I

RESEARCH ARTICLES

DEVELOPING AN AUTOMATED METHOD FOR THE APPLICATION OF LIDAR IN IUMAT LAND-USE MODEL: ANALYSIS OF LAND-USE CHANGES USING BUILDING-FORM PARAMETERIZATION, GIS, AND ARTIFICIAL NEURAL NETWORKS

Mohamad Farzinmoghadam,[1] Nariman Mostafavi,[2]
Elisabeth Hamin Infield,[3] and Simi Hoque[4]*

ABSTRACT

Predicting resource consumption in the built environment and its associated environmental consequences is one of the core challenges facing policy-makers and planners seeking to increase the sustainability of urban areas. The study of land-use change has many implications for infrastructure design, resource allocation, and urban metabolism simulation. While most urban models focus on horizontal growth patterns, few investigate the impacts of vertical characteristics of urbanscapes in predicting land-use changes. In this paper, Building-form variables are introduced as a new determinant factor for investigating effects of vertical characteristics of an urbanscape in predicting land-use change. This work outlines an automated method for generating building-form variables from Light Detection and Ranging (LIDAR) data by using Density-Based Spatial Clustering and normal equations. This paper presents a Land-Use Model that uses Remote Sensing, GIS, and Artificial Neural Networks (ANNs) to predict urban growth patterns within the IUMAT framework (Integrated Urban Metabolism Analysis Tool), which is an analytical platform for quantifying the overall sustainability in the urbanscape. The town of Amherst in Western Massachusetts (for the period of 1971–2005) is used as a case study for testing the model. By isolating the weights of each explanatory variable in models, this study highlights the influence of building geometry on future development scenarios.

KEYWORDS

Land-Use Modeling, Spatial Analysis, LIDAR, Building form Extraction, Artificial Neural Networks

1. Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, mfarzinmoghadam@wpi.edu
2. Department of Civil, Architectural, and Environmental Engineering, Drexel University, sm3892@drexel.edu
3. Department of Regional Planning and Landscape Architecture, University of Massachusetts, Amherst, emhamin@larp.umass.edu
4. Department of Civil, Architectural, and Environmental Engineering, Drexel University, simi@coe.drexel.edu (*Author to whom correspondence should be addressed)

1. INTRODUCTION

Mountain snowpack declination (Mote et al., 2005), unprecedented drought in California (Mann & Gleick, 2015), and Atlantic hurricane trends (Mann & Emanuel, 2006) are some examples of the changing climate. The U.N. Climate Chief has stated "This transformation is unstoppable" (UN releases draft agreement on climate change, 2015). Human activities and rapid urbanization are two major sources of GHG emissions (International Energy Agency, 2008, Grimm et al., 2008). While the world population living in urban or suburban areas is expected to grow 25 percent between 2011 and 2050 (Crossette et al., 2011), more studies provide evidence highlighting strong associations between land-use change and climate change (Melton et al., 2016, Heald & Spracklen, 2015, Pielke et al., 2002). Planners develop policies for minimizing environmental impacts of land-use change like air pollution (Mage et al., 1996), waste (Kennedy et al., 2009), and soil erosion (Chen, 2007) in urban and suburban districts. However, understanding the processes and parameters involved in land-use transition remains one of the most challenging tasks in the planning community. To address this issue, planners have employed advanced methods such as Cellular Automata and Artificial Neural Networks to capture land-use change. Land-use models improve our perception of the causes and consequences of existing spatial patterns, and the extent of future changes (Verburg et al., 2004).

2. LITERATURE REVIEW

2.1 Existing Land-use Models and Modeling Approaches

Early land-use models with deterministic approaches concentrated solely on deforestation modeling (Lambin, 1997). More recent methods implement dynamic methods to simulate complex land cover changes such as urbanization (Carrero et al., 2014). Land-use and urban models can be categorized based on modeling approaches: spatial approaches, dynamics of time and scale, and planning applications (Silva & Wu, 2012). These models investigate the interaction of involved parameters at a micro scale (e.g. TLUMIP (Weidner et al., 2009), UrbanSim (Waddell et al., 2003), ILUMASS (Wagner & Wegener, 2007)), or a macro scale (e.g. LTM (Pijanowski et al., 2002)), or at multiple scales (e.g. WiVsim (Spahn and Lenz, 2007). Land-use changes and urban growth models in the long term (e.g. FEARLUS, Cioffi-Revilla & Gotts, 2003), medium term (e.g. CLUE-S (Verburg et al., 2002), or short term are developed for different planning tasks (Silva & Wu, 2012).

Cellular Automata (CA) based models (e.g. SLEUTH (Jantz et al., 2010), iCity (Stevens et al., 2007), Metronamica (van Delden et al., 2005), Agent-Based Models (e.g. STAU-Wien (Loibl & Toetzer, 2003), SIMPOP (Sanders et al., 1997), Genetic Algorithms (Tseng et al., 2008), and Artificial Neural Networks (e.g. ART-MMAP (Liu & Seto, 2008; Omrani et al., 2012; Pijanowski et al., 2014) offer intelligent approaches to modeling land-use changes. CA models and Agent-Based Models work with spatial data while other methods need to be integrated with other spatial techniques (Carrero et al., 2014). CA is used for capturing the long term conversion of non-urban to urban land in urban growth models. In this dynamic discrete modeling technique, each cell responds to the same set of rules based on states of neighboring cells while ignoring the global spatial context and characteristics of the built environment (Vanegas et al., 2010). In contrast to CA, the interaction between agents in an urban context is included in Agent-Based Models. The latter method still applies simple behavioral rules influenced by localized context, similar to CA models, but there are limited attempts for validating

models by observed data (Vanegas et al., 2010). Genetic Algorithms is a method used to generate and optimize a set of parameters for complex problems with high levels of uncertainty (Jenerette & Wu, 2001); however, the logic behind the rules is difficult to parse (Tseng et al., 2008).

Similar to Genetic Algorithms, Artificial Neural Networks (ANNs) apply "machines with the mathematical logic capacity of human neural systems" to solve sophisticated problems such as land-use changes and urban growth (Basse et al., 2014). ANNs are interconnected networks of neurons comprised of input, output, and hidden layers. Interrelational weights between nodes are updated by implementing different algorithms and an internal transfer function (Aisa et al., 2008). Users are responsible for defining the number of hidden layers, regularization value, learning rate, learning iteration numbers, and data encoding techniques (Tseng et al., 2008).

ANNs, with the aid of spatial analysis methods, are capable of simulating land-use and urban changes by integrating the variety of environmental, social, and political variables. For example, in ART-MMAP, Liu & Seto (2008) predict urban growth by learning from past trends and regularized weights of socioeconomic variables. Tayyebi et al. (2011) use an ANNs-based model for predicting urban growth boundaries, based on variables such as built areas, accessibility to roads, green areas, and service stations. Maithani (2009) proposed coupling ANNs with GIS and remote sensing measurements for generating site variables and reducing subjectivity in urban growth modeling. The non-parametric characteristic of ANNs models may be considered an alternative to estimating land-use transition probabilities in CA simulation models (Almeida et al., 2008). Unlike common statistical methods, ANNs do not make assumptions about the data distribution and can reduce the subjectivity in modeling complex phenomena such as urban growth where there is high nonlinearity between variables (Maithani, 2009). ANNs also perform better in predicting land-use classes change compared to other well-known non-linear models like Classification and Regression Trees (CART) and Multivariate Adaptive Regression Splines (Tayyebi & Pijanowski, 2014). Integration of Multiple Neural Networks in urban growth models could improve the modeling accuracy and enhance modeling capacity in capturing spatial heterogeneity (Wang & Mountrakis, 2011). However, calibration and validation of ANNs models remains a challenge (Basse et al., 2014). Although Triantakonstantis & Mountrakis (2012) believe there is no need for multicollinearity and spatial correlation assumption in ANNs analysis, others like Garg and Tai (2012) assume that ANNs models cannot automatically deal with data interrelationships in training data. One of the main weakness is the "black box" behavior of ANNs models where users cannot specifically extract rules or conclusions from the learning process (Triantakonstantis & Mountrakis, 2012).

The integration of land-use vector data and urban form in urban simulation may improve our understanding of human behavior (Silva & Wu, 2012) in different areas such as transportation and travel behavior (Chao & Qing, 2011; Ewing & Cervero, 2001; Newman & Kenworthy, 2006; Cervero & Gorham, 2009), accessibility (Handy & Clifton, 2001), energy use (Ewing & Rong, 2008), life cycle analysis (Norman et al., 2006), ecological assessments (Bereitschaft & Debbage, 2013), and environmental impacts (Anderson et al., 1996; Ellis, 2002; Ewing & Rong, 2008; Frey, 2003; Gordon & Richardson, 1997; Newton, 2000; Williams., 2011). These studies confirm that the integration of a comprehensive set of land-use and urban form variables (e.g. Hamidi et al., 2015; Chao & Qing, 2011) improve prediction of complex problems, and even provide results that contradict conventional wisdom. For example, Glaeser and Kahn (2010) found that more restrictive land-use regulations increase urban GHG emissions by promoting new developments in the periphery of cities. Urban form variables such

as concentration, dispersal, mixed use (Buxton, 2000; Newton et al., 2000), urban continuity (Bechle et al., 2011), centrality, compactness index, and open space ratio (Huang et al., 2007), in combination with urban sprawl index (Lopez and Hynes, 2003; Sutton, 2003) are also used in analyzing the metabolic performance of cities. For example, Bereitschaft and Debbage (2013) study the relation between urban continuity and shape complexity indices with air pollution. These indices are also integrated into landscape metrics for exploring evolutions of land-use and urban growth (Ji et. al, 2006; Luck & Wu, 2002). In other studies, multi-dimensional sprawl indices (Hamidi et al., 2015, Ewing et al., 2003) integrated with socioeconomic variables and urban form indices are applied for measuring transportation.

Urban and building-form indices play an essential role in modeling human behavior and urban systems. Three-dimensional urban geography research performs better when compared to two-dimensional analysis in capturing the complexity of the built environment (Thill et al., 2011). Researchers investigate the effects of a building's height on different areas such as heat island, rainwater runoff, pollution, and habitability (Lin et al., 2014). For example, building height influences the rainfall run-off process in an urban environment. Integration of building heights in urban hydrological models enhances modeling capacity for capturing the run-off and plays a significant role in stormwater management (Isidoro & Lima, 2014). Combined with other morphological properties of buildings, such as roof area and compactness, building height is used to extract urban land-use categories (Barnsley et al., 2003). The vertical aspect of the urbanscape influences parameters such as humidity, wind direction and speed, and solar radiation; these effects create different microclimates and impact thermal comfort within urban districts (Palme & Ramírez, 2013). In addition to thermal implications, building heights are critical in analyzing visual and acoustical effects of urbanscape. Ko et al. (2011) evaluated impacts of building heights on road traffic noise for identifying areas with excessive environmental noise. Moreover, the vertical growth of urbanscapes significantly influences "livability" in urbanscapes (Lin et al., 2014). A comprehensive 3D geospatial database of an urbanscape not only is a valuable resource for analyzing different aspects of urban systems, but also is useful during emergencies by reducing the response time on multi-level structures (Lee & Zlatanova, 2008; Kwan & Lee, 2005).

2.1 Research questions and Paper structure

Integrating buildings geometric information in land use models is a challenging task. Lack of sufficient databases and advanced methodologies could be among the reasons for the gap in the existing literature. While most urban models focus on horizontal growth patterns, few investigate the impacts of vertical characteristics of the urbanscape into predicting land-use changes. In this study, we explore the use of building-form indices extracted from LIDAR data in land-use modeling. This approach is integrated into the IUMAT Land-Use Model (IUMAT-LUM). The IUMAT, Integrated Urban Metabolism Analysis Tool, is an analytical platform for quantifying the overall sustainability in the urbanscape (Mostafavi et al., 2014). IUMAT-LUM with an ANNs simulation platform models urban growth and future development patterns.

In contrast to environmental, physical, institutional, and cultural data, many planning and design agencies do not have the resources or knowledge to develop a comprehensive vectorized database of urban and building geometry. LIDAR data, which is the 3-dimensional measurement of the built environment, is a valuable resource for creating building-form parameters. Parameterization of roof shape provides enough information about most of the architectural

characteristics of a building such as geometric prototype, footprints, site coverage, courtyard ratio, the number of floors, height, and building orientations. Therefore, building-form is identified as roof shape in this study.

Our research introduces an automated method for generating building-form variables from LIDAR measurements. The main research questions are as follows:

1. How can Light Detection and Ranging data (LIDAR) be transformed into a determinant factor (building-form indices) in land-use modeling?
2. Do building-form indices in combination with other spatial explanatory variables improve the predictive power of land-use modeling?

The remainder of this paper is organized into three sections. Section 3 on Research Methods outlines approaches for generating building geometric variables from LIDAR measurements and other explanatory variables from GIS vectorized databases. It summarizes the structure of an IUMAT-LUM simulator. Section 4 (Results and Discussion) describes the study area and databases used in this study. We explain the implementation of the proposed IUMAT-LUM framework to the town of Amherst. The impacts of the buildings' geometric information in predicting changes in the pattern of the built environment are also explored. In the final and concluding section (5), we discuss the potential for IUMAT-LUM to integrate into other urban metabolism and land-use policy studies.

3. RESEARCH METHODS

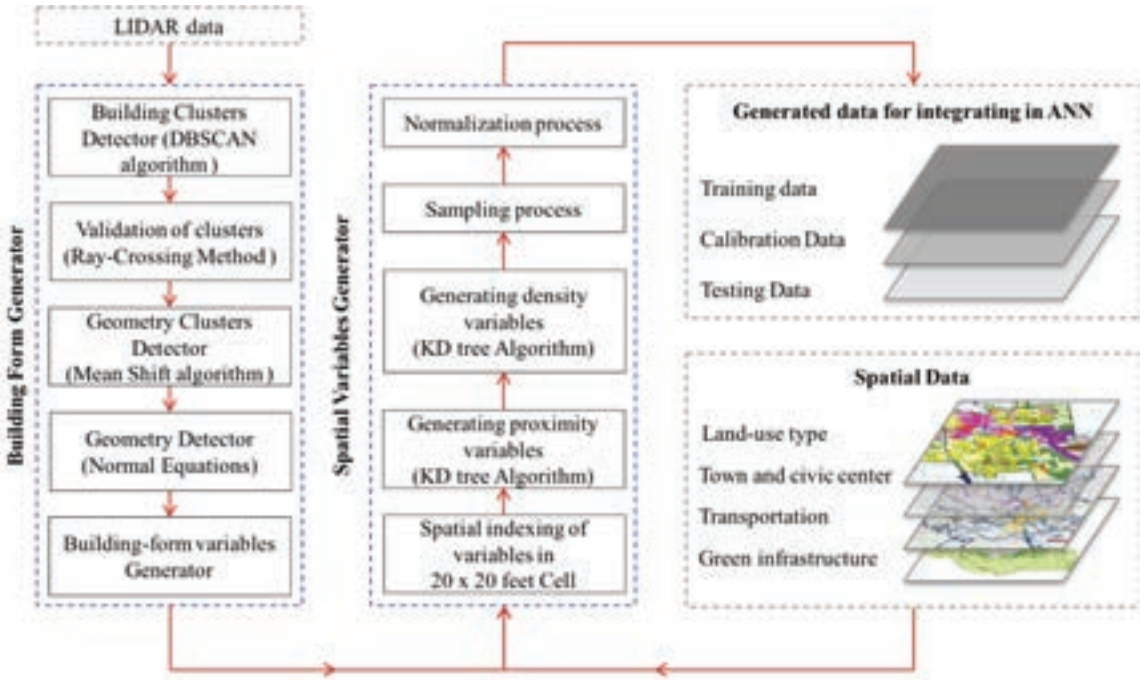
3.1 IUMAT-LUM framework

Land-use models are used to investigate the relationships between socioeconomic characteristics and the built environment in order to analyze and predict land-use change. In addition to physical, institutional, cultural, and environmental parameters commonly studied, in the IUMAT-LUM, building and urban form are used as new determinant factors for modeling land-use change. We investigate the influence of building-form indices extracted from LIDAR measurements on patterns of new developments. The three main components of the IUMAT-LUM framework are the Building-Form Generator, the Spatial Variables Generator, and the Simulator (Figures 1 and 6). The Building-Form Generator applies different algorithms for extracting building-form variables from LIDAR data. The Spatial Variables Generator compiles these variables and GIS vector data into a spatial grid system and employs spatial functions for calculating density, proximity, and land cover estimates. Generated explanatory variables are split into training, calibration, and test data for training and calibrating the model in the Simulator. And finally, the simulator applies optimized ANNs training weights for predicting land-use changes in an urban block. The algorithms of IUMAT-LUM are written in Python to generate, process, and analyze the data.

3.2 Building-Form Generator

The Building-Form Generator in the IUMAT-LUM (Figure 1) generates a vectorized database about the architectural characteristics of buildings such as geometric prototype, footprint, site coverage, mass to space ratio, the number of floors, height, and building orientations. The building-form complexity index and Building height variables are extracted from LIDAR data, processed in Spatial Variables Generator and the IUMAT-LUM simulation platform.

FIGURE 1. Conceptual framework of IUMAT-LUM data preparation process.



Airborne LIDAR is a remote survey technology, which generates 3D points with coordinates (*x* and *y*) and elevation information (*z*) about the natural and built environment without any projection and shadow distortions (Yan et al., 2015). Compared to aerial and satellite images, LIDAR data is more useful for extracting building 3D models especially when dealing with large sets of objects (Zhang et al., 2006). Building and urban 3D models have many applications in planning and urban design such as measuring energy performance, creating virtual urban models, and assessing urban heat island (Jensen, 2009). For the last decade, researchers have developed several algorithms for converting LIDAR data to building 3D models at the urban scale (Grammatikopoulos et al., 2015; Yan et al., 2015; Palmer & Shan, 2005; You et al., 2003). Schwalbe et al. (2005) categorized these algorithms into model-driven and data-driven methods. Data-driven algorithms identify planes in cloud points or combine LIDAR measurements with other data sources like imagery to extract building 3D models, while in model-driven approaches, limited predefined geometry models are fitted to the LIDAR measurements.

The Building-Form Generator integrates a model-driven approach and employs five steps towards converting LIDAR measurements to building-form variables. In the first step, the Building Clusters Detector applies a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm introduced by Ester et al. (1996). LIDAR points provided by public agencies can have a classification type such as *ground*, *low vegetation*, or *building* assigned to it, which define the type of object represented by that point. The Clusters Detector selects LIDAR points in the *building* class within an urban block, classifies points with higher density together as buildings, and sets low-density regions as outliers. The DBSCAN algorithm is suitable for detecting arbitrary shapes with high efficiency on large databases without specifying numbers of clusters for the algorithm. The Cluster Detector uses the DBSCAN algorithm in Python Scikit-learn library. Users are responsible for adjusting *min-samples* and *eps* parameters,

which define the minimum density of points in clusters (of buildings). The *min-sample* specifies the minimum number of points in a region and the *eps* parameter represents the maximum distance allowed by the algorithm between points in each cluster.

If building boundary vector data exist in spatial databases, the Building-Form Generator checks the positions of cluster points relative to those boundaries. In step 2, using a Ray Crossing Number method (Shimrat, 1962), the model assigns zero to points inside a building footprint and one to points outside. The Cluster Detector outcomes are then used in the Geometry Cluster Detector (step 3), which applies Mean Shift and Fuzzy clustering algorithms for identifying geometric components in each building cluster. Mean shift is a non-parametric technique for detecting modes of a density (Eq. 1 and Eq. 2) and was originally proposed for image segmentation and analysis of multidimensional spaces (Comaniciu & Meer, 2002). Within a given building cluster, the algorithm initially selects centroid candidates and updates candidates' positions to be the means of points in each iteration:

$$m(x) = \frac{\sum_{i=1}^{n} K\left(x - x_i\right) x_i}{\sum_{i=1}^{n} K\left(x - x_i\right)} \tag{1}$$

$$K(x) = \exp\left(-||x||^2\right) \tag{2}$$

where K is a Gaussian kernel density estimation function of squared distances between points and the cluster centroid. The algorithm ends when the difference between $m(x)$ and x is small ($m(x) \rightarrow x$). The numbers of geometric components are defined based on the size of a building and examined by Fuzzy Clustering. This soft clustering method gives each point a degree of belonging to different clusters instead of assigning concretely to a particular group. If the fuzzy partition coefficient is more than 0.9, the Cluster Detector breaks a building component down into multiple ones (Figure 2).

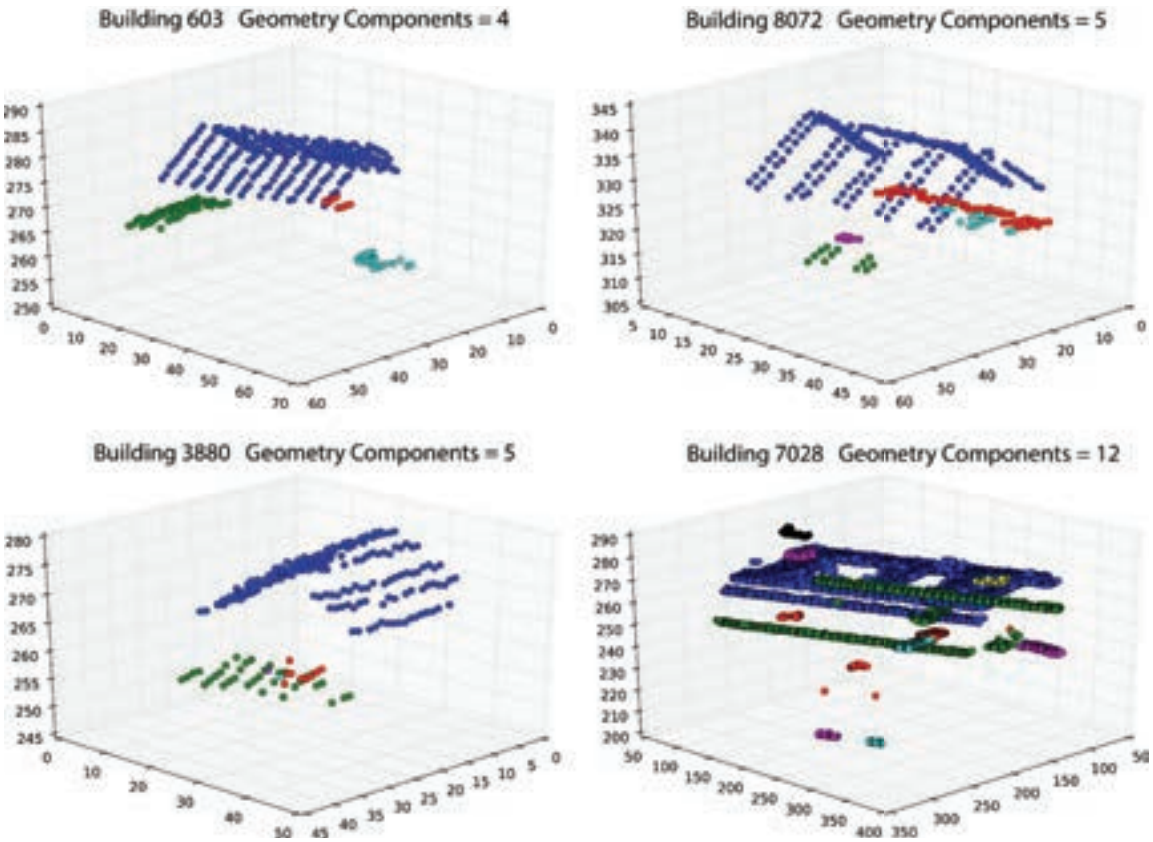
In step 4, the Geometry Detector uses three predefined geometry models to identify geometric types for each building component: one linear model for a flat roof, two linear models for a gable and single sloped roof, and one non-linear model for a gambrel roof (Figures 3 and 4). For each component, the algorithm fits three predefined models and selects a model with the lowest mean squared error (MSE). For the optimization of computing performance, we use the normal equations that perform faster compared to other non-vectorize least square regression techniques. The normal equations apply matrix derivatives for minimizing the model's error to calculate the fitting parameters (Eq. 3 and Eq. 4):

$$\theta = \left(X^T \times X\right)^{-1} \times X^T \times y \tag{3}$$

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left(y - \left(\theta \times X^T\right)^{-1}\right)^2 \tag{4}$$

where X is a matrix $M^{m \times n+1}$ of coordinates (x and y values), y is a m-dimensional vector V^m of elevation information. θ is a n+1-dimensional vector V^{n+1} of fitting parameters, m is the number of points in each geometric component, and n is the number of coordinates (two in our study).

FIGURE 2. Examples of the Cluster Detector results: Mean shift and Fuzzy Clustering algorithms are used for grouping LIDAR points in each building and defining geometry components. Each color represents one geometry component.



In the final step, for each building, the Building-Form Variables Generator calculates building-form variables (Eq. 5) and generates building information variables such as area, number of floors, and height.

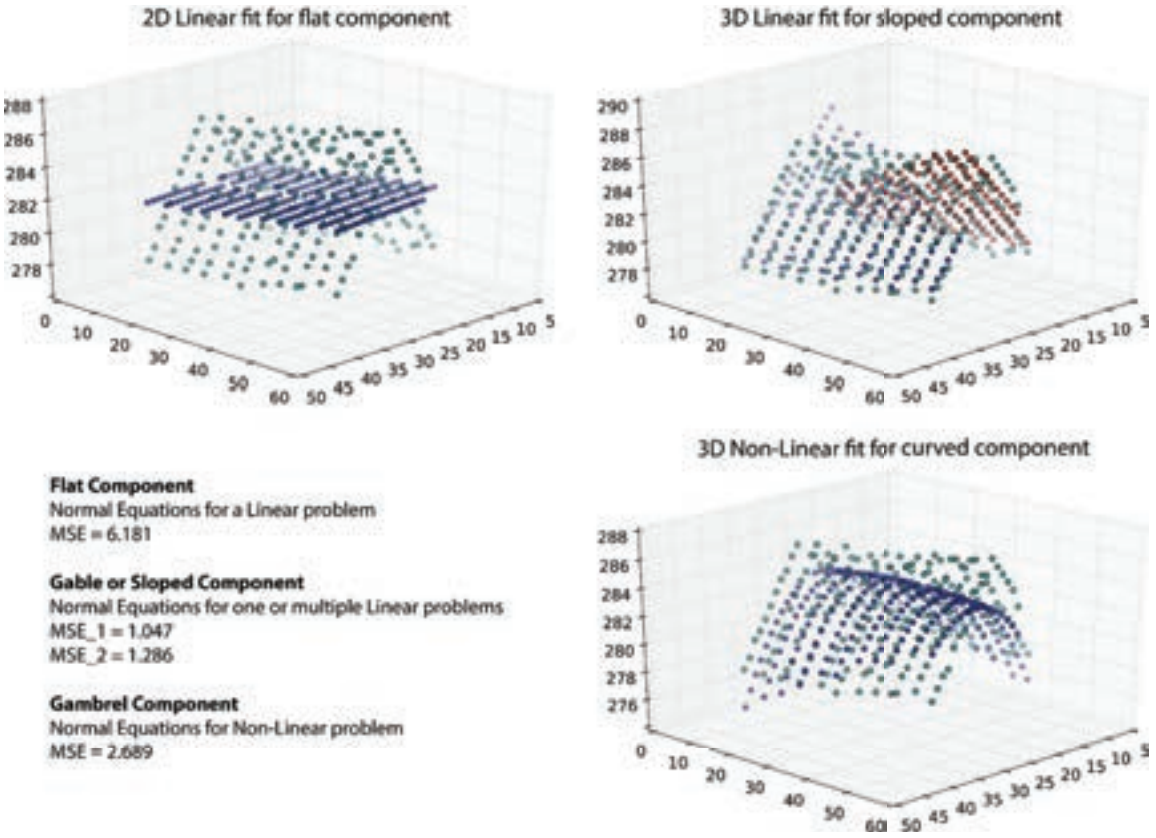
$$C_{pv} = n_{\text{cluster}} + \sum_{i=1}^{n}\left(cl_{\text{type}_i} \times cl_{p_i}\right) + \left(\text{if } cl\_ext_i, 1\right) \tag{5}$$

where n_{cluster} is the number of geometric components in a building, cl_type_i is a categorical value for each geometry type, cl_{p_i} is the portion of building for each type, and cl_ext_i is a binary value for the existence of overhang. $C_{pv}$is a continuous variable indicating the complexity level for a building's geometry. Higher C_{pv} values specify that a building has diverse cluster types, more overhangs, and complex components.

3.3 Spatial Variables Generator

The Spatial Variables Generator converts building-form variables in addition to other physical, institutional, cultural, and environmental parameters into proximity and density variables. These external and internal driving forces are fed into the IUMAT-LUM Simulator for land-use

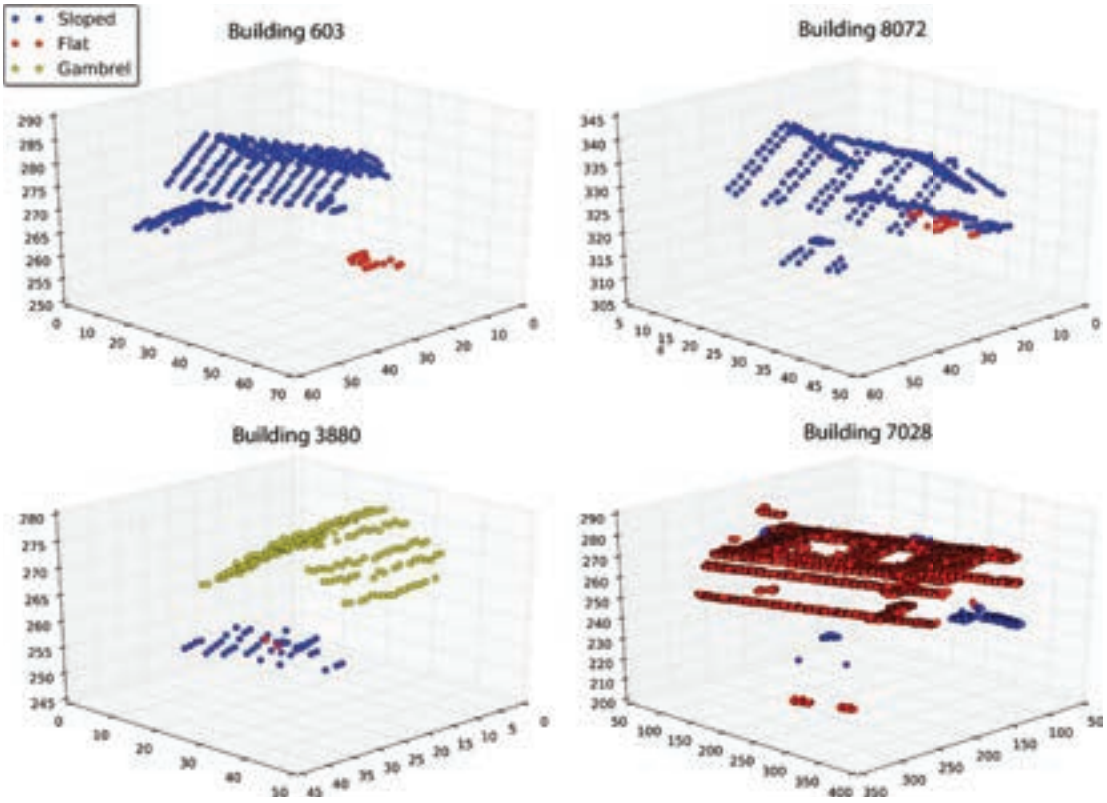
FIGURE 3. An example for Geometry Detector fitting results: three predefined models fitted to a building component and the algorithm selects a model with the lowest MSE value.



changes. The algorithm converts building-form variables, which have coordinate information to GIS vector data and combines it with spatial databases. IUMAT-LUM Spatial Variables Generator employs five steps in coding the GIS vector data to the input data required for training IUMAT-LUM simulator (Figure 1). The Spatial Variables Generator converts all parameters into the same projection. These parameters are then converted into a spatial grid system with a 6x6 meters cell resolution so that the smallest structure is represented in one cell while optimizing the simulation process and memory requirement. The Spatial Variables Generator creates ten descriptive variables (Table 1) that include a land-use type of each cell. Similar to other studies (Almeida et al., 2008), we convert related land-use types into one category, e.g. different residential densities are transformed into one. In doing so, we reclassified 21 land-use classes into 8 groups: residential, commercial, educational, industrial, recreational, urban infrastructure, and non-urban. The Spatial Variables Generator uses the k-dimensional tree algorithm or KD tree (Bentley, 1975) for searching the nearest neighborhood and calculating the proximity variables (Table 2). A KD tree is suitable for avoiding inefficiencies in brute-force computations as the required number of calculations are reduced by encoding the k-dimensional data into new partitioned regions. The algorithm then calculates Euclidean distance between each cell and nearest neighboring cell (Pijanowski et al., 2002) for each parameter (e.g. commercial).

Next, the Spatial Variables Generator produces eleven candidate parameters for density variables (Tables 3 and 4) using Kernel density estimation (Scott, 2015) with a Gaussian

FIGURE 4. Examples of the Geometry Detector outcomes in Amherst: algorithm fits three predefined models to each geometry component and detects roof types.



function. Kernel density estimation is a nonparametric spatial agglomeration for pre-smoothing data, especially with large samples and variables. This smoother density method demonstrates some level of clustering for capturing the distribution and spatial relations (Freisthler, 2013). For each variable in Table 3 and 4, the algorithm places a kernel (Eq. 6) over each cell and calculates the density estimate for the distribution:

$$f(x,y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(d_i, h\right) \tag{6}$$

where n is number of cells, h is the bandwidth, K is kernel function, $f(x, y)$ is density estimate at the location of (x, y), and the distance between cell (x, y) with each neighboring cell. Bandwidth and grid size are two important parameters that affect the outcome of density variables (Anderson, 2009). Using the Gaussian-kde algorithm in a Scipy-Stats library (Python), the algorithm applies Eq. 7 (Scott, 2015) for calculating bandwidth:

$$h = 1.06 \times Std \times n^{-\frac{1}{5}} \tag{7}$$

TABLE 1. Summary of descriptive variables.

	Variable	Description
1	*x*	Latitude converted to feet
2	*y*	Longitude converted to feet
3	*z*	Elevation in feet
4–9	conservation, vegetation, water, recreational, railroad, educational	Binary variables
10	Trans	Binary variable: impermeable surfaces related to transportation network (paved road, parking, driveway)

where the *std* is the sample standard deviation and *n* is the sample size. The algorithm uses a specified distance from a cell's center to estimate the probability density variables. These variables indicate relations of each cell with local actions and global patterns. The algorithm uses small distance (100-meter bandwidth) for exploring the local effects and a wider search range (3000-meter bandwidth) for the overall pattern (Xie & Yan, 2008). For maximizing computing performance, the Spatial Variables Generator normalizes all descriptive, proximity, and density

TABLE 2. Proximity variables (Euclidean distance).

	Variable	Description
11	d_residential	Distance to nearest residential areas (land-use class: multi-family residential, high-density residential, medium density residential, and low-density residential)
12	d_commercial	Distance to nearest commercial areas (land-use class: commercial)
13	d_m_comercial	Distance to main commercial district
14	d_city_center	Distance to the nearest city center
15	d_rec	Distance to recreation spaces (land-use class: participation recreation, spectator recreation, and water-based recreation)
16	d_ind	Distance to industries (land-use class: mining, industrial, and waste disposal)
17	d_edu	Distance to educational facilities (university, college, and school)
18	water	Distance to water bodies
19	d_m_road	Distance to primary roads (transportation networks: paved road, tunnel, and bridge)
20	d_s_road	Distance to roads (transportation networks: unpaved road)
21	d_p_trans	Distance to public transportation

TABLE 3. Summary of density variables (Kernel density estimation).

	Variable	Description
22	agri_kde	Kernel density of agricultural land (land-use class: cropland, pasture)
23	forest_kde	Kernel density of forest (land-use code: forest and non-forested wetland)
24	water_kde	Kernel density of water bodies
25	res_kde	Kernel density of residential districts
26	com_kde	Kernel density of commercial areas
27	rec_kde	Kernel density of recreational regions
28	ind_kde	Kernel density industrial areas
29	edu_kde	Kernel density of educational spaces (university, college, and school)
30	trans_kde	Kernel density of paved surface (transportation networks including driveway and parking lot)
31	walk_kde	Kernel density of sidewalks & bike-path (transportation networks type: bike or walk path, lead walk, detach sidewalks, and attached sidewalks)

variables (ranging from 0.00 to 1.00) by subtracting each variable from the minimum value and dividing the product by the maximum value. In the final step, the algorithm creates a binary transition variable for phase transition from one state to another. It detects land-use changes in different periods and assigns zero to non-change conditions and one to land-use changes. In most cases, we might have an unbalanced database due to the small ratio of land-use changes compared to stable states, which results in skewed model outcomes. In IUMAT-LUM, we use a downsampling method (Omrani et al., 2016; Provost, 2000) to deal with unbalanced databases, which is explained in section 4.1.

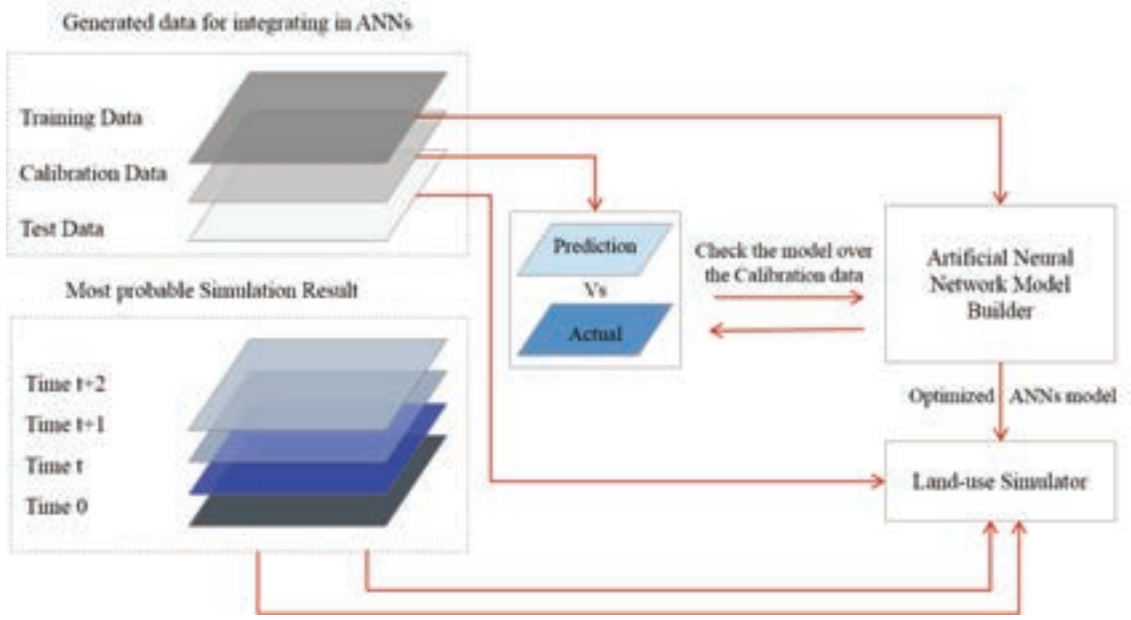
3.4 IUMAT-LUM Simulation Platform

IUMAT-LUM employs an ANNs-based land-use change model, which is a robust machine learning tool for recognizing complex patterns in data (Skapura, 1996). In the IUMAT-LUM framework, we employ the backpropagation algorithm (Rumelhart et al., 1986) for multi-layer network architecture. ANNs models are composed of several layers of nodes called multi-layer perceptrons, an input layer, one or multiple hidden layers, and an output layer. Associated weights control mapping from one node to connected nodes and the activation operation or

TABLE 4. Building-form variables extracted from LIDAR measurements and converted to density variable (Kernel density estimation).

	Variable	Description
32	b_height_kde	Kernel density of building height
33	comp_v_kde	Kernel density of Building-form complexity index

FIGURE 5. Conceptual structure of IUMAT-LUM Simulator.



squashing function is a nonlinear function, which regulates relations between nodes and keeps cell output between certain limits (Graupe, 2013). Mean Squared Error (MSE) is used for training the ANNs model, which is a procedure for updating the weights and bias in the ANNs model. Since ANNs has the tendency to overfit training data (Triantakonstantis & Mountrakis, 2012), in the IUMAT-LUM framework, we divide a given dataset into three subsets of training, validation, and test sets. The ANNs Model Builder runs a learning process over the training data, and checks the overfitting (validates models) by calculating the MSE over a calibration set. In doing so, the algorithm selects an optimal ANNs model that is not overfitted to the training data. And finally, the Land-use Simulator runs the optimized ANNs model over test data for simulating changes over time specified by users (Figure 5).

Multicollinearity refers to the strong correlation between independent variables and is one of the main challenges in machine learning, especially in ANNs algorithms, which cannot automatically exclude relevant parameters (Garg & Tai, 2012). To deal with multicollinearity, instead of using data transformation methods like Principal Component Analysis, in IUMAT-LUM, we evaluate the correlation between variables and keep one predictor from highly correlated variables. Since Pearson's r method only measures linear relationships, we also use Spearman's rank-order correlations to evaluate possible monotonic relationships between parameters.

4. RESULTS AND DISCUSSION

4.1 Study Area and databases

The town of Amherst, Massachusetts, for the period of 1971–2005 is used to test the IUMAT-LUM framework. Located in the Connecticut River Valley, Amherst has three institutes of higher education—the University of Massachusetts, Hampshire College, and Amherst College.

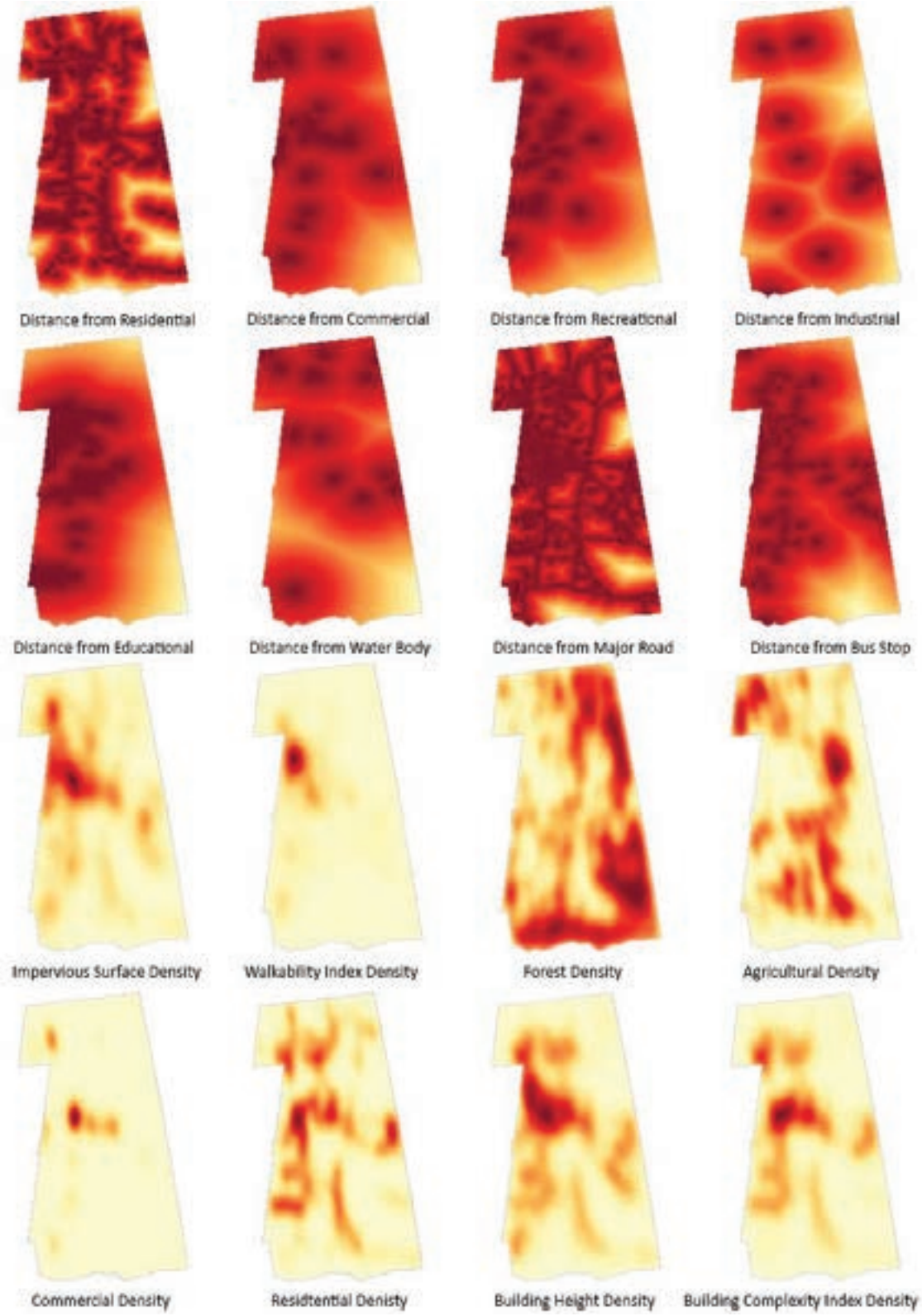
The town has an area of 71.9 km^2 and has experienced a steady population growth, from 26,331 total population in 1970 to 37,819 in 2010. Amherst has diverse land-use classes ranging from relatively high-density commercial to forest and cropland. A random distribution of urban and non-urban lands, as well as a consistent trend of land-use changes since 1971, are valuable resources for generating and testing the land-use model. Additionally, the Amherst Planning Board in collaboration with UMass Campus Planning has comprehensive remote sensing and GIS vector data available. LIDAR measurements provided by the Town of Amherst are used to generate the building-form variables. Zoning, vegetation, hydro system, building boundaries produced by the town, as well as transportation networks, land-use (1971–2005), topography, educational institutes' boundaries provided by MassGIS are compiled in Spatial Variable Generator for producing explanatory variables (Figure 6).

In most urban areas, the number of cells without change is usually more than ones with land-use change. One common approach to deal with imbalanced data is to alter the balance artificially by upsampling or downsampling datasets (Provost, 2000). Dividing the Amherst dataset into three time-intervals (e.g. 1971–1985, 1985–1999, and 1999–2005), we downsampled (Huang et al., 2009) or ignored cells from the majority. In each set, the algorithm assigns a value of one to cells that change in that time interval and zero to those that do not. The algorithm then measures the number of cells with changes and resamples from the no-change cells (Figure 7). In doing so, the adopted 6x6 meters cell resolution in Amherst with a total of 1,933,022 cells is adjusted in different datasets (Table 5). For example, in the 1971–1985 set, 20,132 cells were selected from the land-use map where 10,066 cells (assigned a value of one) belong to the cells that transitioned from non-urban in 1971 to built-up in 1985, while another half (assigned a value of zero) is sampled from non-urban cells in both 1971 and 1985. We also check the balance at different intervals by measuring the transition probabilities of datasets. Rather than using discrete sets, the algorithm combines three datasets and randomly divides it into 60% data as a training set, 20% data for a cross-validation, and 20% data as a testing set (Raj et al., 2010). In this way, the ANNs model is trained based on a continuous historical trend (from 1971 to 2005), not a discrete snapshot and can simulate future patterns with higher accuracy. Calibration and testing datasets are used to check overfitting and to determine if the ANNs model predictions over untrained data are reliable.

4.2 IUMAT-LUM Simulation structure

The IUMAT-LUM simulator uses ANNs classification model (ANNs-Cl) and ANNs regression model (ANNs-Rg) performances in predicting land-use change in Amherst. The number of hidden layers in the ANNs model depends on different factors, including the number of input and output layers, the complexity of the problem, the noise level in the data, the training algorithm, and the regularization (Sarle, 2000). Although in many studies, including Isik et al. (2013), a trial-error method was commonly used for determining the optimum number of hidden layers, Kavzoglu & Mather (2003) suggest 5–10 times the training size as a suitable number. Some studies, such as Omrani (2015) and Tayyebi et al. (2011), used only one hidden layer in predicting urban growth and travel mode respectively, while Wang & Mountrakis (2011) and Kia et al. (2012) developed ANNs with two hidden layers for simulation of urban growth and flood simulation. Like Zhao & Peng (2014), and Maduako et al. (2016), in the IUMAT-LUM Simulator, ANNs-Cl has three hidden layers; each has nodes equal to the numbers of dependent variables, with one rectifier and two sigmoid activation functions. The normalized exponential function is applied to the output layer, so ANNs-Cl generates a zero or one value

FIGURE 6. Thematic maps of normalized proximity and density variables in Amherst from 1971 to 1985; values ranging from zero to one (yellow to red).



for outcomes. With a similar structure, ANNs-Rg uses a sigmoid activation function for the output layer, so the outcomes range between zero and one, which determines the probability of land-use change for cells. A value of one indicates a maximum potential for a future change while a value of zero indicates a low probability. The algorithm initially runs ANNs models over the training datasets for a hundred iterations and updates the weights. After initial training, weights and bias are used in a separate computational loop, which uses training data for updating weights and the calibration set for calculating the MSE. Once it identifies a specified MSE value (over calibration subset), the training process is halted. The optimum ANNs model with the lowest error on the calibration data is checked with the test dataset for overfitting and

FIGURE 7. Maps of downsampled datasets present patterns of land-use change in Amherst from 1971 to 2005. Top: Transition pattern from non-urban to urban types in different time intervals. Bottom: Land-use change patterns within different urban classes.

TABLE 5. Transition probabilities in ANNs datasets for Amherst.

	Number of cells	Global Transition probabilities	Global Transition probabilities Urban Cells	Global Transition probabilities Non-Urban Cells
Training	20,132	0.0451	0.0318	0.0496
Calibration	21,308	0.0497	0.0306	0.0574
Test	12,758	0.1187	0.2252	0.0583

underfitting of the ANNs model. Higher MSE of the test data indicates that the model is overfitted to the training data, while lower MSE of the test data demonstrates the reliability of the model in predicting untrained data.

For isolating the effect of each explanatory variable in the land-use model, we use the inverse version of the "drop one out" approach (Washington et al., 2010; Tayyebi et al., 2011). We run a series of normal equations for measuring variable effects in modeling land-use change within all six databases. For each iteration, the algorithm adds a new variable to matrix X, updates theta (Eq. 3), and measures the MSE value (Eq. 4). Figure 8 shows the trend of MSE values. For the first run, latitude (x), longitude (y), and height (z) are initially used in matrix X. For the second iteration, the binary variable of conservation is added to the matrix X. All independent variables listed in Table 1–Table 4 have positive impacts on model MSE values that vary from a dataset to a dataset (Figure 8). For example, the distance to residential has more impact on land-use transformation of non-urban transition data compared to urban data. In urban databases, residential districts, educational institutes, and forestlands improve the predictability of the model, while in non-urban datasets, agricultural lands, green infrastructure, and transportation networks have significant effects on land-use transformation. In another analysis, the relative effect of building-form variables on land-use modeling is separately explored. After the

TABLE 6. Mean squared error of land-use models with included variables for the statistical analysis. x, y, and z as basic variables are used for the initial run and dependent variables (Table 1–Table 4) are included in the final run.

	Initial Run	Building Height	Building Complexity Index	All dependent variables included
Non-urban Training set	0.2436	0.2378	0.2158	0.1540
Non-urban Calibration set	0.2442	0.2395	0.2191	0.1524
Non-urban Testing set	0.2442	0.2376	0.2142	0.1526
Urban Training set	0.2316	0.2310	0.1884	0.1391
Urban Calibration set	0.2294	0.2289	0.1909	0.1421
Urban Testing set	0.2322	0.2316	0.1885	0.1381

first run with basic variables (x, y, z), building height and building complexity indices are added to the list of variables for the next two iterations (Table 6). In Amherst, these building indices improve prediction of the land-use model by 11% in non-urban and 19% in urban datasets.

4.3 Simulation results

After successful training and calibration, bias and weights of ANNs models are used for forecasting land-use change in the test data. The accuracy value of each model is measured by comparing predicted outcomes applied to the test data versus expected values. Multiple replications of learning, validating, and testing subsets have been conducted in order to avoid the selection of a particular subset by random sampling (called randomness burden). The variation among the simulation results is 5.00 %, and one of the outcomes is presented in Table 7. In non-urban datasets, the MSE over testing data for the simulation is 0.085 in the ANNs regression model and 0.077 in the ANNs classification model with a 0.53 F1 Score, while in urban datasets, ANNs models (regression and classification) have better predictions compared to non-urban

FIGURE 8. Improvement trend of model MSE values in predicting land-use change while adding one independent variable in each iteration.

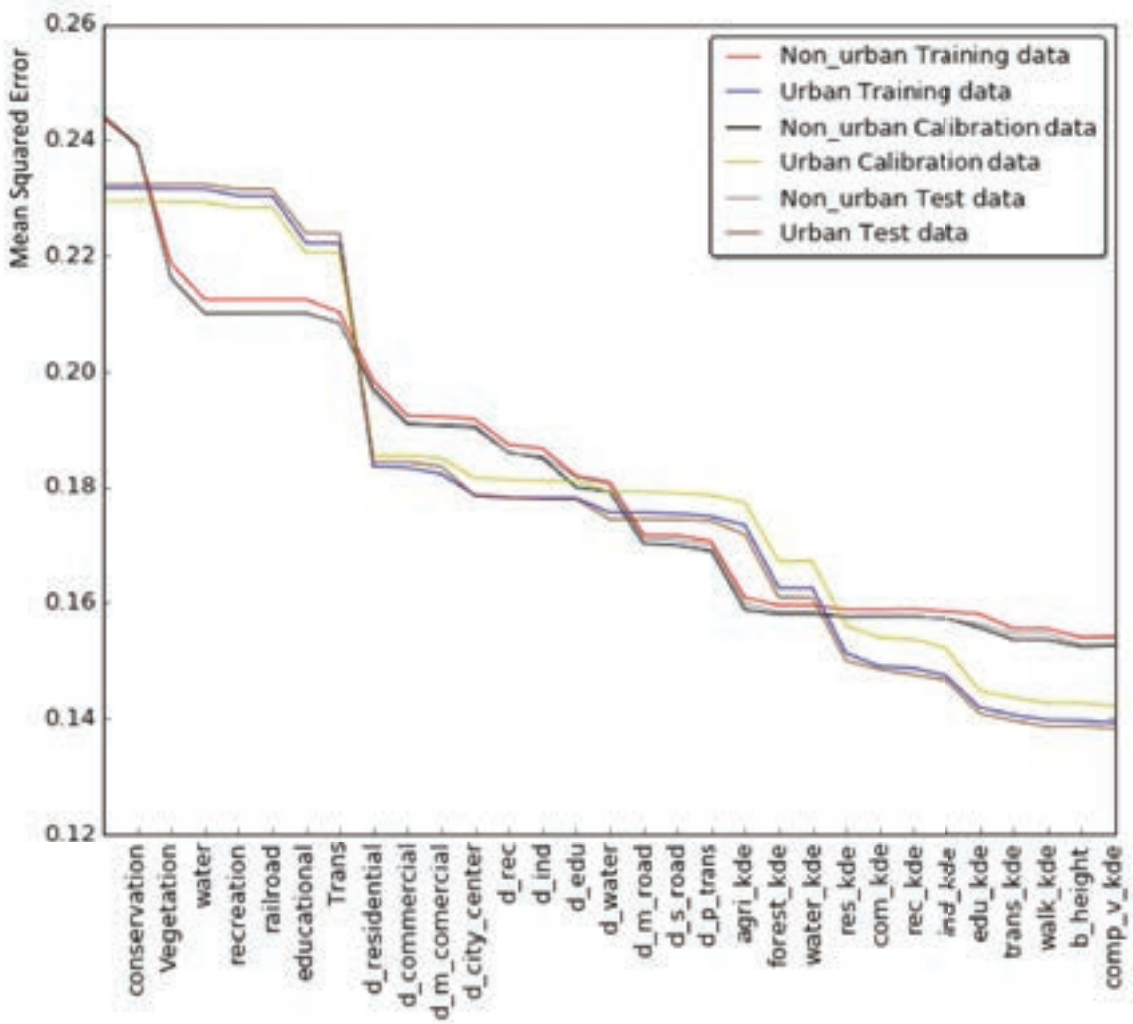


TABLE 7. The goodness of fit in the IUMAT-LUM ANNs-classification models after 300 iterations with regards to different subsets.

	MSE	Precision	Recall	F1 Score
Non-urban Training set	0.052	0.404	0.809	0.539
Non-urban Calibration set	0.074	0.399	0.795	0.532
Non-urban Testing set	0.077	0.400	0.804	0.534
Urban Training set	0.011	0.496	0.987	0.660
Urban Calibration set	0.021	0.481	0.977	0.645
Urban Testing set	0.017	0.490	0.983	0.654

datasets, where MSE is 0.033 in the ANNs regression model and 0.017 in the ANNs classification model with a 0.65 F1 Score (Figure 9). Since the ANNs models have trained over balanced datasets, the performance of the models is not significantly different in predicting the state of non-change and change cells. It is also observed that the Simulator performs better in predicting state changes in urban cells and differences between MSE values of ANN classification and regression models steadily decrease until they reach to the global minimum. However, for non-urban sets, the models arrive at the global optimum more quickly and have higher differences between MSE values of ANNs-Cl and ANNs-Rg models.

For visual comparison, model simulation outcomes were converted into color-coded maps, while in ANNs-Rg models, which generate local transition probabilities ranging from zero to one, the outcomes were transformed into thematic maps for better visualization (Figure 10). The results indicate that the IUMAT-LUM can produce satisfactory predictions about the patterns and scope of changes with slight differences between simulated results and the observed situation. One reason for these discrepancies is the complex spatial interactions and behavioral differences of land-use classes within urban systems, such as green infrastructure or transport networks. Another reason is the interaction between land-use types not being included in IUMAT-LUM. The emergence or evolution of a particular class in a region creates different situations for neighboring cells, which results in changing ultimate land-use pattern (Basse et al., 2014). In addition, socioeconomic characteristics as another deterministic factor have not been integrated into the IUMAT-LUM framework at this stage.

5. CONCLUSIONS AND FUTURE DEVELOPMENT

In this paper, we described how the IUMAT-LUM framework applies Remote Sensing, GIS, and Artificial Neural Networks to simulate urban growth patterns. In IUMAT-LUM, the Building-Form Generator integrates vector GIS routines and LIDAR data to building variables in five steps. We outlined a method for extracting building geometry variables by implementing Density-Based Spatial Clustering, Mean Shift, and Fuzzy clustering algorithms for detecting geometric clusters. We fit three predefined normal equation models (using a model-driven approach) to identify the form of each component in the Geometry Detector. In addition to physical, environmental, cultural, and institutional parameters commonly explored in land-use

FIGURE 9. Mean Squared Error decay curve regarding the IUMAT-LUM ANNs models after 300 iterations. Top: MSE trend of ANNs classification and regression models in non-urban datasets. Bottom: MSE trend of ANNs models in urban datasets.

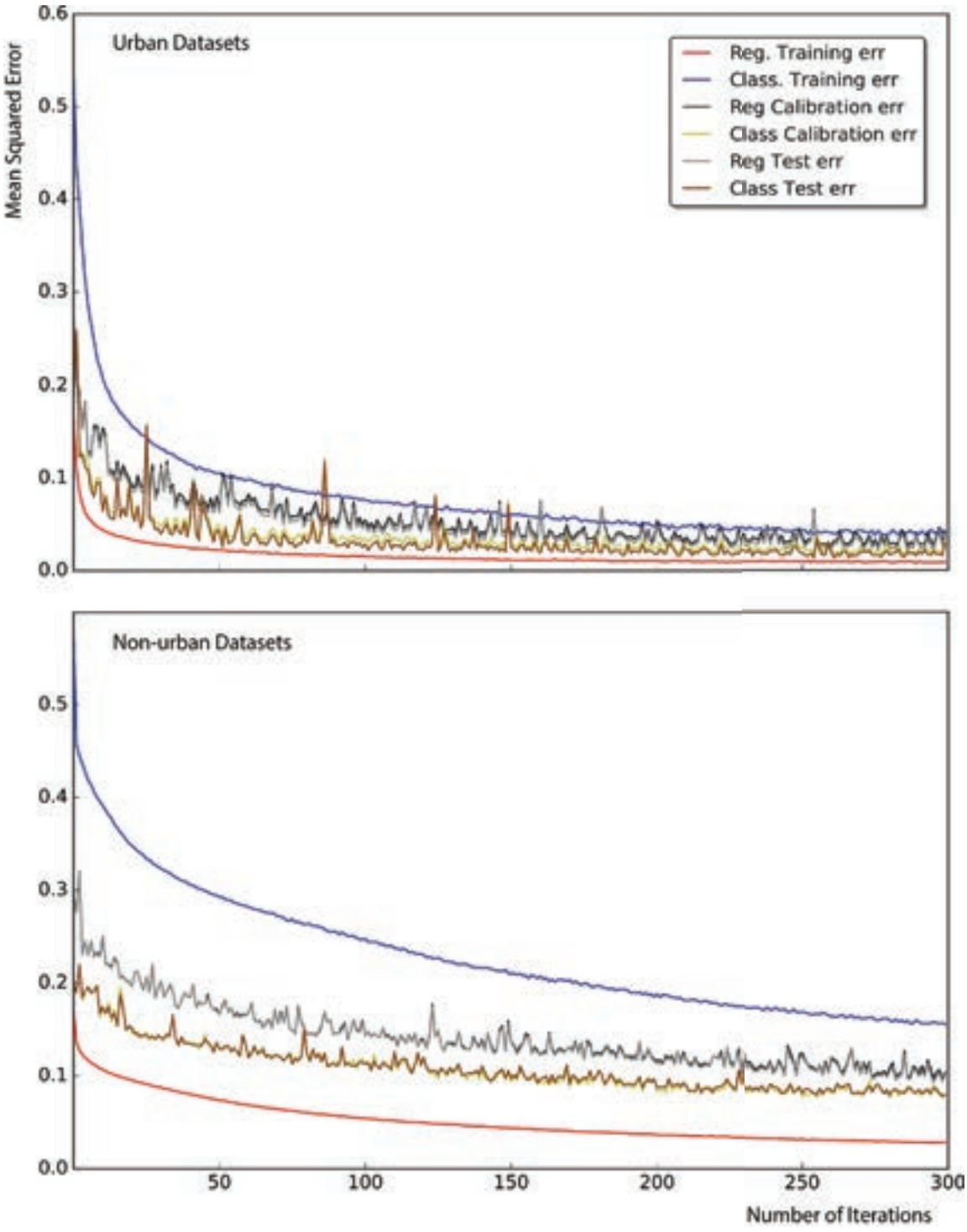
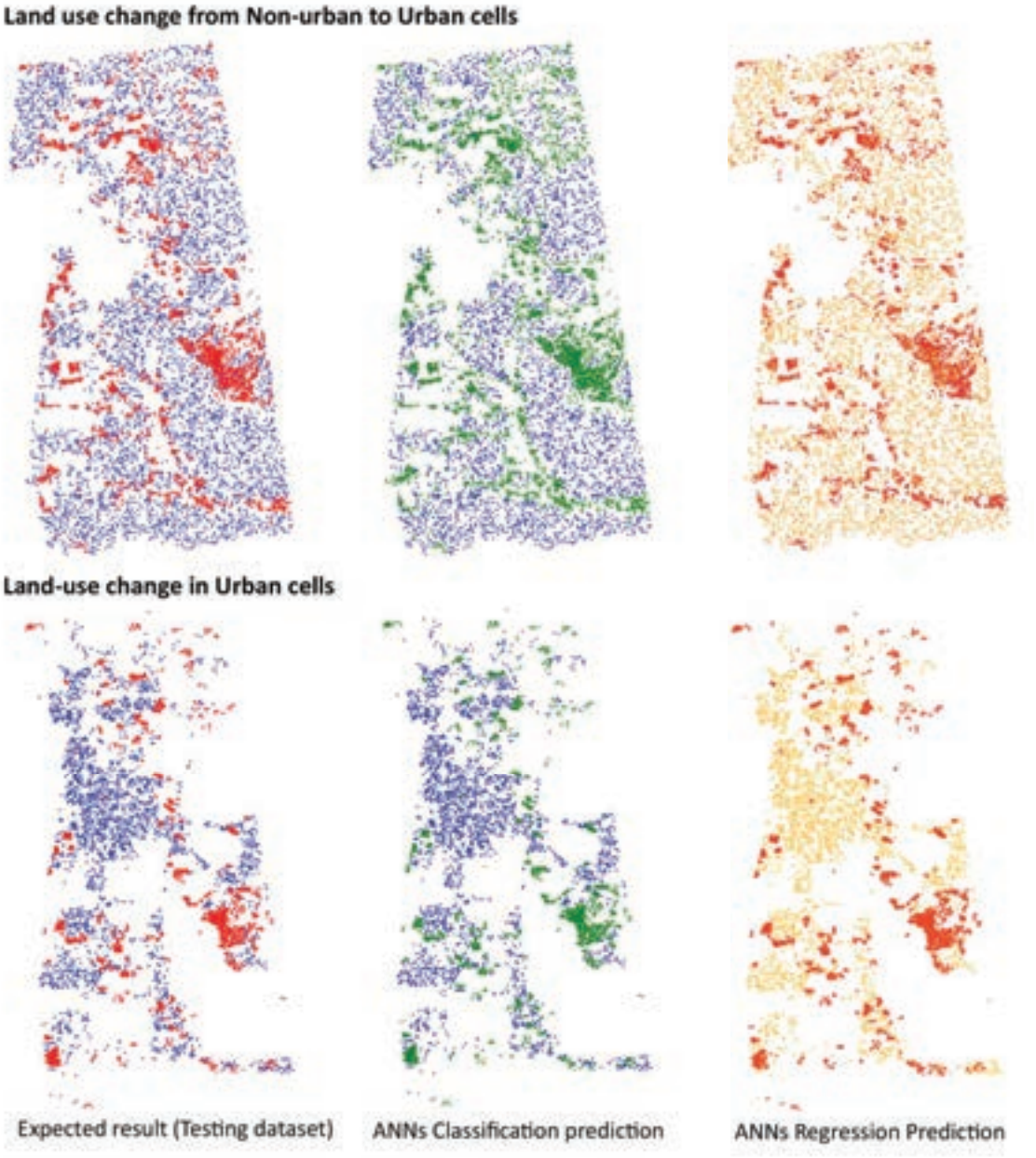


FIGURE 10. Land-use simulation results for the town of Amherst from 1971 to 2005. Left: Map of the testing dataset that shows the expected value in urban and non-urban cells; red presents cells with a transition and blue shows stable cells. Center: Map of ANNs-Cl predicted values; green represents areas with changes and blue shows areas without any changes. Right: Thematic map of ANNs-Rg predicted values ranging from zero to one (yellow to red) that represent land-use change probability.



modeling, we introduced building-form indices as a new determinant factor in simulating land-use change. We applied the IUMAT-LUM framework to the town of Amherst, Massachusetts. Outcomes suggested where land-use transition will be more likely. IUMAT-LUM distinguishes urban regions (residential, commercial, educational, and industrial areas) from non-urban lands (forest, water bodies, agriculture, conservation), and predicts transition probabilities in each group. Our results indicate that building-form indices in combination with other spatial explanatory variables improve the predictive power of land-use modeling. In the town of Amherst, building-form indices improve model predictions by 11% in non-urban and 19% in urban datasets. Regions with higher building geometry indices have higher probabilities of land-use transitions, in other words, more built-up urban areas will have more land-use change compared to less built-up areas. As such, the effects of building-form indices are more noticeable in urban zones than non-urban areas.

The IUMAT-LUM indicates positive effects of the building-form indices on the land-use model performance in the town of Amherst. Since the case study area is relatively small, the proposed model requires further development before it can have any practical applications. In the next stage, IUMAT-LUM will be examined in metropolitan areas (such as Boston and Philadelphia), where the role of the vertical aspect of urbanscape may contribute more to urban growth patterns.

IUMAT-LUM, in its current stage cannot recognize the type of changes in the modeling process. In the future, our focus will be on predicting the type of new development. Impacts of explanatory variables might alter from one type to another land-use type (Basse et al., 2014); for example, a higher walkability index promotes residential developments rather than industrial, or distance to major roads have similar impacts on residential and commercial sectors. Like Carrero et al. (2014) and Tayyebi & Pijanowski (2014), we believe that single ANNs modeling methods cannot solely provide a robust approach for simulating different land-use types. In the next stage of this research, multiple ANNs will be integrated into IUMAT-LUM for modeling different types of land-use change.

Although the predefined geometry models used in this study can recognize most common building-forms, we need to develop a comprehensive archive of predefined models for detecting complex geometries. The Building-Form Generator in the IUMAT-LUM framework provides information about architectural characteristics of buildings such as geometric prototype, footprint, and site coverage. Although IUMAT-LUM does not include all the building-form variables in land-use modeling, these variables in association with other attributes such as building Energy Use Intensity, orientation, and number floors, will be used in other IUMAT models (Mostafavi et al., 2017) for capturing resource consumption. The proposed method may assist planning and design agencies to produce a comprehensive vectorized database of urban and building geometries. This unsupervised method for parameterizing building geometry can also be automated and integrated into other urban metabolism analytical tools similar to the IUMAT framework.

Decision makers and city planners can use the IUMAT-LUM model for determining roles of explanatory parameters on land-use changes and studying future patterns. They can prioritize the planning resources for future scenarios. The IUMAT-LUM approach to predicting future growth patterns within cities' borders is based on historical trends. Comparing simulation results with observed outcomes after implementing a policy could provide new insights into impacts of a particular planning policy. Planners can predict possible developments in environmentally sensitive regions, and regulate non-urban conservation policies accordingly. As an analytical tool

for land-use modeling, it is hoped that IUMAT-LUM can be integrated with urban metabolism analyses for developing sustainable land-use policies that account for the complex spatial relationships of dependent parameters.

REFERENCES

Aisa, B., Mingus, B., & O'Reilly, R. (2008). The Emergent neural modeling system. *Neural Networks*, *21*(8), 1146–1152.

Almeida, C. M., Gleriani, J. M., Castejon, E. F., & Soares-Filho, B. S. (2008). Using neural networks and cellular automata for modelling intra-urban land-use dynamics. *International Journal of Geographical Information Science*, *22*(9), 943–963.

Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis and Prevention*, *41*(3), 359–364.

Anderson, W. P., Kanaroglou, P. S., & Miller, E. J. (1996). Urban form, energy and the environment: a review of issues, evidence and policy. *Urban studies*, *33*(1), 7–35.

Barnsley, M. J., Steel, A. M., & Barr, S. L. (2003). Determining urban land use through an analysis of the spatial composition of buildings identified in LIDAR and multispectral image data. *Remotely sensed cities*, 47–82.

Basse, R. M., Omrani, H., Charif, O., Gerber, P., & Bódis, K. (2014). Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale. *Applied Geography*, *53*, 160–171.

Bechle, M. J., Millet, D. B., & Marshall, J. D. (2011). Effects of income and urban form on urban NO2: Global evidence from satellites. *Environmental Science and Technology*, *45*(11), 4914–4919.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509–517.

Bereitschaft, B., & Debbage, K. (2013). Urban Form, Air Pollution, and CO 2 Emissions in Large U.S. Metropolitan Areas. *The Professional Geographer*, *65*(March 2015), 612–635.

Buxton, M. (2000). Energy, transport and urban form in Australia. *Achieving Sustainable Urban Form*, 54–63.

Carrero, R., Navas, F., Malvarez, G., & Guisado-Pintado, E. (2014). Artificial intelligence-based models to simulate land-use change around an estuary. *Journal of Coastal Research*, (70), 414.

Cervero, R., & Gorham, R. (2009). Commuting in Transit Versus Automobile Neighborhoods. *Journal of the American Planning Association*, *61*(2), 210–225.

Chao, L., & Qing, S. (2011). An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers, Environment and Urban Systems*, *35*(5), 347–357.

Chen, J. (2007). Rapid urbanization in China: A real challenge to soil protection and food security. *Catena*, *69*(1), 1–15.

Cioffi-Revilla, C., & Gotts, N. (2003). Comparative analysis of agent-based social simulations: GeoSim and FEARLUS models. *Journal of Artificial Societies and Social Simulation*, *6*(4).

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

Crossette, B., Kollodge, R., Puchalik, R., & Chalijub, M. (2011). The State of World Population 2011. *United Nations Population Fund*, 1–132.

Ellis, C. (2002). The New Urbanism: Critiques and Rebuttals. *Journal of Urban Design*, *7*(3), 261–291.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. *In Kdd* (Vol. 96, No. 34, pp. 226–231).

Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*(1), 87–114.

Ewing, R., Pendall, R., & Chen, D. (2003). Measuring Sprawl and Its Transportation Impacts. *Transportation Research Record*, *1831*(1), 175–183.

Ewing, R., & Rong, F. (2008). The impact of urban form on US residential energy use. *Housing Policy Debate*, *19*(1), 1–30.

Freisthler, B. (2013). Need for and access to supportive services in the child welfare system. *GeoJournal*, *78*(3), 429–441.

Frey, H. (2003). *Designing the city: towards a more sustainable urban form*. Taylor & Francis.

Garg, A., & Tai, K. (2012). Comparison of regression analysis, artificial neural network and genetic programming in handling the multicollinearity problem. In *Modelling, Identification & Control (ICMIC), 2012 Proceedings of International Conference on* (pp. 353–358). IEEE.

Glaeser, E. L., & Kahn, M. E. (2010). The greenness of cities: Carbon dioxide emissions and urban development. *Journal of Urban Economics*, *67*(3), 404–418.

Gordon, P., & Richardson, H. W. (1997). Are compact cities a desirable planning goal? *Journal of the American Planning Association*, *63*(1), 95–106.

Grammatikopoulos, L., Kalisperakis, I., Petsa, E., & Stentoumis, C. (2015, June). 3D city models completion by fusing lidar and image data. In SPIE Optical Metrology (pp. 95280O-95280O). *International Society for Optics and Photonics*.

Graupe, D. (2013). Principles of artificial neural networks (Vol. 7). World Scientific.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., et al. (2008). Global change and the ecology of cities. *Science*, *319*(5864), 756–760.

Hamidi, S., Ewing, R., Preuss, I., & Dodds, a. (2015). Measuring Sprawl and Its Impacts: An Update. *Journal of Planning Education and Research*, *35*, 35–50.

Handy, S. L., & Clifton, K. J. (2001). Evaluating neighborhood accessibility: Possibilities and practicalities. *Journal of Transportation and Statistics*, *4*(2/3), 67–78.

Heald, C. L., & Spracklen, D. V. (2015). Land use change impacts on air quality and climate. *Chemical reviews*, *115*(10), 4476–4496.

Huang, B., Xie, C., Tay, R., & Wu, B. (2009). Land-use-change modeling using unbalanced support-vector machines. *Environment and Planning B: Planning and Design*, *36*(3), 398–416.

Huang, J., Lu, X. X., & Sellers, J. M. (2007). A global comparative analysis of urban form: Applying spatial metrics and remote sensing. *Landscape and Urban Planning*, *82*(4), 184–197.

International Energy Agency (2008) World Energy Outlook 2008 (Paris: International Energy Agency)

Isidoro, J. M. G. P., & Lima, J. L. M. P. (2014). Laboratory simulation of the influence of building height and storm movement on the rainfall run-off process in impervious areas. *Journal of Flood Risk Management*, *7*(2), 176–181.

Isik, S., Kalin, L., Schoonover, J. E., Srivastava, P., & Lockaby, B. G. (2013). Modeling effects of changing land use/cover on daily streamflow: an artificial neural network and curve number based hybrid approach. *Journal of Hydrology*, *485*, 103–112.

Jantz, C. A., Goetz, S. J., Donato, D., & Claggett, P. (2010). Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Computers, Environment and Urban Systems*, *34*(1), 1–16.

Jenerette, G. D., & Wu, J. (2001). Analysis and simulation of land-use change in the central Arizona–Phoenix region, USA. *Landscape ecology*, *16*(7), 611–626.

Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective 2/e*. Pearson Education India.

Ji, W., Ma, J., Twibell, R. W., & Underhill, K. (2006). Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. *Computers, Environment and Urban Systems*, *30*(6), 861–879.

Kavzoglu, T., & Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, *24*(23), 4907–4938.

Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., Havranek, M., et al. (2009). Greenhouse gas emissions from global cities. *Environmental science & technology*, *43*(19), 7297–7302.

Kia, M. B., Pirasteh, S., Pradhan, B., Mahmud, A. R., Sulaiman, W. N. A., & Moradi, A. (2012). An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia. *Environmental Earth Sciences*, *67*(1), 251–264.

Ko, J. H., Chang, S. I., & Lee, B. C. (2011). Noise impact assessment by utilizing noise map and GIS: A case study in the city of Chungju, Republic of Korea. *Applied Acoustics*, *72*(8), 544–550.

Kwan, M. P., & Lee, J. (2005). Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments. *Computers, Environment and Urban Systems*, *29*(2), 93–113.

Lambin, E. F. (1997). Modelling and monitoring land-cover change processes in tropical regions. *Progress in physical geography*, *21*(3), 375–393.

Lee, J., & Zlatanova, S. (2008). A 3D data model and topological analyses for emergency response in urban areas. *Geospatial information technology for emergency response*, *143*, C168.

Lin, J., Huang, B., Chen, M., & Huang, Z. (2014). Modeling urban vertical growth using cellular automata—Guangzhou as a case study. *Applied Geography*, *53*, 172–186.

Liu, W., & Seto, K. C. (2008). Using the ART-MMAP neural network to model and predict urban growth: A spatiotemporal data mining approach. *Environment and Planning B: Planning and Design*, *35*(2), 296–317.

Loibl, W., & Toetzer, T. (2003). Modeling growth and densification processes in suburban regions—Simulation of landscape transition with spatial agents. *Environmental Modelling and Software*, *18*(6), 553–563. http://doi.org/10.1016/S1364-8152(03)00030-6

Lopez, R., & Hynes, H. P. (2003). Sprawl In The 1990s: Measurement, Distribution, and Trends. *Urban Affairs Review* (Vol. 38).

Luck, M., & Wu, J. G. (2002). A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology*, *17*(4), 327–339.

Maduako, I. D., Yun, Z., & Patrick, B. (2016). Simulation and Prediction of Land Surface Temperature (LST) Dynamics within Ikom City in Nigeria Using Artificial Neural Network (ANN). *Journal of Remote Sensing & GIS*, 2016.

Mage, D., Ozolins, G., Peterson, P., Webster, A., Orthofer, R., Vandeweerd, V., et al. (1996). Urban air pollution in megacities of the world. *Atmospheric Environment*, *30*(5), 681–686.

Maithani, S. (2009). RESEARCH ARTICLE A Neural Network based Urban Growth Model of an Indian City. *Journal of the Indian Society of Remote Sensing*, *2021*(September), 363–376.

Mann, M. E., & Emanuel, K. A. (2006). Atlantic hurricane trends linked to climate change. Eos, *Transactions American Geophysical Union*, *87*(24), 233–241.

Mann, M. E., & Gleick, P. H. (2015). Climate change and California drought in the 21st century. *Proceedings of the National Academy of Sciences*, *112*(13), 3858–3859.

Melton, F., Xiong, J., Wang, W., Milesi, C., Li, S., Quackenbush, A., . . . & Nemani, R. (2016). Potential Impacts of Climate and Land Use Change on Ecosystem Processes in the Great Northern and Appalachian Landscape Conservation Cooperatives. *Climate Change in Wildlands: Pioneering Approaches to Science and Management*, 119.

Mostafavi, N., Farzinmoghadam, M., & Hoque, S. (2017). Urban residential energy consumption modeling in the Integrated Urban Metabolism Analysis Tool (IUMAT). *Building and Environment*, *114*, 429–444.

Mostafavi, N., Farzinmoghadam, M., & Hoque, S. (2014). A framework for integrated urban metabolism analysis tool (IUMAT). *Building and Environment*, *82*, 702–712.

Mostafavi, N., Farzinmoghadam, M., Hoque, S., & Weil, B. (2014). Integrated urban metabolism analysis tool (IUMAT). *Urban Policy and Research*, *32*(1), 53–69.

Mote, P. W., Hamlet, A. F., Clark, M. P., & Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America*. *Bulletin of the American meteorological Society*, *86*(1), 39–49.

Newman, P., & Kenworthy, J. (2006). Urban Design to Reduce Automobile Dependence. *Opolis: An International Journal of Suburban and Metropolitan Studies*, *2*(1), 35–52.

Newton, P. (2000). Urban form and environmental performance. *Achieving Sustainable Urban Form*, 46–53.

Newton, P., Tucker, S., & Ambrose, M. (2000). Housing form, energy use and greenhouse gas emissions. *Achieving Sustainable Urban Form*, 74–83.

Norman, J., Maclean, H. L., Asce, M., & Kennedy, C. a. (2006). Comparing High and Low Residential Density: Life-Cycle Analysis of Energy Use and Greenhouse Gas Emissions. *Journal of Urban Planning and Develpment*, *132*(March), 10–22.

Omrani, H., Judge, V., Antoni, J. P., Klein, O. (2016). Modelling challenge of small amounts of change: a new sampling strategy in land use science. CAMUSS conference, Canada.

Omrani, H. (2015). Predicting Travel Mode of Individuals by Machine Learning. *Transportation Research Procedia*, *10*, 840–849.

Omrani, H., Charif, O., Gerber, P., Bódis, K., & Basse, R. M. (2012). *Simulation of land use changes using cellular automata and artificial neural network* (RPRT). Technical report, CEPS/INSTEAD working paper.

Palme, M., & Ramírez, J. G. (2013). A critical assessment and projection of urban vertical growth in Antofagasta, Chile. *Sustainability*, *5*(7), 2840–2855.

Palmer, T. C., & Shan, J. J. (2001). Urban modeling from LIDAR data in an integrated GIS environment. In *St Louis: ASPRS Annual Conference* (pp. 23–27).

Pielke, R. A., Marland, G., Betts, R. A., Chase, T. N., Eastman, J. L., Niles, J. O., & Running, S. W. (2002). The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1797), 1705–1719.

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. *Environmental Modelling & Software*, *51*, 250–268.

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

Provost, F. (2000, July). Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets* (pp. 1–3).

Raj, K. R., Kardam, A., Arora, J. K., & Srivastava, S. (2010). Artificial Neural Network (ANN) design for Hg–Se interactions and their effect on reduction of Hg uptake by radish plant. *Journal of radioanalytical and nuclear chemistry*, *283*(3), 797–801.

Rumelhart, D. E., McClelland, J. L., & Group, P. D. P. R. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1–2. Cambridge, MA.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., & Bura, S. (1997). SIMPOP: A multiagent system for the study of urbanism. *Environment and Planning B: Planning and Design*, *24*(2), 287–305. http://doi.org/10.1068/b240287

Sarle, W. S. (2000). Neural Network FAQ Part III.

Schwalbe, E., Maas, H. G., & Seidel, F. (2005). 3D building model generation from airborne laser scanner data using 2D GIS data and orthogonal point cloud projections. *Laser Scanning* (ISPRS), (November), 209–214.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Shimrat, M. (1962). Algorithm 112: position of point relative to polygon. *Communications of the ACM*, *5*(8), 434.

Silva, E., & Wu, N. (2012). Surveying models in urban land studies. *Journal of Planning Literature*, *27*(2), 139–152.

Skapura, D. M. (1996). *Building neural networks*. Addison-Wesley Professional.

Spahn, M., & Lenz, B. (2007). Application of a multi-level freight transport demand model to estimate the effects of political measures. *11th World Conference on Transport Research*, 1–10.

Stevens, D., Dragicevic, S., & Rothley, K. (2007). iCity: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling and Software*, *22*(6), 761–773.

Sutton, P. C. (2003). A scale-adjusted measure of "Urban Sprawl" using nighttime satellite imagery. *Remote Sensing of Environment*, *86*(3), 353–369.

Tayyebi, A., & Pijanowski, B. C. (2014). Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *International Journal of Applied Earth Observation and Geoinformation*, *28*, 102–116.

Tayyebi, A., Pijanowski, B. C., & Tayyebi, A. H. (2011). An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran. *Landscape and Urban Planning*, *100*(1–2), 35–44.

Thill, J. C., Dao, T. H. D., & Zhou, Y. (2011). Traveling in the three-dimensional city: applications in route planning, accessibility assessment, location analysis and beyond. *Journal of Transport Geography*, *19*(3), 405–421.

Triantakonstantis, D., & Mountrakis. (2012). Urban growth prediction: a review of computational models and human perceptions. *Journal of Geographic Information System*, *04* (December), 555–587.

Tseng, M.-H., Chen, S.-J., Hwang, G.-H., & Shen, M.-Y. (2008). A genetic algorithm rule-based approach for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(2), 202–212.

UN releases draft agreement on climate change. (n.d.). Retrieved October 5, 2015, from http://www.cbsnews.com/news/un-releases-draft-agreement-to-combat-climate-change/

van Delden, H., Escudero, J. C., Uljee, I., & Engelen, G. (2005). METRONAMICA: A dynamic spatial land use model applied to Vitoria-Gasteiz. In *Virtual Seminar of the MILES Project. Centro de Estudios Ambientales, Vitoria-Gasteiz*.

Vanegas, C. A., Aliaga, D. G., Wonka, P., Müller, P., Waddell, P., & Watson, B. (2010). Modelling the appearance and behaviour of urban spaces. *Computer Graphics Forum*. Blackwell Publishing Ltd.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002). Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, *30*(3), 391–405.

Verburg, P. H., de Nijs, T. C., van Eck, J. R., Visser, H., & de Jong, K. (2004). A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, *28*(6), 667–690.

Waddell, P., Borning, a, Noth, M., Freier, N., Becke, M., & Ulfarsson, G. F. (2003). Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*, *3*, 43–67.

Wagner, P., & Wegener, M. (2007). Urban Land Use, Transport and Environment Models. *disP—The Planning Review*, *43*(170), 45–56

Wang, J., & Mountrakis, G. (2011). Developing a multi-network urbanization model: A case study of urban growth in Denver, Colorado. *International Journal of Geographical Information Science*, *25*(2), 229–253.

Washington-Ottombre, C., Pijanowski, B., Campbell, D., Olson, J., Maitima, J., Musili, A., . . . Mwangi, A. (2010). Using a role-playing game to inform the development of land-use models for the study of a complex socio-ecological system. *Agricultural Systems*, *103*(3), 117–126.

Weidner, T., Knudson, B., Picado, R., & Hunt, J. (2009). Sensitivity testing with the Oregon Statewide Integrated Model. *Transportation Research Record: Journal of the Transportation Research Board* (2133), 109–122.

Williams, E. (2011). Environmental effects of information and communications technologies. *Nature*, *479*(7373), 354–358.

Xie, Z., & Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, *32*(5), 396–406.

Yan, J., Zhang, K., Zhang, C., Chen, S., Member, S., & Narasimhan, G. (2015). Automatic Construction of 3-D Building Model From Airborne LIDAR Data Through 2-D Snake Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(1), 3–14.

You, S., Hu, J., Neumann, U., & Fox, P. (2003). Urban Site Modeling From LiDAR. *Computational Science and Its Applications—ICCSA 2003*, 1–10.

Zhang, K., Yan, J., & Chen, S. C. (2006). Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(9), 2523–2533.

Zhao, L., & Peng, Z. R. (2014). LandSys II: Agent-Based Land Use–Forecast Model with Artificial Neural Networks and Multiagent Model. *Journal of Urban Planning and Development*, *141*(4), 04014045.