

Data Descriptor

Long-Term Land Cover Data for the Lower Peninsula of Michigan, 2010–2050

Amin Tayyebi ^{1,*}, Samuel J. Smidt ² and Bryan C. Pijanowski ³

¹ Monsanto, St. Louis, MO 63146, USA

² Environmental Geoscience, Michigan State University, East Lansing, MI 48824, USA; smidtsam@msu.edu

³ Forestry and Natural Resources Department, Purdue University, West Lafayette, IN 47907, USA; bpijanow@purdue.edu

* Correspondence: amin.tayyebi@gmail.com; Tel.: +1-765-412-1591

Academic Editor: Juanle Wang

Received: 23 November 2016; Accepted: 4 May 2017; Published: 5 May 2017

Abstract: Land cover data are often used to examine the impacts of landscape alterations on the environment from the local to global scale. Although various agencies produce land cover data at various spatial scales, data are still limited at the regional scale over extended timescales. This is a critical data gap since decision-makers often use future and long-term land cover maps to develop effective policies for sustainable environmental systems. As a result, land change science incorporates common data mining tools to create future land cover maps that extend over long timescales. This study applied one of the well-known land cover change models, called Land Transformation Model (LTM), to produce urbanization maps for the Lower Peninsula of Michigan in United States from 2010 to 2050 with five year intervals. Long-term urbanization data in the Lower Peninsula of Michigan can be used in various environmental studies such as assessing the impact of future urbanization on climate change, water quality, food security and biodiversity.

Data Set: available as a supplementary file.

Data Set License: CCBY

Keywords: land cover; data mining; urbanization; Lower Peninsula of Michigan; long-term; land transformation model

1. Introduction

Humans alter the landscape to meet their resource needs [1]. One considerable landscape change in response to human interaction includes urbanization [2]. Environmental planners have raised concerns about the amount of land that urbanization occupies, especially in the United States and other developed nations [3]. While only 2.7% of the world's land is currently occupied by urban development, increased urbanization often leads to the intensification of global resources (e.g., water, food, soil). With global demographic trends projecting 2.5 billion more urban residents by 2050 [4], urbanization will continue with the potential negative impacts on environment and various ecosystem services [5]. Long-term land cover data, especially future urbanization data, play a significant role in better understanding the future impacts of land cover changes on multiple ecosystem services [6] such as biodiversity, climate change, water quality and food security [7–11].

A variety of agencies have produced land cover data from the local to global scales, covering different time periods and different spatial resolutions, such as Moderate Resolution Imaging Spectrometer [12], National Land Cover Data (NLCD) [13], International Geosphere–Biosphere Program [14], and GlobCover [15]. Most of these historical land cover products created from remote

sensing sensors are coarse resolution data and limited in long-term time intervals [16,17]. In terms of spatial resolution, for example, Global Land Cover and Moderate Resolution Imaging Spectrometer are at 1 km and 500 m, respectively. In terms of temporal resolution, NLCD has land cover data for entire United States only for 1992, 2001, 2006, and 2011. Decision-makers often use future land cover maps to develop effective policies for sustainable future environments [18]. To overcome the limitation of coarse resolution land cover data and their long-term time intervals, we incorporated a land cover change model called Land Transformation Model (LTM) to create future high resolution land cover maps (e.g., 30 m) from 2010 to 2050 with five year intervals for the Lower Peninsula of Michigan in United States [19].

Land cover change is a very complex process driven by non-linear factors, including public policy, behavior, economics, and a variety of biophysical and geographic factors, operating at a variety of spatial and temporal scales [20–22]. The complex nature of land cover change relationships requires use of modern data-mining tools to extract underlying patterns in land cover change data [23]. A variety of data mining approaches including empirical [24], dynamic [25], rule-base [26], agent-based [27] and machine learning techniques [1,2,15] have been applied in land change science.

A variety of studies have compared existing land cover change models with each other across the globe. Many of these studies have shown that machine learning techniques, such as artificial neural network (ANN), performed more accurately than other models [23,28]. ANN is inspired by the way biological nervous systems (e.g., brain) process information. ANN is composed of a large number of interconnected processing elements called neurons working together to solve specific problems. Within the ANN, the learning process involves adjustments to the synaptic connections that exist between the neurons.

The Land Transformation Model (LTM) as a data mining approach simulates large-scale land cover change by integrating remote sensing and GIS data within an ANN [28]. The LTM has been used to forecast land cover change in a variety of areas across the world, such as the United States [1,29], Europe [30] eastern Africa [23] and Asia [31]. Primarily, the LTM has been used to (1) determine the uncertainty levels of land change model outputs at a variety of spatial-temporal scales and land change contexts [32]; (2) couple other process-based models to understand how land cover alters climate [33,34] and water [26,35] and ecosystem dynamics; and (3) generate baseline data layers for online decision making tools [36,37]. Here, we applied LTM to produce urbanization maps for the Lower Peninsula of Michigan in United States from 2010 to 2050 with five-year intervals.

2. Data Description

2.1. Study Area

We used the Lower Peninsula of Michigan as the primary boundary (Figure 1), which includes the areas that drain into the adjacent Great Lakes, incorporating parts of the Chicago metropolitan area, northern Indiana, and the Toledo metropolitan area. Nearly 10 million people live in Michigan's Lower Peninsula, and more than 15 million people when including the boundary extent [38]. The largest metropolitan areas include Chicago, Detroit, Grand Rapids, Kalamazoo, Lansing, South Bend, Toledo, and Traverse City. The main commodities produced include corn, soybeans, wheat, hay, cherries, apples, blueberries, potatoes, cucumbers, dry beans, and sugar beets. These eleven commodities account for nearly 98% of the agricultural land across the boundary, and 36% of the entire boundary region; the remaining area is 14% urban, 25% forest, 12% grassland, and 13% open water.

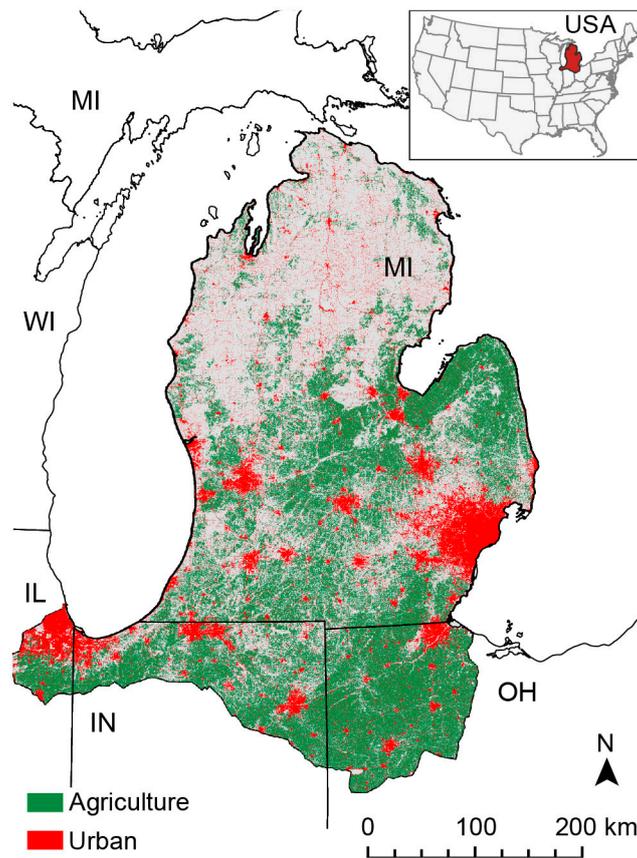


Figure 1. Study area in the Lower Peninsula of Michigan in United States.

2.2. Data

We focused on modeling the conversion of other land cover (e.g., forest, agriculture) to urban (Figure 2B). The output of LTM is a binary land cover change map between two times, which is either coded 1 (for cells converted from other land cover classes to urban class) or coded 0 (for cells remaining in the same land cover class). We used the NLCD in 1992 and 2001 to create the initial binary urbanization maps (Figure 2A). To find cells with a status of 1 and 0, we aggregated NLCDs at each time interval to a binary level covering urban and non-urban classes. By comparing the binary level map in 1992 and 2001, we created a land cover change map between two times, highlighting new urbanization cells in 2001 compared to 1992.

We followed the existing literature in the United States [1,2,23,28] to select the main driving forces that influence urbanization in the Lower Peninsula (Figure 2C), including (1) distance to urban areas; (2) population density; (3) distance to primary highways; (4) distance to secondary highways; (5) distance to rivers; (6) distance to inland lakes; and (7) distance to the coast. Each distance raster was calculated using Euclidean distance in ArcGIS at 30 m \times 30 m resolution. All layers were rasters of 25,000 rows and 25,000 columns in 1992. Within the modelling process, urban areas, roads, parks and water bodies in 1992 were labelled as exclusionary zones, that is, areas where no urban expansion would occur.

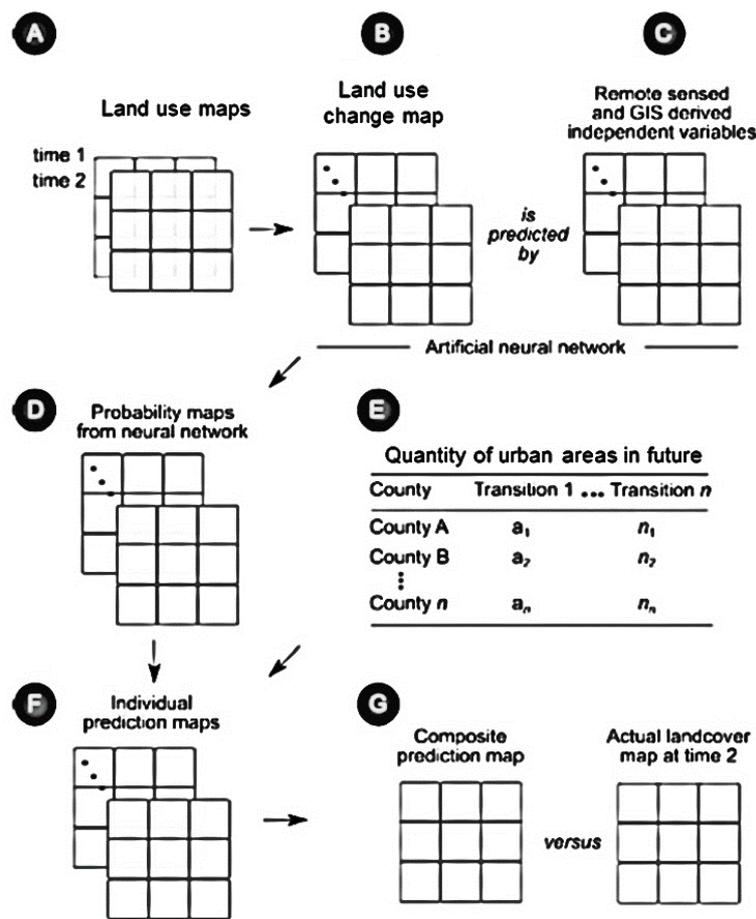


Figure 2. Schematic illustrating the use of the Land Transformation Model.

2.3. Land Transformation Model

LTM uses ANN, which is a machine learning technique, for modelling land cover change [37,39–41]. The multilayer perceptron is one of the well-known ANN forms that is most commonly employed in land cover change science [1]. The multilayer perceptron ANN consists of one input layer, one hidden layer and one output layer. In this study, the number of nodes set for the input and hidden layers was the same as the number of driving forces [36], while we used one node for the output layer. The output node of the ANN was coded with either 1 (for cells converted from other land cover classes to urban classes) or 0 (for cells remaining in same land cover class) between 1992 and 2001. To avoid overfitting, we used a stratified random sampling approach to select 50% of the data for training purposes since the proportion of urban and non-urban areas was not the same. The multilayer perceptron was trained with the most widely utilized algorithm called back propagation (Figure 2B,C) to estimate ANN parameters.

The back-propagation algorithm encompasses two phases. First, the ANN initially assigns random values to ANN parameters (weights and biases), then the multilayer perceptron applies random weights and biases to the input data in order to estimate the outputs. Second, the ANN calculates the mean squared error (difference between the estimated outputs and reference outputs), which is then propagated backward to the previous layers [40]. Optimum model parameters values are estimated by iterating the LTM model through many cycles. A cycle is defined as one complete presentation of all the training data to the ANN [28]. Within the forward and backward process, the error between the reference and estimated outcomes are reduced by updating the ANN parameters (e.g., weights and biases). The difference between mean squared errors of cycles is used as a stopping

condition, where a training run continues until two successive mean squared error differences reach less than 0.05 [23]. The training run stopped after 10,000 cycles.

The estimated parameters are then applied to the independent predictor variables in 1992 (Figure 2C) to calculate the likelihood of urbanization in 2001 (Figure 2D). The result of this step is called a suitability map, where cell values vary from 0 (least likely to convert to urban) to 1 (most likely to convert to urban) and describe the probability of urbanization [26]. While the suitability map determines the potential for urbanization, the LTM also requires a quantity of urbanization to convert the suitability map to a simulated binary map with values of 0, indicating non-urban, or 1, indicating urbanization [36]. These quantities calculated as a result of comparison between the land cover maps in 1992 and 2001 (Figure 2E). The cells in the suitability map are sorted from 1 to 0, and the LTM converts the cells with higher suitability values to urban class until the quantity of urbanization is met (Figure 2F).

2.4. Model Calibration

To evaluate the predictive ability of the LTM, the transformation map in 2001 was combined with NLCD in 1992 to create a composite land cover map that was then compared to the NLCD in 2001 (as a reference map) (Figure 2G). The performance of the LTM for modeling urbanization was evaluated using Percent Correct Match statistics (PCM) and Relative Operating Characteristic curves (ROC) [23,28,42]. PCM describes the proportion of the reference map where urbanization, and no-urban, have been correctly predicted by the LTM [29]. In contrast, ROC is capable of calculating the accuracy across a range of thresholds vary from 0 to 1 [30,40]. For each given threshold (e.g., 0, 0.1, 0.2, . . . , 0.9, 1), the suitability map was converted to a simulated urbanization map. We then compared the simulated urbanization map with NLCD 2001 to calculate false positive (disagreements between urban areas) rates and true positive (agreement between urban areas) rates for each threshold [1,30]. False positive rates and true positive rates are plotted along X and Y axes for each threshold. The area under the ROC curve represents the model accuracy [43]. The calculated PCM (90%) and ROC (85%) indicated that the accuracy of trained model was high enough to be used for future land cover prediction [30].

2.5. Future Land Cover Projection

We then used the calibrated LTM between 1992 and 2001 to forecast urbanization from 2010 to 2050 for each 5-year interval (Figure 3). The LTM parameters linearly estimated from urban transitions between 1992 and 2001 were applied to the predictor variables in 2001 to generate the suitability map of urbanization in 2010. The other ingredient for forecasting urbanization was quantity estimates of urbanized areas from 2010 to 2050. These area estimates were required to convert continuous probabilities to the binary simulated land cover map. These values were estimated using future population projection from U.S. Geological Survey (USGS) to estimate the future quantity of urbanization. With area estimates for urbanization, we allocated future urban areas using the suitability map, in the order of high to low suitability values. Urban transition cells with the highest suitabilities were converted until the total urban quantity was satisfied. This process led to predicted land cover maps that superimposed on NLCD 2001, resulting in a land cover map (Figure 3).

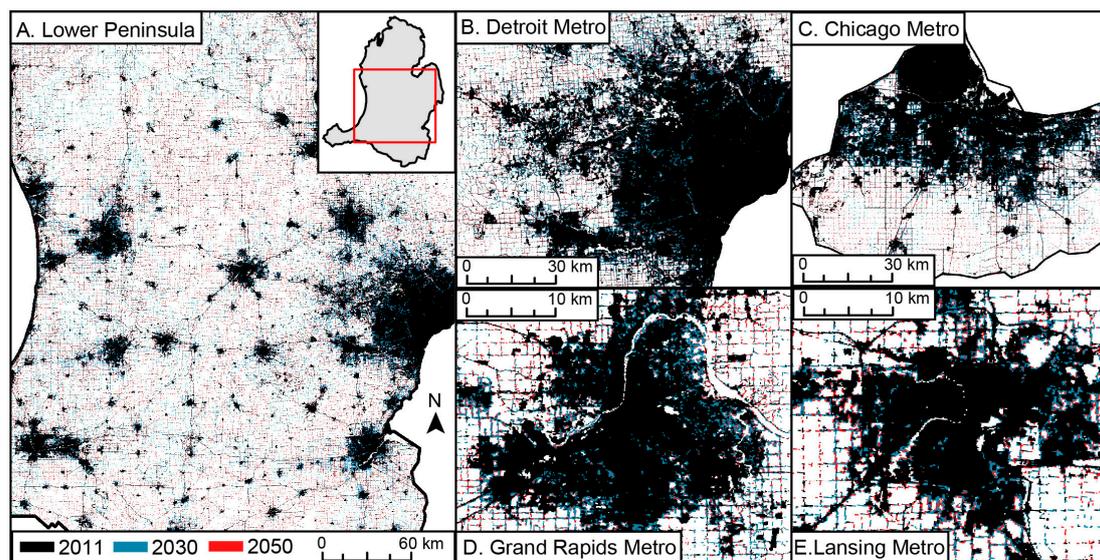


Figure 3. Future urbanized maps for the Lower Peninsula of Michigan in United States from 2010 to 2050 provided by the Land Transformation Model.

3. Conclusions

The calculated PCM (90%) and ROC (85%) indicate that the goodness-of-fit of our trained model was high enough to be used for future land cover prediction. The predicted land cover maps in each time coded either with values of 0, indicating non-urban, or 1, indicating urbanization. The predicted land cover maps are raster files of 25,000 rows and 25,000 columns. Table 1 shows the future urban areas for metropolitan regions in the Lower Peninsula of Michigan. The predicted land cover maps in the Lower Peninsula of Michigan can be used to assess the impact of future urbanization on climate change, water quality, food security and biodiversity into the future.

Table 1. The 2011–2050 simulated areas for major metropolitan regions [$\text{km}^2 \times 10^2$].

Year	Chicago	Detroit	Grand Rapids	Kalamazoo	Lansing	Saginaw	South Bend	Toledo	Traverse City
2011	21.31	39.35	7.84	2.47	3.57	2.14	7.25	5.97	1.12
2015	23.43	42.54	8.59	2.77	3.88	2.27	7.68	6.56	1.20
2020	24.39	44.10	9.06	2.96	4.07	2.34	8.01	6.94	1.43
2025	24.88	45.05	9.39	3.07	4.22	2.43	8.24	7.17	1.50
2030	25.30	45.76	9.66	3.13	4.32	2.52	8.44	7.37	1.52
2035	25.79	46.44	9.96	3.21	4.43	2.60	8.59	7.57	1.58
2040	26.25	47.17	10.15	3.30	4.53	2.67	8.74	7.81	1.72
2045	26.60	47.84	10.36	3.38	4.62	2.72	8.94	7.99	1.84
2050	26.93	48.42	10.58	3.43	4.75	2.79	9.10	8.14	1.94

Author Contributions: All authors contributed to writing the manuscript. Amin Tayyebi conceived the project, ran the LTM, and wrote the initial draft of manuscript. Samuel Smidt prepared spatial and temporal land cover data for modeling. This included data conversion and geographically projecting each land cover as well as processing predictor variables. Bryan C. Pijanowski is the owner of LTM and provided insights during the entire project. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pijanowski, B.C.; Tayyebi, A.; Doucette, J.; Pekin, B.K.; Braun, D.; Plourde, J. 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. *Environ. Model. Softw.* **2014**, *51*, 250–268. [[CrossRef](#)]

2. Tayyebi, A.; Pekin, B.K.; Pijanowski, B.C.; Plourde, J.D.; Doucette, J.S.; Braun, D. Hierarchical modeling of urban growth across the conterminous USA: Developing meso-scale quantity drivers for the Land Transformation Model. *J. Land Use Sci.* **2013**, *8*, 422–442. [[CrossRef](#)]
3. Grimm, N.B.; Faeth, S.H.; Golubiewski, N.E.; Redman, C.L.; Wu, J.; Bai, X.; Briggs, J.M. Global change and the ecology of cities. *Science* **2008**, *319*, 756–760. [[CrossRef](#)] [[PubMed](#)]
4. *Fifth Global Environment Outlook: Environment for the Future We Want*; UNEP: Geneva, Switzerland, 2012.
5. Tayyebi, A.; Meehan, T.D.; Dischler, J.; Radloff, G.; Ferris, M.; Gratton, C. SmartScape™: A web-based decision support system for assessing the tradeoffs among multiple ecosystem services under crop-change scenarios. *Comput. Electron. Agric.* **2016**, *121*, 108–121. [[CrossRef](#)]
6. Meehan, T.D.; Gratton, C.; Diehl, E.; Hunt, N.D.; Mooney, D.F.; Ventura, S.J.; Jackson, R.D. Ecosystem-service tradeoffs associated with switching from annual to perennial energy crops in riparian zones of the US Midwest. *PLoS ONE* **2013**, *8*, e80093. [[CrossRef](#)] [[PubMed](#)]
7. Kalnay, E.; Cai, M. Impact of urbanization and land-use change on climate. *Nature* **2003**, *423*, 528–531. [[CrossRef](#)] [[PubMed](#)]
8. Rudel, T.K.; Coomes, O.T.; Moran, E.; Achard, F.; Angelsen, A.; Xu, J.; Lambin, E. Forest transitions: Towards a global understanding of land use change. *Glob. Environ. Chang.* **2005**, *15*, 23–31. [[CrossRef](#)]
9. Searchinger, T.; Heimlich, R.; Houghton, R.A.; Dong, F.; Elobeid, A.; Fabiosa, J.; Yu, T.H. Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **2008**, *319*, 1238–1240. [[CrossRef](#)] [[PubMed](#)]
10. Lambin, E.F.; Meyfroidt, P. Global land use change, economic globalization, and the looming land scarcity. *Proc. Natl. Acad. Sci. USA* **2001**, *108*, 3465–3472. [[CrossRef](#)] [[PubMed](#)]
11. Baker, A. Land Use and Water Quality. In *Encyclopedia of Hydrological Sciences*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 2005.
12. Schneider, A.; Friedl, M.A.; Potere, D. A new map of global urban extent from MODIS satellite data. *Environ. Res. Lett.* **2009**, *4*, 044003. [[CrossRef](#)]
13. Homer, C.H.; Fry, J.A.; Barnes, C.A. The national land cover database. *US Geol. Surv. Fact Sheet* **2012**, *3020*, 1–4.
14. Belward, A.S.; Estes, J.E.; Kline, K.D. The IGBP-DIS global 1-km land-cover data set DISCover: A project overview. *Photogramm. Eng. Remote Sens.* **1999**, *65*, 1013–1020.
15. Cover, G.L. v2 2008 database. 2008.
16. Arsanjani, J.J.; See, L.; Tayyebi, A. Assessing the suitability of GlobeLand30 for mapping land cover in Germany. *Int. J. Digit. Earth* **2016**, *9*, 1–19. [[CrossRef](#)]
17. Arsanjani, J.J.; Tayyebi, A.; Vaz, E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat Int.* **2016**, *55*, 25–31. [[CrossRef](#)]
18. Ostrom, E. A general framework for analyzing sustainability of social-ecological systems. *Science* **2009**, *325*, 419–422. [[CrossRef](#)] [[PubMed](#)]
19. López, E.; Bocco, G.; Mendoza, M.; Duhau, E. Predicting land-cover and land-use change in the urban fringe: A case in Morelia city, Mexico. *Landsc. Urban Plan.* **2001**, *55*, 271–285. [[CrossRef](#)]
20. Pijanowski, B.C.; Tayyebi, A.; Delavar, M.R.; Yazdanpanah, M.J. Urban expansion simulation using geospatial information system and artificial neural networks. *Int. J. Environ. Res.* **2010**, *3*, 493–502.
21. Lambin, E.F.; Turner, B.L.; Geist, H.J.; Agbola, S.B.; Angelsen, A.; Bruce, J.W.; George, P. The causes of land-use and land-cover change: Moving beyond the myths. *Glob. Environ. Chang.* **2001**, *11*, 261–269. [[CrossRef](#)]
22. Verburg, P.H.; Overmars, K.P. Combining top-down and bottom-up dynamics in land use modeling: Exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landsc. Ecol.* **2009**, *24*, 1167–1181. [[CrossRef](#)]
23. Tayyebi, A.; Pijanowski, B.C.; Linderman, M.; Gratton, C. Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. *Environ. Model. Softw.* **2014**, *59*, 202–221. [[CrossRef](#)]
24. McMillen, D.P. An empirical model of urban fringe land use. *Land Econ.* **1989**, *65*, 138–145. [[CrossRef](#)]
25. Clarke, K.C.; Hoppen, S.; Gaydos, L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environ. Plan. B* **1997**, *24*, 247–261. [[CrossRef](#)]

26. Tayyebi, A.; Pijanowski, B.C.; Pekin, B. Two rule-based urban growth boundary models applied to the Tehran Metropolitan Area, Iran. *Appl. Geogr.* **2011**, *31*, 908–918. [[CrossRef](#)]
27. Parker, D.C.; Manson, S.M.; Janssen, M.A.; Hoffmann, M.J.; Deadman, P. Multi-agent systems for the simulation of land-use and land-cover change: A review. *Ann. Assoc. Am. Geogr.* **2003**, *93*, 314–337. [[CrossRef](#)]
28. Tayyebi, A.; Pijanowski, B.C. Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 102–116. [[CrossRef](#)]
29. Pijanowski, B.C.; Brown, D.G.; Shellito, B.A.; Manik, G.A. Using neural networks and GIS to forecast land use changes: A land transformation model. *Comput. Environ. Urban Syst.* **2002**, *26*, 553–575. [[CrossRef](#)]
30. Pijanowski, B.C.; Alexandridis, K.T.; Mueller, D. Modelling urbanization patterns in two diverse regions of the world. *J. Land Use Sci.* **2006**, *1*, 83–108. [[CrossRef](#)]
31. Tayyebi, A.; Tayyebi, A.H.; Arsanjani, J.J.; Moghadam, H.S.; Omrani, H. FSAUA: A framework for sensitivity analysis and uncertainty assessment in historical and forecasted land use maps. *Environ. Model. Softw.* **2016**, *84*, 70–84. [[CrossRef](#)]
32. Ray, D.K.; Pijanowski, B.C.; Kendall, A.D.; Hyndman, D.W. Coupling land use and groundwater models to map land use legacies: Assessment of model uncertainties relevant to land use planning. *Appl. Geogr.* **2012**, *34*, 356–370. [[CrossRef](#)]
33. Olson, J.M.; Alagaraswamy, G.; Andresen, J.A.; Campbell, D.J.; Davis, A.Y.; Ge, J.; Pijanowski, B.C. Integrating diverse methods to understand climate–land interactions in East Africa. *Geoforum* **2008**, *39*, 898–911. [[CrossRef](#)]
34. Tayyebi, A.; Jenerette, G.D. Increases in the climate change adaption effectiveness and availability of vegetation across a coastal to desert climate gradient in metropolitan Los Angeles, CA, USA. *Sci. Total Environ.* **2016**, *548*, 60–71. [[CrossRef](#)] [[PubMed](#)]
35. Pijanowski, B.; Ray, D.K.; Kendall, A.D.; Duckles, J.M.; Hyndman, D.W. Using backcast land-use change and groundwater travel-time models to generate land-use legacy maps for watershed management. *Ecol. Soc.* **2007**, *12*, 25. [[CrossRef](#)]
36. Azari, M.; Tayyebi, A.; Helbich, M.; Reveshty, M.A. Integrating cellular automata, artificial neural network, and fuzzy set theory to simulate threatened orchards: Application to Maragheh, Iran. *GIScience Remote Sens.* **2016**, *53*, 183–205. [[CrossRef](#)]
37. Shafizadeh-Moghadam, H.; Asghari, A.; Tayyebi, A.; Taleai, M. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Comput. Environ. Urban Syst.* **2017**, *64*, 297–308. [[CrossRef](#)]
38. DeNavas-Walt, C.; Proctor, B.D.; Smith, J.C. *Income, Poverty, and Health Insurance Coverage in the United States: 2009*; US Census Bureau, current population reports, P60-238; U.S. Government Printing Office: Washington, DC, USA, 2010.
39. Song, W.; Pijanowski, B.C.; Tayyebi, A. Urban expansion and its consumption of high-quality farmland in Beijing, China. *Ecol. Indic.* **2015**, *54*, 60–70. [[CrossRef](#)]
40. Shafizadeh-Moghadam, H.; Asghari, A.; Taleai, M.; Helbich, M.; Tayyebi, A. Sensitivity analysis and accuracy assessment of the land transformation model using cellular automata. *GIScience Remote Sens.* **2017**, 1–18. [[CrossRef](#)]
41. Newman, G.; Lee, J.; Berke, P. Using the land transformation model to forecast vacant land. *J. Land Use Sci.* **2016**, *11*, 450–475. [[CrossRef](#)]
42. Tayyebi, A.H.; Delavar, M.R.; Tayyebi, A.; Golobi, M. Combining multi criteria decision making and Dempster Shafer theory for landfill site selection. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 6.
43. Pontius, R.G., Jr.; Batchu, K. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. *Trans. GIS* **2003**, *7*, 467–484. [[CrossRef](#)]

