale [64] and in some cases it is avoided by sampling points at distances greater than the distance where spatial autocorrelation occurs [65]. An alternative solution is autologistic regression that accommodates the autocorrelation effects by using an autocovariate term [66-68]. This additional independent variable captures the spatial variability of the response variable.

Another important characteristic of urban growth is spatial heterogeneity [69]. Different patterns of urban growth may be treated separately using local models instead of a global model into the entire study area [70]. Three modeling techniques may be applied in order to handle spatial heterogeneity: switching regressions, multilevel models and geographically weighted regression [71]. Switching regression model classifies a dataset into a number of mutually exclusive homogenous areas, where a linear regression model is applied in each of them [72,73]. The switching regression model bridges the gap between a local and a global approach in spatial modeling. Multilevel models, also known as hierarchical models, group units of interest (e.g. urban structures) into higher level clusters (e.g. neighborhoods). The motivation of using multilevel models is that they can differentiate heterogeneity between clusters and units nested within clusters [74,75]. Finally, geographically weighted regression is based on assigning weights to all points of dataset according to their distance from a focal point of interest [76].

Despite the fact that these are known issues in UGPMs, from the reviewed manuscripts only six concurrently supported spatial autocorrelation and heterogeneity, and twelve supported either heterogeneity or autocorrelation (Figure 1). The lack of incorporation of these concepts may be attributed to increased mathematical complexity associated with them rather than awareness of their contributions.

2.2. Underlying Modeling Algorithms

As Figure 3 indicates, a wide variety of algorithms has been incorporated in UGPMs with cellular automata tested in the majority of the reviewed manuscripts. In this section we discuss applications, advantages and limitations of currently prevailing and promising methods.

2.2.1 Cellular Automata Modeling

Cellular automata (CA) were introduced by Ulam and Neumann in 1940 and since 1980 numerous models have been developed for simulating urban growth [77]. CA are defined as discrete dynamics systems, represented by a grid of cells, in which local interconnected relationships exhibit global changes [34,78]. Generally, the state of each cell depends on the value of the cell on its previous state as well as the values of its neighbors according to some transition rules. These rules affect the urban growth, indicating environmental and socioeconomic support or limitations. Therefore, the bottom up approach implemented in CA relies on the simulation of local actions that progressively create the global emergent structure [79,80].

CA deals with non-linearity of urban structures and the iterative process leads to produce fractal patterns, which are common characteristics in an urban environment [81].

The applications of CA in urban growth can be classified into: 1) theoretical model developments and 2) applied UGPMs in real data. The first category, which developed in early years of CA, includes theoretical developments
D. TRIANTAKONSTANTIS, G. MOUNTRAKIS

Figure 3. Underlying UGPM algorithms sorted by popularity (percentage of 156 manuscripts, a manuscript may contain multiple algorithms).

of CA models in urban simulation [82-89]. In these studies artificial case exemplars were used to develop theoretical models. The authors also note that urban growth is neither the pure application of The Game of Life nor the pure classical global urban models such as Lowry land use model [34]. Each study area must be examined separately, taking into account the particular conditions which influence urban change. Therefore, a combination of global and local factors must be considered in UGPMs through the appropriate parameterization [84,90,91].

Subsequently, these theoretical approaches found real world implementations. A large number of applications have incorporated CA for UGPM development using real data [91-125]. A combination of CA with Markov models has also appeared in multiple studies [126-130]. A Markov model can not only explain the conversion among land uses, but also calculate the transfer rates among different types. Multi-criteria evaluation techniques [130,131] and weight of evidence [132,133] have been used for estimating the importance of qualitative and quantitative drivers within the CA modelling framework.

An urban growth model, which is widely used by many applications, is the SLEUTH model (slope, land use, exclusion, urban extent, transportation and hillshade). SLEUTH, introduced in [134], is a CA-based UGPM which uses historical data for calibration of variables, achieving successful implementation in regional scale modeling and ability to deal with protection areas. SLEUTH calibration is a computationally intensive process and therefore requires sufficient computing resources [135]. The SLEUTH model is broadly applied with many study areas found around the globe. Over 100 applications in USA and worldwide were accumulated within ten years [136]. Application examples are available at [137-152]. A new version of the SLEUTH model (SLEUTH-3r) was proposed in [153], increasing the performance and the applicability by introducing new fit statistics, which enhance the calibration process. Metronamica, another CA-based model, was used with SLEUTH in [154]. Metronamica is defined by three components: distance decay functions, integration with GIS and constrained cell transitions by calculating a ranked score for each cell.

Other CA-based urban growth models are iCity and SimLand. iCity (Irregular City), which was developed by [155], is an extension of the traditional form of CA, that includes an irregular lattice [82]. SimLand is a simulation CA model based on multicriteria evaluation methods such as analytical hierarchy process. It was developed by [96] in order to facilitate easier retrieval of spatial data, and to integrate multicriteria evaluation methods and CA with GIS (spatial decision support system), applying a more realistic way in defining transition rules. In addition to the above models, a fuzzy inference guided cellular automata approach has been proposed by [156], where fuzzy theory was applied to provide common semantic and linguistic knowledge to urban growth and simplify transition rules. An optimal combination of transition rules has been investigated using Genetic Algorithms within the calibration process [157,158]. Non-linear transitions rules were also examined using Support Vector Machines [159]. Moreover, Artificial Neural Networks were used to generate conversion probabilities from the initial cell to target land use type. [160] applied a Radial Basis Function Neural Network Model (RBFNN) for this
objective. Because Neural Networks cannot explicitly identify the contribution of each variable, less important variables may be included into the model. Due to this limitation, Bayesian Networks were also implemented, where land use drivers have a clear interpretation and the probabilities are easier to understand compared to weights within Neural Networks [161]. The importance of each land use driver (weights) can be also determined using Monte Carlo repetitions [162].

[163] used an Agent Based Model within CA to produce an Entity Based model, in which each household member is considered as a separate entity (agent). The neighborhood infrastructure as well as other neighborhood entities contribute to each household behavior. Vector based CA have been also used by [164] to overcome the difficulties with sensitivity to cell size and neighborhood configuration. They allow the presentation of space by applying vector shapes (polygons), while the neighborhood is semantically described. More specifically, the neighborhood changes through time without having a fixed distance delineating it. Finally, in a logistic CA model, proposed by [165,166], the urban growth can be described by a continuous spatial diffusion process where the dependent variable is continuous ranging from 0 to 1.

CA methods also face challenges in urban simulations. Due to spatial heterogeneity, different parts of cities should be addressed by different transition rules. Therefore, global transition rules applied by CA may be inappropriate for modeling cellular space. Furthermore, spatial heterogeneity dictates that neighborhoods should be described by different shapes and sizes in order to capture better spatial interactions of urban structures. CA methods typically implement regularity in neighborhoods, limiting modeling capabilities. Finally, disadvantages of CA include the assumption of spatial and temporal invariance for transition rules and the inability of CA to deal with stochastic behavior [82]. CA examines the synchronous dynamics of urban environment, in essence all cells update simultaneously at each iterative step. Real cities violate this condition because of their chaotic behavior and therefore, further research in stochastic CA is still needed.

2.2.2 Artificial Neural Networks Modeling
Artificial Neural Networks (ANNs) are incorporated in UGPMs due to their increased modeling capabilities. Unlike most multi-variate modeling techniques, ANNs are not significantly affected by input data relationships, therefore no assumptions about spatial autocorrelation and multi-collinearity must be made. The multi-layer perceptron (MLP) neural network, produced by [167] has gained large applicability. MLP is a system composed by a number of single processing elements, called neurons. The network output is computed using an internal transfer function depends on the input neurons which are connected together with weighted relationships. The ANN learns by the existed input and output data through an iterative way of learning (e.g. back-propagation algorithm). ANNs popularity has substantially increased in recent years due to improved computing power capabilities with applications in many scientific fields [168-170]. Because cities grow in a comprehensive way, the ANNs learning process can produce tools capable to model urban structure complexity [171].

UGPMs often incorporate environmental and socio-economic variables to simulate the change that has occurred. [172,173] produced the Land Transformation Model (LTM), where GIS and ANNs are combined in order to forecast land use changes, taking into account a variety of social, political and environmental factors. Another approach in urban growth prediction is the ART-MMAP [174], which produces a prediction map under different scenarios, using past information of land use driving forces and socioeconomic data. [175] produced an ANN-based urban growth model in order to estimate future urban growth boundaries and complex geometry of cities, based on factors of urban sprawl such as distances from roads, green spaces, service stations and built areas, elevation, slope and aspect. Another ANN-based urban growth model is proposed by [176] in order to reveal how future urban shape or growth patterns relate to site attributes and reduce the subjectivity in urban growth modeling.

ANNs as a non-parametric technique can successfully capture the spatial heterogeneity [177]. [178] designed a multiple neural network, which allows input data to be automatically reallocated into appropriate neural networks, in order to handle spatial heterogeneity. Furthermore, several neural network algorithms have used fuzzy logic [179,180], multivariate analysis [181] and self-organizing maps [182].

ANNs have been also used for calibration and simulation of CA models in urban studies [183,184]. [183] developed an integration of CA and ANNs in order to simulate land use dynamics, using multiple output neurons. Moreover, [98] introduced a land use simulation model which uses a supervised back-propagation ANN and the generated probabilities were input to a CA model. ANN-based cellular automata models were also applied for urban/non urban cell transitions [184,185].

ANNs have the tendency to overfit data; therefore, the training dataset size should be selected carefully with respect to the number of hidden neurons [186]. A usage of at least 5 to 10 times the training size as are the existed weights is generally accepted [187,188]. The demand of a large training size needed to take advantage of ANNs modeling capabilities often limits UGPM incorporation as training data may not be as widely available. Other
different fractal dimensions may be assigned into an urban context, even though different fractal dimensions have been observed for the same object [200]. Therefore, different objects from the same fractal class may have significantly different fractal dimensions [200].纹理的不同可能有不同的分形维数。因此，来自相同分形类别的不同物体可能具有显著不同的分形维度。

2.2.3. Fractal Modeling

Fractal geometry has also been used in urban growth simulation. Classical Euclidean geometry is recognized as inadequate to describe the spatiotemporal patterns in nature. [189] introduced fractal geometry and since then a rapid expansion of fractals in many scientific fields has been observed [190]. Cities can be considered as fractal objects, where the interaction of different spatial components can be described by non-linear relationships [81]. Fractal theory deals with the non-linearity of spatial structural complexity, indicating that urban growth conforms to a multiscale spatial self-organization [191-196]. Self-organization is an important process in environmental phenomena. It is based on the ability of the system to organize its components with an internal power support. In self-organized systems, the organization spontaneously increases without being controlled by external force, but it is triggered by internal variation process, which could be fluctuations or noise. The urban environment has an unexpected behavior, when some regions are isolated and examined separately. Because any region is a component of the urban self-organized system, it must be treated globally rather than as independent parts [3,23,171].

A diffusion-limited aggregation (DLA) method was proposed by [197], where the urban structures were generated in a tree-like form, producing spatial self-similarity in different scales. [81] employed fractal methods to measure the irregularity of urban land parcels, examining the similarity of fractal dimension. [198] applied self-affine fractal structure in order to explain the complex pattern of urban spatiotemporal evolution, followed by an optimization of urban form using fractal dimension. [195] used a Minkowski dilation in order to detect spatial discontinuity or urban agglomeration and applied a threshold which defines the limit of self similarity of the urban system and stops the dilation process.

Although there has been a tendency to see “fractals everywhere”, many objects cannot be considered true fractals. Natural objects and phenomena are not necessarily described by self-similarity [199]. Numerous algorithms have been programmed in order to calculate the fractal dimension. Fractal dimension measurements have limitations such as: a) different techniques of fractal dimension measurement of the same object may yield to different results, b) objects with different morphological characteristics may share the same fractal dimension and c) objects from the same fractal class may have significantly different fractal dimensions [200]. Therefore, different fractal dimensions may be assigned into an urban structure using various software packages. Moreover, urban structures with different texture may produce similar fractal dimension or urban structures with similar texture may have significantly different fractal dimensions. Finally, accurate fractal modeling is highly dependent on the satisfactory assessment of urban complexity.

2.2.4. Linear/Logistic Regression

Linear regression analysis examines the relationships between urban land uses and independent variables. When the dependent variable is dichotomous, logistic regression can be applied to predict the presence or absence of a characteristic based on a matrix of independent variables. For example, a dichotomous dependent variable can be urban change, where a value of 1 indicates change from non urban to urban and 0 indicates no change. The independent variables can be continuous, categorical, or both. Linear and logistic regression models have been widely used in urban growth modeling, accommodating socio-economic and environmental independent variables [3,44,59,66,72,73,76,177,178,181,201-217]. CLUE (Conversion of Land Use and its Effects) is a model, which simulates land use changes through a non-spatial demand module and a spatially explicit allocation module [206,218]. For the land use demand model, different modeling approaches can be implemented, ranging from trend extrapolations to more complex economic modeling techniques. In the spatial allocation model, the relationships between land use and independent variables are evaluated using logistic regression. A spatially explicit procedure is used to create the Land Use Scanner model, in which the residential land use demand on spatial units is allocated [219,220]. In the Land Use Scanner model, a logistic regression is applied to empirically specify weights for the preparation of suitability maps.

Land Suitability Index maps were created by [221] using frequency ratio, analytical hierarchy process, logistic regression and ANNs in order to evaluate the performance of each method. The accuracy results did not reveal any important differences among these methods. Regression analysis combined with Markov chains appeared in [222] to study how urban growth relates to landscape change as well as to population growth. Moreover, an enhanced approach in spatial modelling, Geographically Weighted Regression, tackles spatial non-stationarity in regression analysis and the regression coefficients are calculated by spatially dependent weights within a neighbourhood [223-225].

Unfortunately, linear and logistic regressions do not offer high modeling capabilities and they fail to capture non-linearity in the relationships between the dependent and independent variables or to address correlations between independent variables.
2.2.5. Agent-Based Modeling

Agent-based models apply a bottom-up approach to yield better understanding of urban systems by allowing the simulation of individual actions of agents and measurement of the resulting system behaviour [226-229]. Agents are autonomous units, which exchange information with other agents under an interactive communication. The individual behavior of the agents allows the influence of human decision making to be incorporated into the model.

A framework which allows describing urban dynamics as a function of interaction between mobile “agents” and static CA is very important in simulating urban sprawl [230]. [231] developed a national scale simulation agent-based model that was based on applying the concept of the “agent” as the decision maker, taking into account the available biophysical and human factors. A multi-agent model for the study of urbanism is presented by [232], in which different rules and parameters can change the spatial structure of the urban system. Moreover, another multi-agent model is applied by [233] where the interactions of different agents, such as residents, peasants and governments are simulated. A statistical approach in validating spatial patterns in agent-based models is presented by [234]. [235] examined the spread of urban development, evaluating the effectiveness of greenbelts located beside a developed area. [236] developed a model which simulates the polycentric development of urban systems using household agents who choose the location of their houses according to several properties such as land prices, traffic problem, and landscape attractiveness. Despite the satisfactory applicability of agent-based models in urban growth simulation, there are limitations mostly resulting from the arbitrary definitions of initial conditions and interaction rules of agents, which could lead to highly variable results [237].

2.2.6. Decision Trees Modeling

Decision trees is a top-down classification algorithm, which has been used in land use change modeling [44,238,239] and land use classification from remotely sensed imagery [240,241]. Despite their limited use in urban growth modeling, decision trees are of particular interest due to their ability to generate rules and the easiness to understand the model structure. Decision trees automatically derive a hierarchy of partition rules that are used to split data into sequential segments. The construction of a decision tree involves three basic steps. The first step is related with tree building using recursive splitting of nodes. In the second step a pruning process is applied, where smaller trees are produced with lower complexity [242-244]. The reduction of overfitting by removal the noisy and erroneous data is also achieved through a pruning process [245]. Finally, the optimal tree which yields the lower testing error is selected. There are two different types of decision trees according to the learning algorithm: a) classification trees and b) regression trees. In the former, the results of the predicted variable take only two values, usually 0 and 1, while in the latter the predicted output varies between the values of the dependent variable [243]. Decision trees have been widely used in urban land use image-based classification to map urban structures and urban vegetation cover, which are important components in urban modeling and planning [246-249].

Spatial heterogeneity is an important attribute of urban development [69]. An important limitation of decision trees is the simple-algorithm structure, where the entire area is indiscriminately targeted into a global rule. As a result, a low degree of spatial heterogeneity can be incorporated into the model [250]. [70] investigated an expert-based selection of models applied in different regions and the results showed that this approach performs better than using a global model.

Many environmental variables exhibit spatial autocorrelation causing spatial clustering. The spatially dependent data produce less information, making the degrees of freedom of the sample exaggerated [251]. Therefore, spatial autocorrelation causes underperformance of the decision tree modeling [252-254]. However, this limitation could be overcome using a proper sampling design, where the distance between sampling points is greater than the distance at which spatial autocorrelation occurs [255-257]. [258] introduced a novel method in order to handle spatial autocorrelation. According to this method, the conventional entropy of the decision tree was replaced by spatial entropy, which takes into account the spatial autocorrelation.

Unfortunately, decision trees can create over-complex structures that restrict generalization abilities. This issue, known as overfitting, can be rectified to a certain extent using a pruning process. Moreover, decision trees are unstable algorithms as they can produce dramatically different classifiers by using only slightly different training samples [259]. This instability could be reduced by applying a number of decision trees into the training sample each time the training sample changes. Finally, when data includes categorical variables, information achieved from decision trees is biased in favor of the variables with more categories. [260] presents a bias correction to reduce the difference between numerical and categorical variables using a univariate split method based on several aspects of the data such as sample size, number of variables, and missing values.

3. A Survey on Model Integration to Decision Making

The application of urban planning to address issues related
to sustainable development represents a balancing act between environmental resources and economic demands. The World Planners Congress in Vancouver in 2006 suggested that urban planners should address urban sustainability in developing and poor countries, putting human livelihoods in the core of urban planning [261]. Because of the increasing trend of urbanization along with potential environmental consequences, urban growth modeling should have a protagonistic role in urban planning to provide appropriate decisions for sustainable urban development [12,13,15,16,262]. However, despite significant efforts in developing UGPMs their usage is limited in the planning community. To investigate further potential reasons that restrict this transition from development to practice a survey was conducted. We are further motivated to use the survey results as a guide for future improvements.

3.1. Questionnaire Content

An online questionnaire was constructed and requested information on the respondent’s background and UGPM development and practise. Table 1 summarizes the questions related to the respondent’s background, such as education (highest degree), employment type and GIS/professional/policy experience.

<p>| Table 1. Professional background of respondent and corresponding statistical groupings. |
|---------------------------------|---------------------------------|---------------------------------|</p>
<table>
<thead>
<tr>
<th>Question</th>
<th>Short Name</th>
<th>Response Options (Grouping)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest Degree</td>
<td>Highest Degree</td>
<td>High School (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Associate (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B.Sc./B.A. (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.Sc./M.A. (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ph.D. (B)</td>
</tr>
<tr>
<td>Employment Type</td>
<td>Employment</td>
<td>Academic Employee (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private Employee (B)</td>
</tr>
<tr>
<td>I am more familiar with</td>
<td>Professional Exp.</td>
<td>Independent Contractor (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Government Employee (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other/Unemployed (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urban Planning (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Urban Modeling (B)</td>
</tr>
<tr>
<td>I am a policy maker with influence on urban planning decisions</td>
<td>Policy Exp.</td>
<td>None of the two (*)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both (C)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No (A)</td>
</tr>
<tr>
<td>The spatial extent of my typical urban planning tasks is</td>
<td>Spatial Extent</td>
<td>City/Municipality/County (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multi-County/ Single State (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multi-State (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>National (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Familiarity (A)</td>
</tr>
<tr>
<td>Familiarity with Geographic Information Systems (GIS)</td>
<td>GIS Exp.</td>
<td>Low (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No Familiarity (A)</td>
</tr>
<tr>
<td>Familiarity with algorithms used in UGPM development</td>
<td>Algorithmic Exp.</td>
<td>Low (A)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium (B)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High (B)</td>
</tr>
</tbody>
</table>

*An answer of “None of the two” to this question would result in the respondent being excluded from the analysis. Letters in parenthesis indicate grouping for statistical testing.*
Answers to these questions were used initially to screen participants (e.g. those with no relevant experience) and later to identify patterns in subgroup responses (e.g. whether data limitations were more pronounced depending on respondent’s GIS experience).

The second part of the survey requested the respondent’s opinion on UGPM related questions. The question text and answer type along with their short name used in follow up analysis are listed in Table 2.

### 3.2. Questionnaire Dissemination and Response Rates

The target audience included respondents from both the modeling and planning communities. Modellers were identified through relevant literature manuscripts. Approximately 1000 email addresses were collected. An email was sent out explaining the survey along with a direct link for participation. Responses to the direct link were not tracked as our IRB protocol required complete anonymity but from analyzing the participants’ background the estimated response rate was 14% (app. 140 responses). The access the planning community we solicited help from the American Planning Association. After satisfying their internal review process a direct link of the survey along with a short explanation was included in APA’s electronic newsletter titled APA Interact. The approximate dissemination base was 10,000 members and we estimate 100 responses were obtained leading to a 1% response rate. Considering practical limitations, especially related to privacy, the respondents were deemed a representative sample without a known bias.

In total 242 questionnaires were submitted. After filtering for non-relevant background and those answers belonging to the “Neutral” category the range of usable responses ranged from 84 to 166. Table 3 provides a detailed participation count for each question, also taking account groups outlined in Table 1 as A, B or C.

### 3.3. Questionnaire Findings

Results are aggregated in Figure 4. There is an overwhelming response in the potential of UGPMs (98% positive), however respondents are split on whether UGPMs currently reach that potential in practice (43% positive). The lack of current widespread implementation does not seem explicitly related to UGPMs prediction quality (67% positive). Rather it is constrained by lack of awareness outside the modeling community (92% negative), which is further supported by lack of communication between modeling and planning communities (94% negative). Modellers are found to create models that are not easy to understand (72% negative, with almost identical responses from the planners (74%) and surprisingly the modelers community (71%)), while planners do not identify clear expectations (81% negative with the modelers being slightly more negative (87%) than the planners (79%)).

Further analysis was undertaken to reveal patterns associated with the respondent’s background. Table 4 lists the Pearson chi-square tests that were conducted to assess if an association existed between the respondents’

<table>
<thead>
<tr>
<th>Question</th>
<th>Short Name</th>
<th>Answer Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you agree that Urban Growth Prediction Models (UGPMs) have the potential to provide effective assistance to urban planners and decision makers?</td>
<td>Potential Assistance</td>
<td>Agree/Disagree</td>
</tr>
<tr>
<td>Do you agree that UGPMs are currently reaching that potential in practice?</td>
<td>Assistance in Practice</td>
<td>Agree/Disagree</td>
</tr>
<tr>
<td>How would you rate data availability for UGPM development purposes?</td>
<td>Data Availability</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>How would you rate the quality of UGPM prediction results?</td>
<td>Prediction Quality</td>
<td>InAccurate/Accurate</td>
</tr>
<tr>
<td>How would you rate the communication between the modeling and planning communities?</td>
<td>Communication</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>How would you rate the modellers’ attention to practical considerations such as software compatibility and software user friendliness?</td>
<td>Practical Considerations</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>How would you rate the planners’ clarity on UGPM outcome expectations?</td>
<td>Clear Expectations</td>
<td>Unclear/Clear</td>
</tr>
<tr>
<td>How would you rate the awareness on Urban Growth Prediction Models (UGPMs)</td>
<td>Awareness</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>How would you rate the financial support for UGPM development?</td>
<td>Development Support</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>How would you rate the financial support for UGPM usage/implementation?</td>
<td>Usage Support</td>
<td>Lacking/Adequate</td>
</tr>
<tr>
<td>The internal workings of UGPMs are easy to understand.</td>
<td>Easy to Understand</td>
<td>Agree/Disagree</td>
</tr>
<tr>
<td>Is the collaboration between urban modellers and planners necessary for successful development of UGPMs?</td>
<td>Development Collaboration</td>
<td>Agree/Disagree</td>
</tr>
<tr>
<td>Is the collaboration between urban modellers and planners necessary for successful usage of UGPMs?</td>
<td>Usage Collaboration</td>
<td>Agree/Disagree</td>
</tr>
</tbody>
</table>
Table 3. Number of respondents per group.

<table>
<thead>
<tr>
<th>Question, Respondent Grouping®</th>
<th>All</th>
<th>Policy Exp. (No/SomeYes)</th>
<th>Spatial Extent (≤County/ Multi-County+)</th>
<th>Professional Exp. (Plan/Model/Both)</th>
<th>Algorithmic Exp. (No/Low/Med/High)</th>
<th>Employment (Acad/Other)</th>
<th>Highest Degree (PhD/Other)</th>
<th>GIS Exp. (High/Other)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Assistance (% Agree)</td>
<td>166</td>
<td>109/56</td>
<td>91/72</td>
<td>43/77/46</td>
<td>48/118</td>
<td>98/67</td>
<td>101/65</td>
<td>105/61</td>
</tr>
<tr>
<td>Assistance in Practice (% Agree)</td>
<td>117</td>
<td>73/43</td>
<td>67/49</td>
<td>32/52/33</td>
<td>32/85</td>
<td>65/51</td>
<td>68/49</td>
<td>83/34</td>
</tr>
<tr>
<td>Data Availability (% Lacking)</td>
<td>138</td>
<td>92/46</td>
<td>72/63</td>
<td>32/64/42</td>
<td>35/103</td>
<td>87/50</td>
<td>87/51</td>
<td>92/46</td>
</tr>
<tr>
<td>Prediction Quality (% InAccurate)</td>
<td>84</td>
<td>54/30</td>
<td>43/39</td>
<td>20/41/23</td>
<td>20/64</td>
<td>51/33</td>
<td>54/30</td>
<td>54/30</td>
</tr>
<tr>
<td>Communication (% Lacking)</td>
<td>144</td>
<td>96/47</td>
<td>78/63</td>
<td>37/65/42</td>
<td>45/99</td>
<td>86/58</td>
<td>89/55</td>
<td>94/50</td>
</tr>
<tr>
<td>Practical Considerations (% Lacking)</td>
<td>112</td>
<td>74/38</td>
<td>59/51</td>
<td>24/51/37</td>
<td>30/82</td>
<td>72/39</td>
<td>73/39</td>
<td>78/34</td>
</tr>
<tr>
<td>Clear Expectations (% Unclear)</td>
<td>111</td>
<td>74/37</td>
<td>63/46</td>
<td>29/47/35</td>
<td>35/76</td>
<td>69/42</td>
<td>66/45</td>
<td>74/37</td>
</tr>
<tr>
<td>Awareness (% Lacking)</td>
<td>145</td>
<td>94/50</td>
<td>79/63</td>
<td>39/63/43</td>
<td>46/99</td>
<td>87/57</td>
<td>88/57</td>
<td>89/56</td>
</tr>
<tr>
<td>Development Support (% Lacking)</td>
<td>111</td>
<td>72/38</td>
<td>54/54</td>
<td>24/47/40</td>
<td>28/83</td>
<td>68/43</td>
<td>71/40</td>
<td>76/35</td>
</tr>
<tr>
<td>Usage Support (% Lacking)</td>
<td>124</td>
<td>80/43</td>
<td>67/55</td>
<td>27/56/41</td>
<td>30/94</td>
<td>74/50</td>
<td>78/46</td>
<td>84/40</td>
</tr>
<tr>
<td>Easy to Understand (% Agree)</td>
<td>133</td>
<td>93/40</td>
<td>75/55</td>
<td>31/66/36</td>
<td>37/96</td>
<td>90/42</td>
<td>87/46</td>
<td>88/45</td>
</tr>
<tr>
<td>Development Collaboration (% Agree)</td>
<td>161</td>
<td>106/54</td>
<td>89/68</td>
<td>41/72/48</td>
<td>45/116</td>
<td>97/63</td>
<td>100/61</td>
<td>105/56</td>
</tr>
<tr>
<td>Usage Collaboration (% Agree)</td>
<td>165</td>
<td>111/53</td>
<td>92/70</td>
<td>42/74/49</td>
<td>50/115</td>
<td>102/62</td>
<td>104/61</td>
<td>104/61</td>
</tr>
</tbody>
</table>

Note: Respondent number varies because neutral/no opinion responses were excluded from analysis. Also respondent participation per question varied.

Figure 4. Summary results for survey.
Table 4. Individual question responses for different respondent background groups.

<table>
<thead>
<tr>
<th>Question, Respondent Grouping</th>
<th>All Policy Exp. (No/Some/Yes)</th>
<th>Spatial Extent (≤County/Multi-County+)</th>
<th>Professional Exp. (Plan/Model/Both)</th>
<th>Algorithmic Exp. (No/Low/Med/High)</th>
<th>Employment (Acad/Other)</th>
<th>Highest Degree (PhD/Other)</th>
<th>GIS Exp. (High/Other)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential Assistance (% Agree)</td>
<td>98% 0.492 0.072 (96/100) 0.356 0.197 0.699 0.653 622</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistance in Practice (% Agree)</td>
<td>43% 0.083 (37/53) 0.963 0.404 0.262 0.352 (38/47) 0.462 0.298</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Availability (% Lacking)</td>
<td>67% 1.000 (67/67) 0.067 (74/59) 0.556 0.065 (80/63) 0.196 (63/74) 0.172 (63/75) 0.001 (58/87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prediction Quality (% InAccurate)</td>
<td>33% 1.000 0.941 0.656 0.856 0.636 0.334 (37/27) 0.629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication (% Lacking)</td>
<td>94% 0.026 (97/87) 0.170 0.1 (86/97/95) 0.378 0.334 0.069 (97/89) 0.175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical Considerations (% Lacking)</td>
<td>63% 0.918 (62/63) 0.431 0.998 (63/63/62) 0.912 0.139 (68/54) 0.573 0.044 (56/76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear Expectations (% Unclear)</td>
<td>81% 0.607 0.672 0.321 (79/87/74) 0.078 (71/86) 0.305 0.220 0.607</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness (% Lacking)</td>
<td>92% 0.589 0.235 0.706 0.741 0.978 0.664 0.628</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development Support (% Lacking)</td>
<td>81% 0.704 0.620 0.906 0.695 0.572 0.195 (77/88) 0.398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage Support (% Lacking)</td>
<td>86% 0.563 0.483 0.905 0.497 0.542 0.480 0.165 (83/93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy to Understand (% Agree)</td>
<td>28% 0.430 0.188 (24/35) 0.954 (26/29/28) 0.899 0.137 (23/36) 0.934 0.832</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development Collaboration (% Agree)</td>
<td>94% 0.139 (92/98) 0.283 0.021 (100/89/98) 0.247 0.013 (91/100) 0.088 (92/98) 0.531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage Collaboration (% Agree)</td>
<td>98% 0.162 (96/100) 0.077 (96/100) 0.080 (100/95/100) 0.815 0.114 (96/100) 0.121 0.616</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: P-Value from the Pearson chi-square tests is reported for comparison between groups. A P-value smaller than 0.10 would indicate a statistically significant association between the grouping variable and the response to the survey question. Numbers in parenthesis indicate % of respondents in each group that answered with the response parentheses shown for each question. The “All” column represents all respondents without any grouping.

Analysis of the results identified the following interesting findings:

- A surprising result was that respondents with higher involvement in urban growth model development are more satisfied with data availability than others. Respondents with high GIS experience (58%), PhD degrees (63%), academic employment (63%) and large scale studies (59%) identified data availability as lacking or significantly lacking. That percentage was higher for medium or lower GIS experience (87%), non-PhDs (75%), non-academics (74%) and respondents with no or low algorithmic experience (80%). Policy experience did not seem to matter (67% for both respondent groups, namely with or without relevant experience).
- On the question regarding whether UGPMs reach their potential in practice, respondents with no policy experience agreed or strongly agreed at a lower percentage than respondents with policy experience (37% vs. 53%).
- Respondents would rate communication between modelers and planners as lacking or significantly lacking at a lower degree if they had policy experience (87%), planning background (86%) or did not have a PhD (89%). The corresponding percentage for their complementary groupings was 97%.
- With respect to the modellers’ attention to practical considerations such as software compatibility and user friendliness, 56% of respondents with high GIS experience found these characteristics to be lacking or significantly lacking whereas for other respondents 76% found these characteristics to be lacking or sig-
nificantly lacking. The corresponding percentage was identical for the planning and modeling communities (63%).

- Finally, as expected, respondents with medium or high UGPM algorithmic experience found planners’ clarity on outcome expectations unclear or very unclear at a larger percentage than respondents with no or limited algorithmic experience (86% vs. 71%).

4. Discussion and Future Outlook

Urban land use change is an extensive, accelerating and important phenomenon. It is partially driven by human desire to improve quality of life, but urbanization often has serious environmental consequences. Modeling urban changes is an essential component of urban land use development and planning that can help determine sustainable management strategies. This task, however, is challenging due to unpredictability in decision making or social phenomena such as arbitrary zoning changes or population movement. At best, UGPMs can provide a scenario-based approach, where abrupt changes are integrated in the modeling prediction as external constraints. Another factor contributing to model uncertainty is the high variability in model validation techniques.

The number of temporal intervals used for validation can also significantly affect model quality. Furthermore, studies on one site may exhibit different dynamics, which in theory can be addressed through methods supporting spatial heterogeneity but in practise patterns are difficult to discern.

Realizing these important limitations, UGPMs can provide significant assistance in future planning exercises. Despite the wide acceptance of the potential of UGPMs (98% positive), the survey indicated that UGPMs do not always reach that potential in practice (43% positive). In the authors’ opinion, two major factors significantly constrain UGPM incorporation in planning decision processes. The first is the availability of UGPM results over extensive study areas. UGPMs are typically tested on limited sites (a strong preference was identified to local and regional studies), which makes generalized and wide-spread adoption difficult. This is partially due to cumbersome data acquisitions, where the data may either not be available in other sites or there is a significant collection and/or preprocessing cost. It is also supported by the survey finding that respondents with limited algorithmic and GIS experience found data availability as lacking at a higher proportion compared to GIS/Algorithmic experts (see Table 4), which could indicate that data availability may be constraining more model implementation rather than model development.

On the other hand, the popularity of the SLEUTH model is partially a result of the input variables’ simplicity and availability, a characteristic that not all UGPMs share. At times, researchers act as specialized consultants by developing UGPMs of high value to the specific study area but of limited general applicability. Instead, more general approaches should be pursued. Along these lines the proliferation of remotely sensed image along with derived products (e.g. the National Land Cover Dataset) could significantly assist in creation of input variables of wide-scale availability, but also in the validation process of UGPM performance over multiple time scales (e.g. 40 years with satellite imagery). In summary, we do not advocate for limited input variable creativity but data availability and processing costs should be considered.

In cases where UGPMs are available a second limiting factor becomes prominent, namely the transition of models from the development community to the user community (i.e. from modellers to planners). The survey indicated several transition barriers, including model awareness and lack of communication between interested parties. The “build it and they will come” approach has not been successful and stronger collaborations are necessary, an observation widely shared in the survey by both modellers and planners (see Table 4). Integration of planners early on in the model design process along with development of computer interfaces that increase user-friendliness should be essential components for successful transition. Furthermore, identification of clear expectations from the planning community to the modellers would also achieve positive results.

Despite the fact that both the survey respondents (67%) and the authors believe that prediction results are currently accurate enough for wide-spread implementation there is room for improvement. Many algorithms have been developed in order to simulate urban growth. Cellular automata provide a stochastic approach in modeling urban system dynamics and are widely applicable. This is further evident as cellular automata are present in almost half of the reviewed manuscripts. Looking into the future, the integration of stochastic and deterministic methods, which appear in many papers of cellular automata in this review, could improve simulation of urban growth complexity [263].

The lack of financial support for UGPM development (81%) and implementation (86% negative) is recognized and should serve as motivation for wiser investments from funding sources. A central repository of UGPM data, models and results could significantly increase theoretical development. Furthermore, the creation of benchmarking datasets would help both the modeling and planning communities assess prediction accuracy in a consistent and thorough manner. A conference bringing together modelers and planners for the purpose of creating guidelines for future UGPM characteristics could significantly advance this promising yet underutilized field.
REFERENCES


[79] C. Webster and F. Wu, “Coase, Spatial Pricing and Self-


[153] C. A. Jantz, S. J. Goetz, D. Donato and P. Claggett, “De-


Appendix

Table S1. Summary table with characteristics of the 156 reviewed manuscripts.

<table>
<thead>
<tr>
<th>Paper Authors</th>
<th>Publication Year</th>
<th>Methods</th>
<th>Data Type</th>
<th>Study Size</th>
<th>Spatial Heterogeneity</th>
<th>Spatial Autocorrelation</th>
<th>Temporal Resolution</th>
<th>Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aguayo et al.</td>
<td>2007</td>
<td>L</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>Yes</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>Al Kheder et al.</td>
<td>2008</td>
<td>CA, F</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>Al-Ahmadi et al.</td>
<td>2009</td>
<td>CA, F</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>Alkheder et al.</td>
<td>2006</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>30</td>
<td>NP</td>
</tr>
<tr>
<td>Allen and Lu</td>
<td>2003</td>
<td>L</td>
<td>LU-R</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>Almeida et al.</td>
<td>2008</td>
<td>CA, ANNs</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>14</td>
<td>50</td>
</tr>
<tr>
<td>Almeida et al.</td>
<td>2003</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Antonio</td>
<td>2001</td>
<td>PM</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Barredo &amp; Demicheli</td>
<td>2003</td>
<td>CA</td>
<td>LU-R</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>Barredo et al.</td>
<td>2003</td>
<td>CA, F</td>
<td>LU-R</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Batisani and Yarnal</td>
<td>2009</td>
<td>L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>7</td>
<td>250</td>
</tr>
<tr>
<td>Batty et al.</td>
<td>1989</td>
<td>F</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>NP</td>
<td>50</td>
</tr>
<tr>
<td>Benenson</td>
<td>1998</td>
<td>ABM</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
<td>40</td>
</tr>
<tr>
<td>Benenson et al.</td>
<td>2002</td>
<td>CA, ABM</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>Yes</td>
<td>40</td>
<td>NP</td>
</tr>
<tr>
<td>Berling-Wolf &amp; Wu</td>
<td>2004</td>
<td>CA</td>
<td>LU-V</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>20</td>
<td>60, 120, 240, 480, 1000</td>
</tr>
<tr>
<td>Besussi et al.</td>
<td>1998</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>3 - 5</td>
<td>30</td>
</tr>
<tr>
<td>Candau</td>
<td>2000</td>
<td>CA</td>
<td>BF, SE</td>
<td>International</td>
<td>No</td>
<td>No</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td>Chen et al.</td>
<td>2002</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>22</td>
<td>150</td>
</tr>
<tr>
<td>Cheng &amp; Masser</td>
<td>2004</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>7</td>
<td>100</td>
</tr>
</tbody>
</table>

Copyright © 2012 SciRes.


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Year</th>
<th>Region</th>
<th>Type</th>
<th>Training</th>
<th>Validation</th>
<th>Horizon</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claggett et al.</td>
<td>2004</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Clarke &amp; Gaydos</td>
<td>1998</td>
<td>CA, MC</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>90</td>
</tr>
<tr>
<td>Conway</td>
<td>2009</td>
<td>L</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>7</td>
</tr>
<tr>
<td>Crooks</td>
<td>2006</td>
<td>ABM</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Deng</td>
<td>2010</td>
<td>LA</td>
<td>S</td>
<td>S</td>
<td>Yes</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Deng et al.</td>
<td>2008</td>
<td>MV</td>
<td>BF, SE</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>13</td>
</tr>
<tr>
<td>Dietzel et al.</td>
<td>2005</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>Yes</td>
<td>No</td>
<td>60</td>
</tr>
<tr>
<td>Ding and Zhang</td>
<td>2007</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>38</td>
</tr>
<tr>
<td>Dragicevic &amp; Marceau</td>
<td>2000</td>
<td>FS</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Dubovyk et al.</td>
<td>2011</td>
<td>L</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>Yes</td>
<td>5</td>
</tr>
<tr>
<td>Engelen et al.</td>
<td>1999</td>
<td>CA</td>
<td>LU-V</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Fan et al.</td>
<td>2009</td>
<td>PA</td>
<td>LU-R, SM</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>24</td>
</tr>
<tr>
<td>Fang et al.</td>
<td>2005</td>
<td>CA, L</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>7</td>
</tr>
<tr>
<td>Feng &amp; Xu</td>
<td>1999</td>
<td>ANNs, F</td>
<td>BF, SE</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Feng et al.</td>
<td>2012</td>
<td>CA</td>
<td>BF, SM</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>13</td>
</tr>
<tr>
<td>Fragiokis &amp; Seto</td>
<td>2007</td>
<td>L</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Frankhauser</td>
<td>2008</td>
<td>F</td>
<td>UB</td>
<td>International</td>
<td>No</td>
<td>No</td>
<td>200</td>
</tr>
<tr>
<td>Geertman et al.</td>
<td>2007</td>
<td>CA</td>
<td>BF</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Goldstein et al.</td>
<td>2004</td>
<td>CA</td>
<td>LU-V,</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>93</td>
</tr>
<tr>
<td>Guan &amp; Clarke</td>
<td>2010</td>
<td>CA</td>
<td>UB</td>
<td>International</td>
<td>Yes</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Guan et al.</td>
<td>2005</td>
<td>ANNs, CA</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Haack and Rafter</td>
<td>2006</td>
<td>KVGIS</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>22</td>
</tr>
<tr>
<td>Han et al.</td>
<td>2009</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>21</td>
</tr>
<tr>
<td>Hashem et al.</td>
<td>2010</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>Hathout</td>
<td>2002</td>
<td>Markov</td>
<td>LU-V</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>29</td>
</tr>
<tr>
<td>He et al.</td>
<td>2006</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>13</td>
</tr>
<tr>
<td>Henriquez et al.</td>
<td>2006</td>
<td>Markov, MCE, CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Henriquez et al.</td>
<td>2006</td>
<td>CA, MCE, Markov</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Herold et al.</td>
<td>2005</td>
<td>LDA</td>
<td>SM</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>70</td>
</tr>
<tr>
<td>Hong et al.</td>
<td>2012</td>
<td>R, L</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Hoymann</td>
<td>2011</td>
<td>Land Use Scanner</td>
<td>SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>17</td>
</tr>
<tr>
<td>Hoymann</td>
<td>2010</td>
<td>L, Land Use Scanner</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Hu &amp; Lo</td>
<td>2007</td>
<td>L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>2009</td>
<td>L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>Yes</td>
<td>Yes</td>
<td>18</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>2009</td>
<td>SVM, L</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>Irwin et al.</td>
<td>2006</td>
<td>R</td>
<td>LU-V, LU-R</td>
<td>National</td>
<td>Yes</td>
<td>No</td>
<td>28</td>
</tr>
<tr>
<td>Jantz &amp; Goetz</td>
<td>2005</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Jantz et al.</td>
<td>2004</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Jantz et al.</td>
<td>2010</td>
<td>CA</td>
<td>BF, SE</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Jat et al.</td>
<td>2008</td>
<td>R</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>25</td>
</tr>
<tr>
<td>Jenerette &amp; Wu</td>
<td>2001</td>
<td>CA</td>
<td>LU-V, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>83</td>
</tr>
<tr>
<td>Kim and Batty</td>
<td>2011</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>Reference</td>
<td>Year</td>
<td>Model(s)</td>
<td>Scale(s)</td>
<td>Data</td>
<td>Sim</td>
<td>Type</td>
<td>No 1</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
<td>----------</td>
<td>----------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Kocabas &amp; Dragicevic</td>
<td>2007</td>
<td>CA, BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Li &amp; Yeh</td>
<td>2004</td>
<td>R, SM</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>Li &amp; Yeh</td>
<td>2002</td>
<td>ANNs, CA, BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Li &amp; Yeh</td>
<td>2000</td>
<td>CA, BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Li and Liu</td>
<td>2006</td>
<td>CA, BF</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>Li et al.</td>
<td>2011a</td>
<td>CA, ACO, BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
<td>90</td>
</tr>
<tr>
<td>Li et al.</td>
<td>2011b</td>
<td>CA, ACO, SIMs, BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Li et al.</td>
<td>2010</td>
<td>CA, BF</td>
<td>Regional</td>
<td>Yes</td>
<td>No</td>
<td>16</td>
<td>28,5</td>
</tr>
<tr>
<td>Lin et al.</td>
<td>2008</td>
<td>CA, L</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Liu &amp; Phinn</td>
<td>2003</td>
<td>CA, FS</td>
<td>S, S</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Liu &amp; Seto</td>
<td>2008</td>
<td>ANNs</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Liu and Feng</td>
<td>2012</td>
<td>CA, L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2008</td>
<td>CA, ACO, DT, BF</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>14</td>
<td>NP</td>
</tr>
<tr>
<td>Liu et al.</td>
<td>2010</td>
<td>CA, AIS, BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>Loibl &amp; Toetzer</td>
<td>2003</td>
<td>ABM</td>
<td>BF, SE</td>
<td>Region</td>
<td>No</td>
<td>No</td>
<td>31</td>
</tr>
<tr>
<td>Lopez et al.</td>
<td>2001</td>
<td>R, Markov</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>Luo and Kanala</td>
<td>2008</td>
<td>L</td>
<td>BF, SE</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>13</td>
</tr>
<tr>
<td>Luo and Wei</td>
<td>2009</td>
<td>L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>12</td>
</tr>
<tr>
<td>Magliocca et al.</td>
<td>2012</td>
<td>ABM</td>
<td>SE</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Mahajan &amp; Venkatadalarn</td>
<td>2009</td>
<td>ANNs, CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Maithani</td>
<td>2009</td>
<td>ANNs</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Mandelas</td>
<td>2007</td>
<td>CA, FS</td>
<td>BF</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Meaille &amp; Wald</td>
<td>1990</td>
<td>NM</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Mitsova et al.</td>
<td>2011</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>9</td>
</tr>
<tr>
<td>Moreno et al.</td>
<td>2009</td>
<td>CA</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>3</td>
</tr>
<tr>
<td>Munday et al.</td>
<td>2010</td>
<td>MCE</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>105</td>
</tr>
<tr>
<td>Nijs et al.</td>
<td>2004</td>
<td>CA</td>
<td>LU-R</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Oguz et al.</td>
<td>2007</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>28</td>
</tr>
<tr>
<td>Park et al.</td>
<td>2011</td>
<td>L, ANN</td>
<td>BF, SE</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>27</td>
</tr>
<tr>
<td>Petrov et al.</td>
<td>2009</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Pijanowski et al.</td>
<td>2002</td>
<td>ANNs</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>Pijanowski et al.</td>
<td>2005a</td>
<td>ANNs</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>15</td>
</tr>
<tr>
<td>Pijanowski et al.</td>
<td>2005b</td>
<td>ANNs</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>7</td>
</tr>
<tr>
<td>Poelmann &amp; Rompaey</td>
<td>2010</td>
<td>L, CA</td>
<td>LU-R</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>24</td>
</tr>
<tr>
<td>Poelmanns et al.</td>
<td>2010</td>
<td>L, CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>14</td>
</tr>
<tr>
<td>Pontius &amp; Malanson</td>
<td>2005</td>
<td>R, CA</td>
<td>LU-V</td>
<td>Regional</td>
<td>Yes</td>
<td>No</td>
<td>28</td>
</tr>
<tr>
<td>Portugali et al.</td>
<td>1994</td>
<td>CA</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Rafice et al.</td>
<td>2009</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>34</td>
</tr>
<tr>
<td>Rajan &amp; Shibasaki</td>
<td>2001</td>
<td>ABM</td>
<td>LU-V</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Rand et al.</td>
<td>2002</td>
<td>ABM</td>
<td>BF, SE</td>
<td>NP</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Samat</td>
<td>2007</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Region</td>
<td>Type</td>
<td>Dataset</td>
<td>Validation</td>
<td>Validation</td>
<td>N1</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>---------</td>
<td>------------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td>Sang et al.</td>
<td>2011</td>
<td>CA</td>
<td>LU-R</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>7</td>
</tr>
<tr>
<td>Sangwongse et al.</td>
<td>2005</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>48</td>
</tr>
<tr>
<td>Seto &amp; Kaufmann</td>
<td>2003</td>
<td>R</td>
<td>BF, SE</td>
<td>Regional</td>
<td>Yes</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Shan et al.</td>
<td>2008</td>
<td>CA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>30</td>
</tr>
<tr>
<td>Shariff et al.</td>
<td>2010</td>
<td>R</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Shen</td>
<td>2002</td>
<td>F</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>200</td>
</tr>
<tr>
<td>Silva and Clarke</td>
<td>2005</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>30</td>
</tr>
<tr>
<td>Stevens et al.</td>
<td>2007</td>
<td>CA</td>
<td>LU-V</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Sullivan</td>
<td>2001</td>
<td>CA</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Syphard et al.</td>
<td>2005</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>55</td>
</tr>
<tr>
<td>Syphard et al.</td>
<td>2009</td>
<td>SI</td>
<td>BF, SE</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Syphard et al.</td>
<td>2011</td>
<td>CA</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>39</td>
</tr>
<tr>
<td>Tannier et al.</td>
<td>2011</td>
<td>F</td>
<td>BF</td>
<td>International</td>
<td>No</td>
<td>No</td>
<td>112</td>
</tr>
<tr>
<td>Tayyebi et al.</td>
<td>2011</td>
<td>ANNs</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>12</td>
</tr>
<tr>
<td>Thapa and Murayama</td>
<td>2011</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>22</td>
</tr>
<tr>
<td>Thomas et al.</td>
<td>2008</td>
<td>F</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Thomas et al.</td>
<td>2010</td>
<td>F</td>
<td>UB</td>
<td>International</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Torrens</td>
<td>2006</td>
<td>ABM</td>
<td>UB</td>
<td>Region</td>
<td>No</td>
<td>No</td>
<td>160</td>
</tr>
<tr>
<td>Triantakonstantis</td>
<td>2012</td>
<td>F</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Triantakonstantis et al.</td>
<td>2011</td>
<td>DT</td>
<td>BF, SE</td>
<td>Local</td>
<td>Yes</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>Vaz et al.</td>
<td>2009</td>
<td>MCE, CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>28</td>
</tr>
<tr>
<td>Vaz et al.</td>
<td>2012</td>
<td>CA</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>Verburg et al.</td>
<td>2004</td>
<td>L, CA</td>
<td>LU-R</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>7</td>
</tr>
<tr>
<td>Verburg et al.</td>
<td>2002</td>
<td>L</td>
<td>BF, SE</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>Vliet et al.</td>
<td>2009</td>
<td>CA</td>
<td>LU-R</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Wang &amp; Mountrakis</td>
<td>2011</td>
<td>ANNs, L</td>
<td>BF, SE</td>
<td>Local</td>
<td>Yes</td>
<td>Yes</td>
<td>20</td>
</tr>
<tr>
<td>Wang and Li</td>
<td>2011</td>
<td>CA, RBFN</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Weber</td>
<td>2003</td>
<td>SIM</td>
<td>LU_R</td>
<td>Region</td>
<td>No</td>
<td>Yes</td>
<td>12</td>
</tr>
<tr>
<td>Weber and Puissant</td>
<td>2002</td>
<td>Potential</td>
<td>LU-R</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Webster &amp; Wu</td>
<td>2001</td>
<td>CA</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Wentz et al.</td>
<td>2010</td>
<td>SI</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>Yes</td>
<td>20</td>
</tr>
<tr>
<td>Westervelt et al.</td>
<td>2011</td>
<td>RUG</td>
<td>BF</td>
<td>Regional</td>
<td>No</td>
<td>No</td>
<td>20</td>
</tr>
<tr>
<td>White &amp; Engelen</td>
<td>2000</td>
<td>CA</td>
<td>LU-R</td>
<td>National</td>
<td>No</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>Wu</td>
<td>2002</td>
<td>CA</td>
<td>LU-V</td>
<td>Local</td>
<td>Yes</td>
<td>Yes</td>
<td>20</td>
</tr>
<tr>
<td>Wu</td>
<td>1998</td>
<td>CA, MCE</td>
<td>LU-R</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Wu &amp; Webster</td>
<td>2000</td>
<td>CA</td>
<td>S</td>
<td>S</td>
<td>No</td>
<td>No</td>
<td>S</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>2009a</td>
<td>LA</td>
<td>BF, SE</td>
<td>Local</td>
<td>No</td>
<td>Yes</td>
<td>NP</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>2009b</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>2010</td>
<td>CA</td>
<td>BF</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>16</td>
</tr>
<tr>
<td>Wu et al.</td>
<td>2006</td>
<td>Markov, R</td>
<td>LU-R</td>
<td>Local</td>
<td>No</td>
<td>No</td>
<td>20</td>
</tr>
</tbody>
</table>
Continued

| Authors            | Year | Data Type | Location | S-RA | R-RA | L-RA | CA | ANNs | DT | F | ABM | ACO | RBFN | SIMs | AIS | FS | MCE | MC | KVGIS | RUG | MV | NM | PA | PM | SIM | SLDM | SVM | SI |
|--------------------|------|-----------|----------|------|------|------|----|------|----|---|-----|-----|------|------|-----|----|    |     |     |     |     |     |    |    |    |    |     |      |     |    |
| Xie et al.         | 2009 | CA        | BF, SE   | Local| No   | No   | 16 | 60   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Xian et al.        | 2005 | CA        | BF       | Regional| No   | No   | 46 | 60, 120, 240 |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Xie                | 1996 | CA        | BF, SE   | Local| No   | No   | NP | 50   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Xu et al.          | 2006 | CA        | BF, SE   | Regional| No   | No   | 21 | NP   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yang               | 2008 | CA        | BF, SE   | Regional| No   | No   | 16 | NP   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yang & Lo          | 2003 | CA        | BF       | Regional| No   | No   | 26 | 240  |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yang et al.        | 2012 | CA, Markov, ACO | BF | Regional| No   | No   | 20 | 30   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yang et al.        | 2008 | CA        | BF       | Local | No   | No   | 16 | NP   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yeh & Li           | 1998 | SLDM      | BF, SE   | Regional| No   | No   | 5  | 30   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Ying et al.        | 2009 | CA, MCE, Markov | BF, SE | Local | No   | No   | 59 | 500  |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Yu and Qingyun     | 2011 | L         | BF, SE   | Regional| No   | No   | 5  | 100  |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Zhang et al.       | 2011 | CA, Markov | BF       | Regional| No   | No   | 10 | 30   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Zhang et al.       | 2010 | ABM       | BF, SE   | Local | No   | No   | 7  | 30   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |
| Zhao and Murayama  | 2007 | CA        | LU-R     | Regional| No   | No   | 5  | 10   |    |   |     |     |      |      |     |    |    |    |    |     |      |     |    |    |

Table S2. Abbreviation legend for Table S1.

Methods | Data Types
---|---
R-RA | Regressive-Autoregressive | BF | Biophysical
L-RA | Logistic-Autologistic | SE | Socioeconomic
CA | Cellular Automata | LU-V | Land Use Vector
ANNs | Artificial Neural Networks | LU-R | Land Use Raster
DT | Decision Trees | SM | Spatial Metrics
F | Fractals | UB | Urban Boundaries
ABM | Agent Based Models |
ACO | Ant Colony Optimization |
RBFN | Radial Basis Function Neural |
SIMs | Swarm Intelligent Models |
AIS | Artificial Immune System |
FS | Fuzzy Sets |
MCE | Multicriteria Evaluation |
MC | Monte Carlo |
KVGIS | Kathmandu Valley GIS |
RUG | Regional Urban Growth |
MV | Multivariate Analysis |
NM | Numerical Model |
PA | Pattern Analysis |
PM | Potential Model |
SIM | Spatial Interaction Model |
SLDM | Sustainable Land Development Model |
SVM | Support Vector Machines |
SI | Spatial Interpolation |

S = Simulated; NP = Not Provided.
Table S3. PRISMA 2009 checklist regarding the analysis data and manuscript location.

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>#</th>
<th>Checklist Item</th>
<th>Reported on Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>1</td>
<td>Identify the report as a systematic review, meta-analysis, or both.</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured summary</td>
<td>2</td>
<td>Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.</td>
<td>1</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>3</td>
<td>Describe the rationale for the review in the context of what is already known.</td>
<td>2</td>
</tr>
<tr>
<td>Objectives</td>
<td>4</td>
<td>Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).</td>
<td>2</td>
</tr>
<tr>
<td>METHODS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol and registration</td>
<td>5</td>
<td>Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.</td>
<td>3</td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>6</td>
<td>Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.</td>
<td>3</td>
</tr>
<tr>
<td>Information sources</td>
<td>7</td>
<td>Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.</td>
<td>3</td>
</tr>
<tr>
<td>Search</td>
<td>8</td>
<td>Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.</td>
<td>3</td>
</tr>
<tr>
<td>Study selection</td>
<td>9</td>
<td>State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).</td>
<td>3</td>
</tr>
<tr>
<td>Data collection process</td>
<td>10</td>
<td>Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.</td>
<td>3</td>
</tr>
<tr>
<td>Data items</td>
<td>11</td>
<td>List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.</td>
<td>2</td>
</tr>
<tr>
<td>Risk of bias in individual studies</td>
<td>12</td>
<td>Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.</td>
<td>N/A</td>
</tr>
<tr>
<td>Summary measures</td>
<td>13</td>
<td>State the principal summary measures (e.g., risk ratio, difference in means).</td>
<td>2</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>14</td>
<td>Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., F) for each meta-analysis.</td>
<td>2</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>15</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).</td>
<td>N/A</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>16</td>
<td>Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.</td>
<td>4</td>
</tr>
<tr>
<td>RESULTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study selection</td>
<td>17</td>
<td>Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.</td>
<td>2,3</td>
</tr>
<tr>
<td>Study characteristics</td>
<td>18</td>
<td>For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.</td>
<td>Table S1</td>
</tr>
<tr>
<td>Risk of bias within studies</td>
<td>19</td>
<td>Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).</td>
<td>N/A</td>
</tr>
<tr>
<td>Section/topic</td>
<td>#</td>
<td>Checklist item</td>
<td>Reported on page</td>
</tr>
<tr>
<td>Results of individual studies</td>
<td>20</td>
<td>For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.</td>
<td>N/A</td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>21</td>
<td>Present results of each meta-analysis done, including confidence intervals and measures of consistency.</td>
<td>N/A</td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>22</td>
<td>Present results of any assessment of risk of bias across studies (see Item 15).</td>
<td>N/A</td>
</tr>
<tr>
<td>Additional analysis</td>
<td>23</td>
<td>Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Copyright © 2012 SciRes. JGIS
Text S1. Identification of Input Data Sources of UGPMs

In the last decade significant progress has been made in urban data collections, mostly led by governmental agencies and recently followed by private entities. Examples of urban data acquisition include tax offices, transportation departments, utility companies, and emergency management departments [1]. Despite this recent information explosion, collected data may not satisfy urban modeling requirements related for example to spatial and temporal accuracy [2]. In addition, geographical isolation of most collecting efforts makes it challenging to compile data from multiple sites to statistically strengthen confidence on obtained results. Furthermore, the temporal extent of data is constrained, with data from older periods often aggregated at coarser scales.

Data used as input variables in UGPMs are mainly acquired by remotely sensed data and field/sampled data. Each of these sources is discussed in detail in the following subsections.

Remotely Sensed Based Data

Remotely sensed data derived from satellite images and aerial photographs can significantly lower acquisition costs while allowing for broader spatial coverage. An image forms a permanent record that could be revisited at a later time for further information extraction. Even though the United States offers the longest temporal record of remotely sensed data for civilian applications (Landsat series), data collections from other countries have started to offer comparable or improved capabilities, especially for recent data. In addition, several private companies have launched satellite sensors providing costly data but of significantly better spatial resolution, in some cases offering pixel sizes smaller than 1 m.

One of the uses of remotely sensed imagery is the automatic creation of land cover maps. Of particular interest to UGPMs, several remote sensing approaches have been proposed for urban monitoring [3] using methods such as multivariate regression [4,5], spectral mixture analysis [6,7], and machine learning models [8,9]. Even though the majority of these algorithms may reach high classification accuracy, a typical problem is that this accuracy is not spatially distributed in a uniform manner (as shown in [10]) which may generate substantial errors in UGPM calibration.

Landscape topography is another important factor in urban growth. Digital elevation models (DEMs) are derived from stereoscopic analysis of remotely sensed data providing information on elevation, slope, aspect and orientation of the land surface. Because terrain does not change rapidly, DEMs are typically updated every 5 to 10 years and in many UGPMs are assumed to remain constant over time.

In addition to satellite imagery, aerial photographs have been available for more than a century and therefore may extend significantly the temporal coverage of UGPM data, especially if combined with satellite images [11]. Different spatial resolutions along with geometric and radiometric distortions make the integration of aerial photographs with satellite images cumbersome.

Visual interpretation is often implemented to produce a plethora of UGPM inputs. For example, the city of Denver, Colorado obtains its parcel data from visual interpretation of panchromatic aerial photography in order to build a billing system for the wastewater service charge [12]. In the UK, Ordnance Survey produces cadastral data for the entire country, using remote sensed data with ground verification. Numerous transportation departments update their road network vectors, an important urbanization indicator, using remotely sensed data and manual digitizing.

Further spatial processing often takes place on information extracted from remotely sensed data. Typical GIS functions indicating proximity to feature (e.g. distance to nearest road) or feature density (e.g. road density) create high value UGPM inputs [13]. In addition, direct analysis of urban patterns using spatial metrics such as patch size, patch density, edge density, contagion index and fractal
dimension have been used as independent variables in UGPMs [14-20].

Satellite images can be also used to derive some useful UGPM inputs expressing socioeconomic data, such as population estimation and life quality indicators (e.g. house value, median family income, average rent) [21,22]. These data can be as accurate as traditional census methods if calibration in situ takes place. Another example of population distribution is LandScan. Produced by Oak Ridge National Laboratory, LandScan is using GIS and Remote Sensing to create a global population distribution with approximate pixel size 1 Km.

Field and Sampling Based Data

Field and sampling based data acquire the precise location of specific points, i.e., coordinates, where GPS receivers can play an important role. The requirements for the types of data as well as their accuracy are different from one study to another, because they are based on the general philosophy of each study. The researcher will decide the kind of data and their accuracy according to expert knowledge and experience of the study area. Moreover, each study requires different spatial and temporal resolutions. For example, different spatial dimensions could be applied in local, regional, national or international levels.

Field surveys help establish a relationship between remote sensing data and the real (ground) environment. Ground data can provide a first-hand view of urban development, useful for understanding urban growth dynamics. The ground control testing corrects the remotely sensed imagery in positions where there is no clear pixel information, e.g. land use boundaries, where spectral mixing is occurring [23].

The classification of built areas requires field-based data. For example, from a field survey, residential and non-residential built areas can be easily discriminated into the following categories: individual houses, buildings, estate housing (residential) and facilities, public buildings, industrial and commercial units (non-residential) [24]. Moreover, some other data can be incorporated to discriminate residential areas such as single-family-multi-family residential [25] and high density-medium density residential, hotel-motel-resort and large lot-small lot residential [26]. [27] acknowledges the importance of human use in land use changing and selected the following field-based data for analyzing urban growth: 1) population density, 2) gross value of industrial output, 3) gross value of agricultural output and d) gross domestic production. Variables related to technology, political and economic institutions as well as cultural values are difficult to include in these analyses because of the complexity of the statistical methods [28].

Field-based data can also provide information about economic activity leading to sprawl [28]. For example, development of service centers such as shopping centers, cafeterias, and bars along a road could become an attraction point for urban development. Therefore, field-based data becomes valuable information for urban planners and decision makers, once they can visualize such types of growth patterns.

Census and other types of data can be collected by government and other local planning agencies [29,30]. For example, the US Census Bureau provides census and survey data, the most common of which is the population census of the US, called the decennial census, which is conducted every 10 years from each household. Decennial census collects data about income, education, homeownership etc. Additionally, the American Community Survey produces population and housing data every year instead of every ten years. Economic census data of 2007 are available concerning business activities in industries and communities across US; data useful not only for policy planners but also for businesses in order to decide for example a new factory or office location. Annual economic surveys are conducted covering annually, quarterly and monthly time periods for various economy sectors [31]. Moreover, some other census data are also available from US Census Bureau such as statistics about governmental activities, demographic, social, economic and housing characteristic of US population, business and industries. The above census data are collected in different spatial scales: census tract, census block and block groups. Census block is the smallest geographic area, while a block group is a combination of census blocks and a subdivision of a census tract. Census tracts are subdivisions contained within counties. Generally, census data are usually limited in their temporal resolution and consistency, as well as they described by restricted availability in many areas, especially outside developed countries [2].

REFERENCES FOR TEXT S1


