ISEIS

Journal of Environmental Informatics 25(2) 71-80 (2015)

Journal of Environmental Informatics

www.iseis.org/jei

Predicting Urban Growth of the Greater Toronto Area - Coupling a Markov Cellular Automata with Document Meta-Analysis

E. Vaz[1,*] and J. J. Arsanjani[2]

[1]*Department of Geography and Environmental Studies, Ryerson University, Toronto, M5B 2K3, Canada*
[2]*Institute of Geography, University of Heidelberg, Heidelberg, D-69120, Germany*

Received December 16, 2013; revised July 25, 2014; accepted August 15, 2014; published online March 19, 2015

ABSTRACT. Toronto's Census Metropolitan Area (CMA) has faced on-going challenges concerning its demographic shifts in the urban and rural fringe tending to become a megacity over the coming decades, due to rapid population increase and urban amalgamation. For this research we examine past urban land use transitions in Toronto's CMA based on collected remote sensing data between 1973 and 2010. A Markov Cellular Automata approach is used deriving the CMA urban future based on the existing and planned strategies for Ontario. This is done by a combination of multi-criteria evaluation processes originating transition probabilities that allow a better understanding of the regions urban future by 2030. While the transition probabilities are incorporated from the traditional Markov Chain process, the variables for suitability are measured through a text mining approach, by incorporating several planning documents. The result offers a more integrative vision of policymaker's preference of future planning instruments, allowing for the creation of a better integration of propensity of future growth indicators. The northern part of Toronto is expected to register continuous growth in the coming decades, while agricultural land will continue to decrease. Urban areas after 2020 tend to become more clustered suggesting an importance of planning of green spaces within the Toronto.

Keywords: urban growth, cellular automata, text mining, Greater Toronto Area, Toronto, multi-criteria evaluation

1. Introduction

1.1. Motivation and Literature Review

Urbanization has become a ubiquitous reality throughout the world (Newbold and Scott, 2013). This is often in detriment of agricultural regions, rangelands, as well as environmental zones (Francis et al., 2012). Many urban areas in the developed world are increasing considerably in their extent, given their occurrence as economic centres and resulting demographic shifts (Brueckner, 2000). This is particularly the case of southern Ontario, where rapid urban expansion has resulted from significant economic growth in metropolitan areas (Vaz and Bowman, 2013). This increment has in large, resulted from rapid demographic changes felt across Canada (Moore and Rosenberg, 1995), and growing migration to Canada. According to the United Nations (United Nations, 2012), the pattern of population growth in urban regions is expected to escalate to 72% by 2050. Southern Ontario nests a total of 26% of the population of Canada and supports a population as of 2011 of 8.67 million inhabitants. This region, labelled as the "Golden Horseshoe" due to its shape, has been identified as the fastest growing region in North America (Cadieux et al., 2013). The continued growth in the Golden Horseshoe is further expressed by the Ministry of Infrastructures through the *Places to Grow* Act, prophesizing a projected population increase to 13.48 million by 2041 (Hemson Consulting, 2013). The Greater Toronto Area holds a population of over six million inhabitants and the census metropolitan area (CMA) and a total of 5,583,064 per Canadian Census in 2011. The CMA represents well the urban core of the city of Toronto and its surrounding urban areas, which have progressively grown in the last decade.

The Toronto Census Metropolitan Area (CMA) is on the brink of becoming a megacity as it is only a question of time to have a population larger than ten million inhabitants. Megacities are a part of the many urban landscapes, posing a set of novel challenges both for planning as well as for sustainable growth (Ginkel, 2008). While these urban regions remain fundamental mechanisms of prosperity and innovation for its urban cores (Nijkamp and Kourtit, 2013), they pose great challenges to the environment, resulting in loss of biodiversity and environmental degradation. Large urban areas have accrued consequences on urban sprawl, leading to greater complexity in planning (Johnson, 2001; Clifton et al., 2008). North America has witnessed excessive urban sprawl in the last decade

* Corresponding author. Tel: +1 416 9795000 ext. 3121.
E-mail address: evaz@geography.ryerson.ca (Eric Vaz).

ISSN: 1726-2135 print/1684-8799 online
 doi: 10.3808/jei.201500299

(Schneider and Woodcock, 2008). This also seems to be the case in southern Ontario, where similar to the United States, a fragmented urban landscape exists (Razin and Rosentraub, 2000). Recurrent leapfrogging activity of residential development alongside agricultural zones is an increasing concern throughout the world (Fazal, 2000). The definition by Ewing, Pendall, and Chen (2002) represents some of the key issues of Toronto's CMA urban landscape: a) a dispersed population and a donut shaped reflection of low-density developments, b) disconnected and separate urban hubs alongside the Toronto core and, c) increasing amount of new developments that dwell on the outskirts of the urban fringe, fostering new immigration patterns for young families. It is this urban profile that makes it particularly susceptible to present and future urban sprawl. The Toronto region is facing unmatched challenges for stakeholders and policy makers to deal with the presservation of southern Ontario's Greenbelt, while maintain liveability in a growing region as is the case of Toronto. This issue has been only sparsely addressed for the Greater Toronto Area and deserves further attention (Ali, 2008). As in detriment of rapid urbanisation and agricultural land use, the southern part of the Greenbelt may be increasingly jeopardized. By calculating the urban density of the Greater Toronto Area and co-relating this with the boundaries of Ontario's Greenbelt, it is asserted that part of the Greenbelt has been afflicted by a growing propensity of urban density increase (Figure 1).

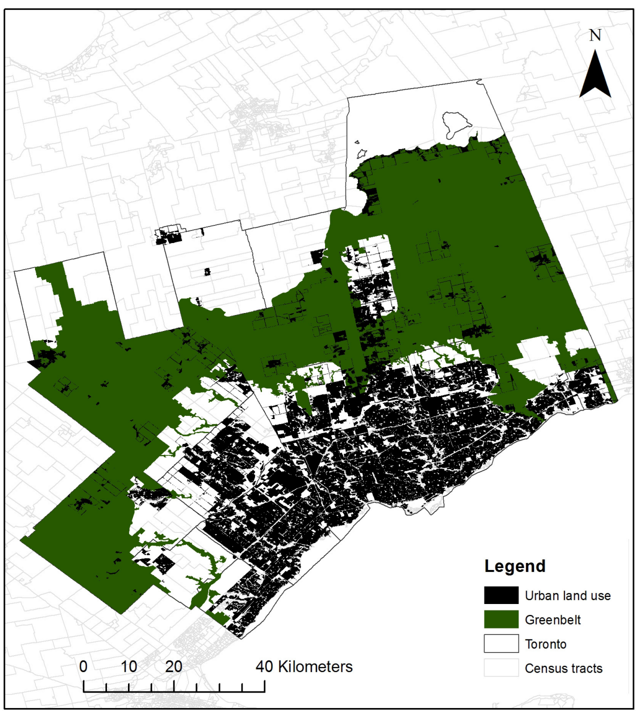

Figure 1. Greenbelt boundaries and urban extension.

Urban planning and land use change in line with an expanding transportation system in the Toronto core, has incremented additional pressure in the downtown Toronto core, felt in the pollution levels felt in the urban metropolis (Nazzal et al., 2013). It is through a combined effort of having information on past and present land use, that planning and monitorring of the Greater Toronto Area can become more efficient. This must be considered with the changing demographic and land-use profiles of Toronto's growing economy (Buliung and Hernandez, 2013). Ancillary strategies of monitoring spatial allocation and land use change must be considered to its fullest extent leading to: a) integration of better land use decisions, b) generate more effective infrastructures coping with the regions hinterland and rural areas, c) diversify commercial activity to local level and local consumption, and d) develop scenarios predicting urban processes and their carrying capacity on land use types and their land use change (Koomen et al., 2007). Spatial information has a vital role in responding to the complexity of urban sprawl. This is especially the case when linked to the many advances of Geographic Information Systems (GIS) and Remote Sensing (RS) (Patino and Duque, 2013). Scenario based modelling approaches of urban and land use change allow thus to establish a more purposeful understanding of future urban change (Vliet et al., 2009). The availability of decades of digital data in large urban as well as mid-resolution satellite imagery greatly contribute to scenario-based analysis of urban dynamics and land use change, in particular in more complex urban environments (Arsanjani et al., 2013a). The complexity in modelling such urban regions is responded thanks to the advances in geocomputational methods, that either explore spatiotemporal characteristics through logistic regression (see Hu and Lo, 2007), Agent Based Models (Arsanjani et al., 2013a), or Multi-Criteria Evaluation (Vaz et al., 2012). Additionally, these geocomputational assessments allow reporting crucial information to legislators and planners, as well as to state and local governmental entities. This combined effort allows to: (i) determine better land use policies and improving transporttation and utility demand, (ii) identify future development pressure points and areas, and (iii) implement effective plans for regional development. This fosters a conciliatory and often efficient effort for sustainable development at the regional level, heightened by the systemic understanding of spatial change at different levels (Clarke, 2003). The policy dimension found in urban growth models as well as the regulatory role of current governmental actions on Ontario's Greenbelt area, suggest a new regional sustainability paradigm where local, regional and national intervention must collaborate (MacDonald and Keil, 2012). This study enables a scenario based approach to understand land use change in the Toronto region through a conciliatory Multi-Criteria Evaluation assessment (Vaz et al., 2012). The following questions are addressed:

(1) Is land use becoming increasingly afflicted by urban land classes in the CMA? What are the main charactерristics of urban growth and present land use change in the Toronto region?

(2) What future land use patterns can be predicted through Markov Transition Chains and cellular Automata, and, does the current documentation help to establish a notion of priorities on urbanisation processes for the Toronto CMA?

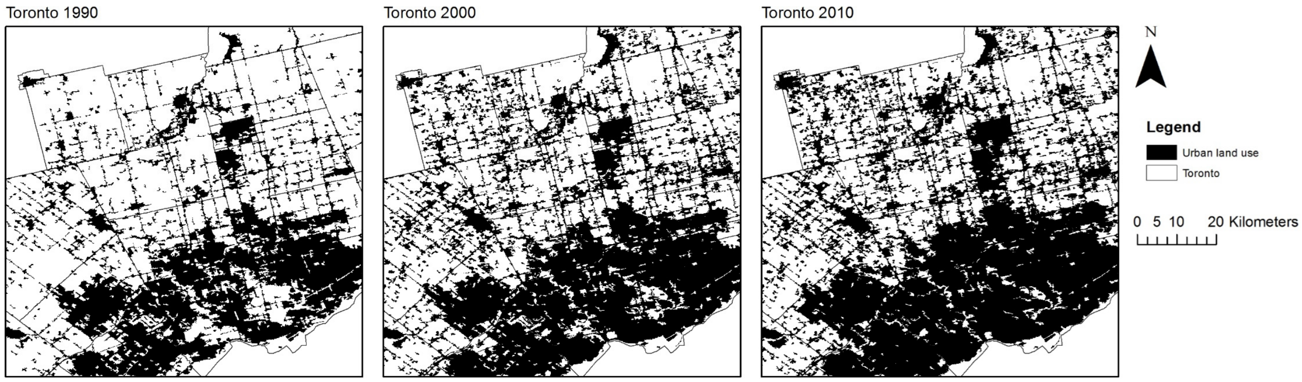

Figure 2. Urban growth of the Greater Toronto Area 1990-2010.

(3) May we expect the Toronto region to merge pre-existent urban land in other areas in the Golden Horseshoe leading to a new megacity in the coming decades?

2. Materials

2.1. Study Area

Toronto's Census Metropolitan Area (CMA) has its core at 43°38′33″ N, 79°23′14″ W. Its main metropolitan region is Toronto, and it harbours three regional municipalities: Durham, Peel and York. For the purpose of our study we have also considered Hamilton, which in recent decades Hamilton has been included as part of the region (Greater Toronto and Hamilton Area). This region has as such become after this addition one of the top fifty largest cities in the world. The CMA extends a total of 7,124.5 km^2, and holds at present (Census Canada, 2011) a total population of 6,054,191, with a population density of 850 inhabitants per km^2. Concentration of population density varies greatly in Ontario. While Toronto has a very high population density, the rest of the province only shows a population density of 14.1 per km^2. The existing infrastructure is marked by two major public transportation systems which cater the Toronto region: The GO Transit and the Toronto Transit Commission. Commutes from the surrounding municipalities to downtown Toronto are frequent, as the metropolitan capital is the current cradle of services and industry. The economic importance of Toronto has shaped the patterns of immigration and location of anthropogenic activity, leading to a rapidly changing urban environment. The fluctuations in the economy and demography have further boosted the rapid urbanisation Toronto's CMA faces (Figure 2). The metropolitan core has thus become, a highly diversified region, with great economic potential for future growth. Toronto has been marked by a successful recovery from the 2009 economic climate felt in North America (Hernandez and Jones, 2005). The investors and stakeholders interest thanks to this rapid recovery, have led in the recent year to a growing interest in investing in the region (Buliung et al., 2007). Presently, the metropolitan area of the Toronto region to be the fourth largest economic centre in North America, and recently classified as one of the top 10 according to the Global Financial Centres Index (GFCI).

2.2. Data

Combinations of different spatial data sets were used to generate the urban growth model (Table 1). Landsat imagery provided an inexpensive option to assess land use for spatial and temporal monitoring (Patino and Duque, 2013). The multi-temporal availability of Landsat imagery makes this imagery particularly adequate for regional assessments, especially given its mid-spatial resolution characteristics (Sexton et al., 2013a), and the availability of data at the Global Land Cover Facility repository (Sexton et al., 2013b). The raw data, was combined with existing classified land use inventories for ground truth, and enabled to create an accurate classification of land use over a larger metropolitan as is the case of Toronto's CMA. Higher resolution satellite imagery was used to create training sites for adequate classification of urban extents and later accuracy assessment, for a complete review on the integration see Vaz and Bowman (2013). For the purpose of this research the following time-stamps were considered: 1992, 1995, 2001, 2010, and 2011. The combined Landsat mosaics allowed to cover the complete metropolitan extent, enabling a classification of land-use types for different decades (1980, 1990, 2000, and 2010). The resulting land-use maps were resampled to 30 m to match a common spatial resolution of all different Landsat imagery (MSS, TM, and ETM+). This allowed for a consistent assessment of land-use of Toronto's CMA. The abundant availability of data in North America, as well as the interest of stakeholders to foment rapid and economically viable decisions on land use planning, make such data sources with longer time frames of spatial information of land use of particular interest. Ancillary data was combined as to create a spatial model of land-use change. The data sets include the variation of population density for 2006 and 2011 (Krakover, 1985) disaggregated at census tract (CT), a digital elevation model (DEM) with 30 metres of spatial resolution per cell, and vector shapefiles of the transporttation network, which were generalized to consider the most important routes (Table 1). The available data sets were projected into the Universal Transverse Mercator coordinate system for the North American Datum 1983, zone 17N, as to allow spatial consistency for the different data sources.

Table 1. Data Sources and Types

Data set	Source	Date
Road network	CanMap RoutLogistics 2010.3	2011
Land use data	CanMap RoutLogistics 2010.3	2010
Landsat imagery	U.S. Geological Survey (USGS)	1992, 1995, 2001, 2010, 2011
Population data	Census Canada	2011
Digital Elevation Model	ASTER (METI/NASA)	2006

Addition of physical geographical data such as elevation and derived slope is pertinent to model urban growth (Clarke and Hoppen, 1997; Rienow and Goetzke, 2014). Regions with low elevation have increased building potential and therefore suggest more accentuated urban growth (Sunde et al., 2014). Slope, derived from the information of elevation also suggests more favourable built-up areas, where increased demand for housing may push pressure to steeper slopes (Syphard et al., 2005).

3. Methodology

In this section, we describe the main components used for the urban growth model of Toronto's CMA. A methodlogical overview is presented in Figure 3, consisting of the workflow that was carried out: a) land use classification of initial multi-temporal satellite imagery, b) creation of suitability maps of land use change relating a multi-criteria evaluation process. And finally, c) integration of a Cellular Automata based algorithm to prospect urban processes for 2020 and 2030.

The methodology assessed the correct land use classifycation for the Greater Toronto Area for the three time stamps used. Markov chains are calculated for the intervals of 1990 to 2000, as to predict a stochastic outcome of land use transition as well as allocation. The model then incorporates additional data sources used in urban growth models in North America land use, excluded areas, distance from urban extent, transporttation network and slope were considered. Past urban extent, the transportation network and slope work as drivers for urbanisation (Silva and Clarke, 2002). These variables were then processed through a Multi-criteria Evaluation (MCE), responding by an Analytical Hierarchical Analysis (AHP) approach that assessed all parameters and generating independent extracted weights. The weights were calculated based on a text mining approach of the existing documents for planning of Toronto's CMA. This allowed filtering the weights of the AHP offering a pairwise assessment in the process of decision making and establishing a priority of importance within the current planning strategy. To allow the text mining integration, a corpus of present literature on planning was used, and examined as a text corpus in R (Feinerer, 2013). This text corpus was then standardized, as to remove the entire fields of numbers, punctuation, and white spaces. A document term matrix was then generated for the text, of which a term analysis was constructed after removing sparse terminologies and allowing an agglomerative cluster of text (Voorhees, 1986). An accounting of search terms followed accordingly, as to allow a perception of words within the Toronto's planning textual corpus.

3.1. Land Use Production and Analysis

Land use and land cover classes were mapped for 1990, 2000 and 2010 based on the available Landsat Thematic Mapper path/rows, and comprising the imagery sets downloaded from the USGS (Path: 18 and 17; Row: 29 and 30). A 30 m ground resolution was adopted for all layers and classified in (1) Urban, (2) Barren, (3) Agriculture, (4) Rangeland, (5) Forest, (6) Water, following the land classification scheme by the USGS (Anderson, 1976) (Figure 4).

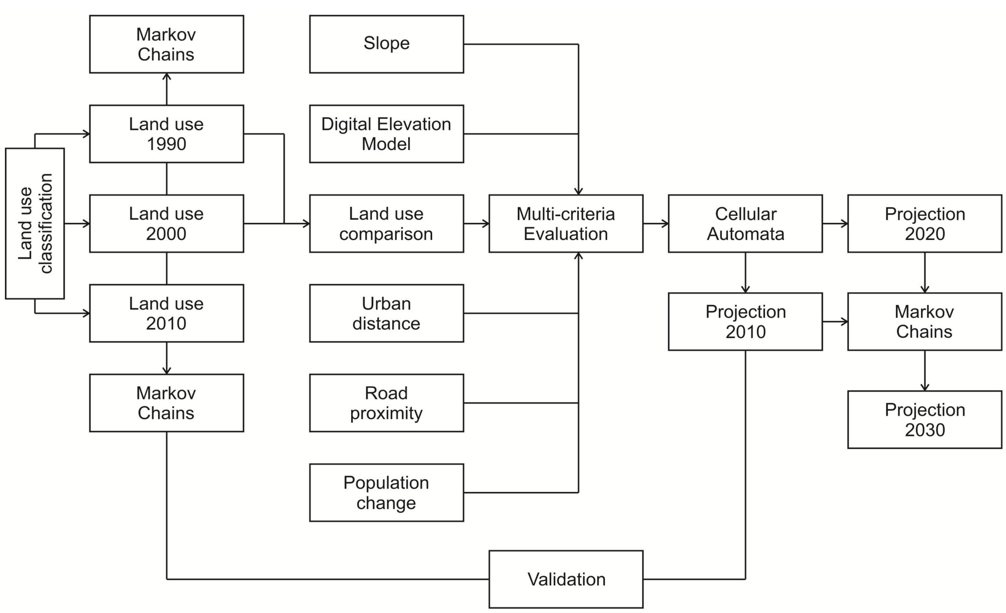

Figure 3. Methodological structure.

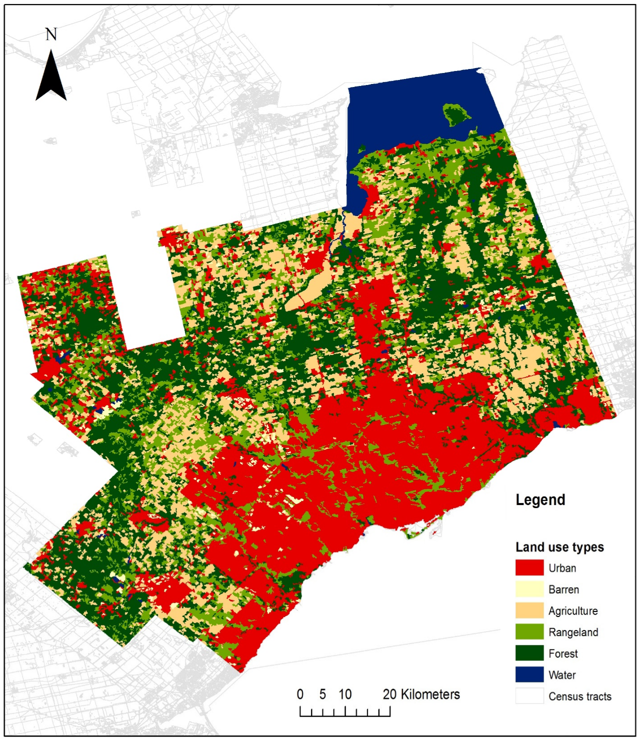

Figure 4. The Land use classification map of 2010.

A supervised classification based on the Maximum Likelihood algorithm was conducted taking advantage from a set of stratified randomly samples resulting in a land use map for each set of satellite images. A Maximum Likelihood algorithm was applied for all imagery. For each individual time stamp, the following steps were taken carefully into account: (i) identification of the most representative training sites for each land use class resulting from an iterative process of selection, (ii) classification of the multispectral images at a pixel level of all land cover classes simultaneously, and finally, the production of a thematic land use map by integrating expert knowledge of the study area with remote sensing images consistently identifying land use and land cover for the region of the Golden Horseshoe in Canada. This was done by selecting 100 training sites for known land use areas based on ground truth data. The overall accuracy calculation led to an overall accuracy of 80% for 2010 and 75% for 2000.

3.2. Markov Transition Chains

The classified land use covers were imported as raster data, which corresponds to a matrix representation of a grid of cells with a given size. The matrix representation of this spatial information can thus, add quantified information per land use class, aggregated in rows and columns as an output of a table format. This is assessed by a Markov Transition Chain which allows measuring the state of each cell individually in n given states. The n states are assembled by a column vector where the component i indicates the probability of a given state in one of the time stamps. A Markov transition chain can thus be expressed as follows:

$$P\left(X_{t_1} \mid X_{t_k}, X_{t_j}, X_{t_i}, \cdots\right) = P\left(X_{t_1} \mid X_{t_k}\right) \Rightarrow t_1 > t_k > t_j, t_i, \cdots \quad (1)$$

A transition probability matrix was built with a stochastic probability of change for examination of the land use classes defined.

3.3. Simulating Future Urban Sprawl

Assembly of cellular automata for predicting urban growth has gained exceptional interest among planners and urban geography scholars (e.g., Arsanjani et al., 2013b). One of the main advantages has been the connection of predicted urbanisation and land use transition processes in the policymaking framework (Pontius et al., 2004). The assembly of urban growth models rely on the technological advances of geocomputation which enable a multi-criteria approach of adding additional complexity on urbanisation processes. Multi-temporal stamps are an asset for determining urbanisation in line with available data inventories, that in developed countries becoming increasingly available. Integrated models bring great advantages when dealing with uncertainty of urban prediction, particularly in regions where spatial change is of utmost importance for sustainable regional and metropolitan planning. A combination of Multi-criteria with Cellular Automata carries interesting results regarding spatial allocation and observance of multi-dimensional consequences of land use change. The quantitative prediction of past temporal stamps of land use, are incorporated with a Cellular Automata (CA) algorithm (White and Engelen, 1997), simulating a composite of urban growth. The thirty metre grid transition rules calculated through the Markov Transition Chains were applied to all cells per iteration factors. The center cell was considered in a 3 × 3 Moore neighbourhood. The determination of a given state of a cell follows as:

$$S_{i,j}^{t+1} = a_n \times N_{i,j}^{t} + M_{i,j} + a_W \times W_{i,j} \quad (2)$$

where $N_{i,j}^{t}$ corresponds to the diffusion factor for its neighbouring cell, $M_{i,j}$ adds the Markov Transitions, $W_{i,j}$ represents the weights of the calculated propensity map. The apparently linear rules that lead a cell to change neighbouring place by automating their probability given a stochastic rule of expansion in $W_{i,j}$, and are then linked to the weighted variables that hierarchically cater different weights creating a landscape of potential transitions for land use change. These transitions are then adjusted with the Markov Transition Chains calculated in the earlier step, allowing for a seemingly and effective interpretation of future urban growth and land use transitions. This integrative approach proves quite useful, as it considers a combinatory approach of spatio-temporal drivers for urban systems (Han et al., 2009). The combination of these data sets is then defined as to create urban change scenarios building on a suitability map of factors and constraints for urbanisation (Pontius and Schneider, 2001). The scenarios themselves are then weighted on by a Multi-Criteria Evaluation process.

Table 2. Markov Chain Probabilities of Transition per Land Use Class

		Probability of changing into				
		Urban	Barren	Agriculture	Rangeland	Forest
Given	Urban	75.04%	2.24%	4.19%	7.39%	10.96%
	Barren	54.66%	22.95%	8.82%	5.66%	2.72%
	Agriculture	19.94%	1.56%	36.14%	24.16%	18.10%
	Rangeland	13.83%	0.59%	23.85%	28.72%	32.91%
	Forest	17.03%	0.14%	5.73%	16.45%	60.26%

4. Results

4.1. Urbanisation in the Greater Toronto Area

A closer inspection on the scattering of land use change (Figure 5) between 2000 and 2010 allows assessing some relevant results from a land use planning perspective. The Toronto downtown core as expected has radically expanded, having in 2000 only 21.71% of urban land use, while in 2010, this value has increased to 30.03%. The pattern of growth justifies further attention, as urban land has increased as a result of urban sprawl (Figure 6), reducing agricultural land from 25.73% in 2000 to 17.53% in 2010. While Barren areas have slightly increased, Forest zones have been noted to increase as well, as a result of agricultural land abandonment to the urban areas, and new creation of forest and park regions within the urban cores (Figure 7).

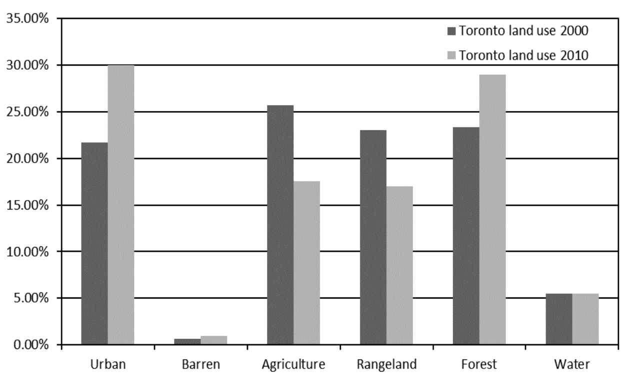

Figure 5. Land use transition between 2000 and 2010.

4.2. Spatio-temporal Change in the Greater Toronto Area

Growth in the Greater Toronto Area has been witnessed since the post-war period, especially given the increase of immigration and population dynamics in Toronto's periphery. This growth in the GTA has been a result of rapid demographic increase, which has led to the diversity of the Greater Toronto Area's landscape. The municipal plans to develop regional infrastructures must take into account land-use change allied to Toronto's urban metabolism (Sahely et al., 2003). At present, the Toronto's CMA is rapidly growing, and as indicated by the Ministry of Infrastructure "Places to Grow" Act, will continue to do so in the coming decades. The vision of the Ministry is to cater the sustainability of urbanisation processes in the Golden Horseshoe and sustain the unique ecological diversity as well as great economic potential of continued growth. While the metropolis will continue to grow, it is important to consider the sustainable strategy of comprising agricultural and rangeland in future as unmanned urban sprawl may jeopardize fragile agricultural regions. The complexity of this urbanisation process is examined through past anthropogenic land-use profiles, which suggests a concerning amount of conversion of agricultural land into urban (19.94%) in the future (Table 2).

The Markov Transition Chain allows a probability interpolation of land use change in the Greater Toronto Area, in particular, looking at the distribution of urban land use several important findings are registered: (i) economic growth and in particular, the growing service sector in the GTA is leading to increased urbanisation , particularly in barren areas, followed by agriculture land, (ii) Forest areas have a tendency to increase in urban areas due to creation of leisure facilities and parks in the urban cores, (iii) Agricultural land has a tendency to be transformed either to Rangeland or Urban areas. Previous analysis carried out for 1990 in comparison to 2000 confirms the tendency of agricultural land. As expected, urban land use is the most dominant land use in the Greater Toronto Area landscape, followed by impacts on Forest, Rangeland and Agriculture. In fact, urban sprawl is predicted to continue in the Greater Toronto Area in the currently rural and agricultural areas of the region's hinterlands. Given this information, the following relative weights were generated through the Analytical Hierarchy Process.

Table 3. Results of Word Matrix

Amend	Area	Avenue	Build
City	Community	Design	Develop
Housing	Land	Map	Nature
Neighbourhood	New	Park	Space
Street	Transit		

4.3. Model Calibration, Evaluation, and Simulation of Future Urban Growth

The underlying factors for urban expansion were exposed by means of the transitional probability, ancillary spatial datasets and available governmental information on future urban planning strategies for southern Ontario. These multiple criteria were integrated in an Analytical Hierarchy Process (AHP) where weights of each criterion were established. This followed a fuzzy standardization process as discussed by and further explored in line with larger urban areas (Feng et al., 2011), giving a set of function derivatives (J-shaped, linear or sigmoid) as well as combination of intervening variables such as land use change promoted by the Markov probabilities of transitions of land use types and weights (Table 3).

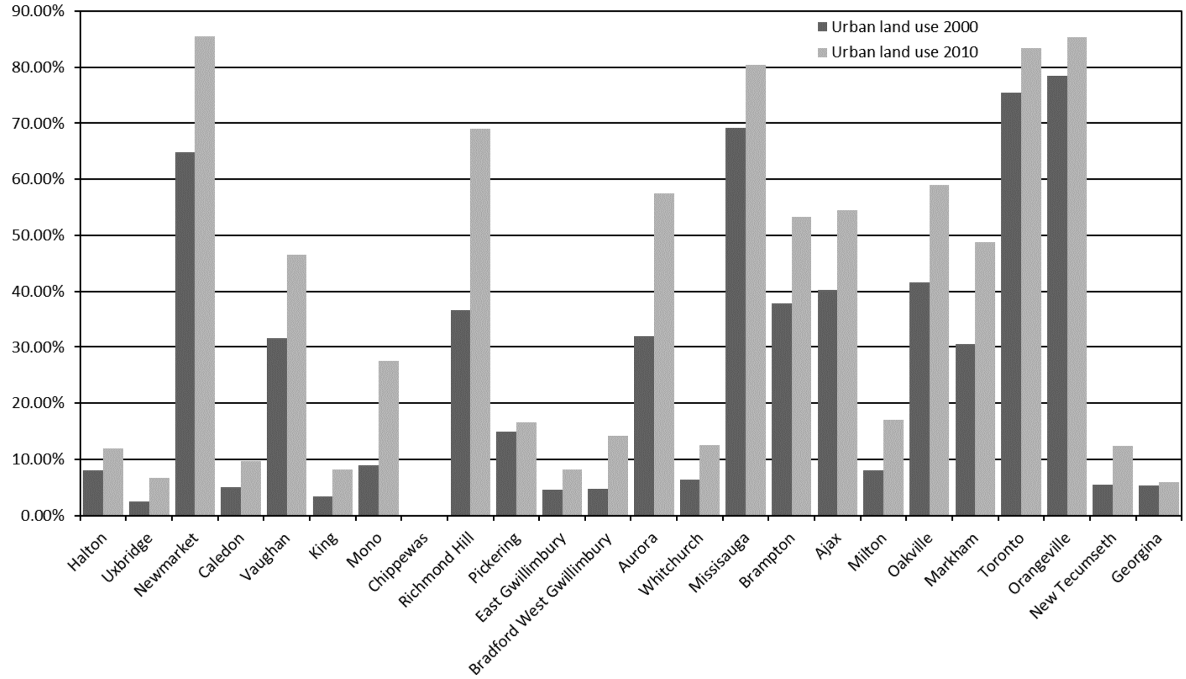

Figure 6. Urban land use change between 2000 and 2010.

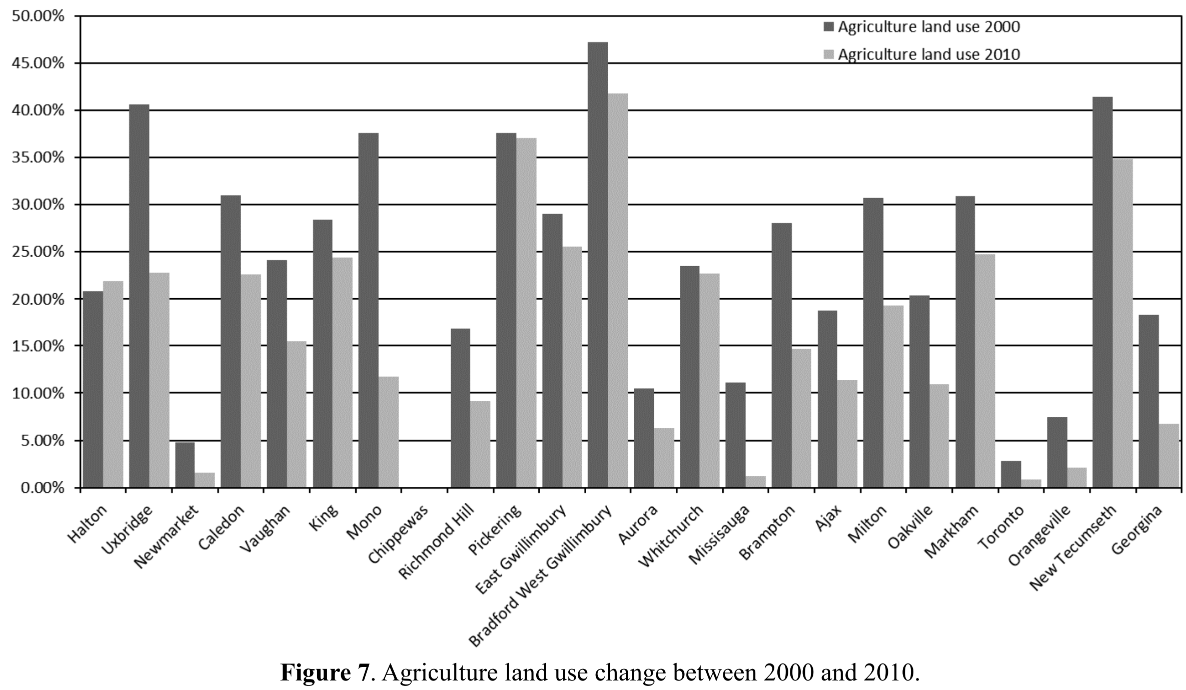

Figure 7. Agriculture land use change between 2000 and 2010.

The variables were chosen through a text mining approach, considering the following documents: "Places to Grow" and "Toronto's Growth Plan", both available through the Ministry of Infrastructure. These documents were parsed, filtered and converted into corpus matrices for textual mining. The textual mining was carried out to find the words that were present more than 75 times in the documents. These words were then organized by typology and geographical weights were thus considered (Table 3).

In this sense, the underlying pattern suggests a higher focus on the existent neighbourhoods, while integrating new urban cores. Improvement is presented by the addition of parks, and transit as well as street optimization, suggests a central role regarding the importance of infrastructure. The expansion of the Greater Toronto Area seems to be community driven.

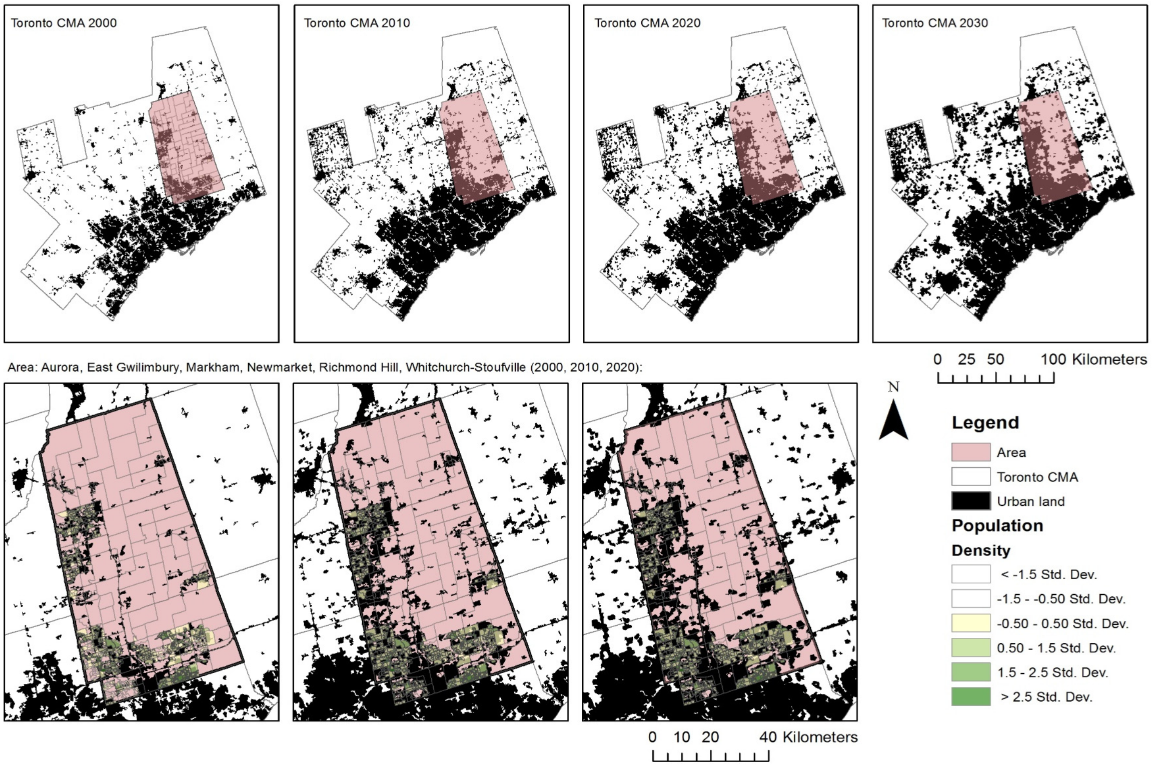

Figure 8. Urban change in the Toronto CMA from 2000 to 2030.

The text mining approach allowed to forward the most adequate factors, and their influence over generating functions and control points, that manifested a more qualitative perception of potential weights. In this sense, and given the conclusions found in the existing description of plans, distance from built-up areas were considered of great importance, followed by distance from roads and the preference of scenic landscapes, following the pertinence of city design. A Multi-criteria Evaluation of weights was thus elaborated based on the preference of the narrative planning structure (Table 4).

Table 4. Distribution of Weights per Factor

Factors	Functions	Control points	Weights
Distance from roads	J-shaped	0-50 m highest suitability 50 m - 1 km decreasing suitability > 1 km no suitability	0.262
Distance from water bodies	Linear	0-50 m no suitability 50 m - 300 m highest suitability >300 decreasing suitability	0.187
Distance from built-up areas	Linear	0-10 km decreasing suitability > 10 km no suitability	0.332
Slope	Sigmoid	0 % highest suitability 0 - 15% decreasing suitability > 15% no suitability	0.091
Land use categories	(Table 2)	n. a.	0.128

As to allow adequate prediction of future land use, evaluation of land use classification was carried out. An integrative approach of predicting land use based on the land use of 1990 and 2000 was conducted. The simulated land use for 2010 was cross compared with the ground truth of land use for 2010. An overall kappa index (Pontius et al., 2004; Arsanjani et al., 2013a) of 82% was achieved. According to Landis and Koch (1977), a kappa index larger than 0.8 refers to almost perfect accordance. The obtained kappa index validated the calibration of the approach in quantifying urban sprawl. By using previously generated transition area matrices of sprawl within 2000-2010, the land use map of 2020 and 2030 are simulated as illustrated in Figure 8.

5. Conclusions

The combination of Cellular Automata with a text mining approach for weight calculation and Markov chains, allowed for the prospection of future urban growth of the Greater Toronto Area for 2020. The figure above (Figure 8) shows urbanisation from 2000 to 2030 (upper part). While urban growth seems to continue to increase in density in the Toronto core, this may be expected to decrease in coming decades. Urbanisation will however increase significantly in the northern area of the Greater Toronto Area, particularly in Aurora, East-Gwilimbury, Markham, Newmarket, Richmond Hill, and Whitchurch-Stoufville (lower part of Figure 8). This is of particular concern given Ontario’s green belt, while still protected, may register accruing change in the forthcoming decades if clear legislation and continued population growth is present. Agricultural land will be lost significantly by 2020, in particular in the perimeters of Toronto's hinterlands, given continued urban pressure in the next years. The connectivity of urban Toronto with other regions in southern Ontario along the Greater Golden Horseshoe (Vaz and Bowman, 2013), suggest that this region must be carefully planned and that the Toronto core plays a vital role in the sustainable development of southern Ontario (Girard, 2007). The province of Ontario is rapidly changing (Muller and Middleton, 1994) and should be analysed through integrated approaches of urban modelling and remote sensing techniques (Banzhaf et al., 2009) thanks to the technical advances in the last decades. Geographic Information Systems play an increasingly important role in sustainable planning (Birkin et al., 1996). Land-use dynamics is a complex phenolmenon which should be addressed through combined methods (Lambin et al., 2001). The usage of non-linear modelling and decision opinions through text mining allow to approach new combinatory assessments of land-use change, in particular in regions where fragmentation seems high and recurrent (Li et al, 2010).

It must be noted that due to rapid economic and social developments of Canadian cities in the last decades, further physical developments in the Canadian cities are theoretically expected, however this investigation confirms this matter in practice as well. Although the importance of rapid urban expansions in Canada is high, a few studies on monitoring the spatiotemporal developments of Canadian cities have been carried out and reported, hence an urgent demand from the urban planners and policy makers have been called for further studies taking these issues into account. More importantly, more investigation on adapting novel modelling techniques integrating GIS algorithms and remote sensing data, which take the domestic circumstances of the urban patters, must be carried out in order to correspond the identified research gap. Because, the essence of land use dynamism in the Canadian cities is unique due to its social, economic, and physical landscape characteristics. In this study, the simple Markov cellular automata model was customized in order to take correspondding socio-economic variables into account to be able to sense the patterns of land use changes.

Acknowledgement. The authors would like to thank the reviewers for their constructive comments. Eric Vaz would like to acknowledge the support of Ryerson University's, Faculty of Arts, through his received New Initiative Award. Jamal Jokar Arsanjani wishes to acknowledge the Alexander von Humboldt foundation for his funding.

References

Ali, A.K. (2008). Greenbelts to contain urban growth in ontario, canada: promises and prospects. *Plann. Practice Res.,* 23(4), 533-548. http://dx.doi.org/10.1080/02697450802522889

Anderson, J.R., Hardy, E.E., Roach, J.T., and Wttmer, R.E. (1976). A land use and land cover classification system for use with remote sensor data. *U.S. Geol, Survey Circ.,* 16, 964.

Banzhaf, E., Grescho, V., and Kindler, A. (2009). Monitoring urban to peri-urban development with integrated remote sensing and GIS information: A Leipzig, Germany case study. *Int. J. Remote Sens.,* 30(7), 1675-1696. http://dx.doi.org/10.1080/0143116080264 2297

Birkin, M., Clarke, G., Clarke, M., and Wilson, A. (1996). Intelligent GIS: Locations decisions and strategic planning, *Geoinformation International,* Cambridge, UK.

Brueckner, J.K. (2000). Urban sprawl: diagnosis and remedies, *Int. Reg. Sci. Rev.,* 23(2), 160-171. http://dx.doi.org/10.1177/0160017 00761012710

Buliung, R., and Hernandez, T. (2013). Retail development in urban Canada: exploring the changing retail landscape of the greater Toronto area (1996-2005). *Int. J. Appl. Geospatial Res.,* 4(1), 32-48. http://dx.doi.org/10.4018/jagr.2013010103

Buliung, R., Hernandez, T., and Mitchell, J. (2007). Exploring the relationship between big-box retail and consumer travel de- mand in the Greater Toronto Area. *Can. J. Transpor.,* 1 (2), 121- 141.

Cadieux, K.V., Taylor, L.E., Buncec, M.F. (2013). Landscape ideology in the Greater Golden Horseshoe Greenbelt Plan: Negotiating material landscapes and abstract ideals in the city's countryside. *J. Rural Stud.,* 32, 307-319. http://dx.doi.org/10.1016/ j.jrurstud.2013.07.005

Clarke, K.C., and Hoppen, S. (1997). A self-modifying cellular automaton model of historical urbanization in the San Franci- sco Bay area. *Environ. Plann. B: Plann. Des.,* 24, 247-261. http:// dx.doi.org/10.1068/b240247

Clarke, K.C. (2003). The Limits of Simplicity: Toward Geocomputational Honesty in Urban Modeling, *7th International Conference on GeoComputation,* University of Southampton, UK.

Clifton, K., Ewing, R., Knaap, G. J., and Song, Y. (2008). Quantitative analysis of urban form: A multidisciplinary review. *J. Urbanism,* 1(1), 17-45. http://dx.doi.org/10.1080/1754917080190 3496

Ewing, R., Pendall, R., and Chen, D. (2002). Measuring sprawl and its impact: the character and consequences of metropolitan expansion, *Smart Growth America*.

Fazal, S. (2000). Urban expansion and loss of agricultural land-a GIS based study of Saharanpur City, India. *Environ. Urban.,* 12(2),133-149.http://dx.doi.org/10.1177/095624780001200211

Feinerer, I. (2014). Introduction to the tm Package Text Mining in R.

Francis, C.A., Hansen, T.E., Fox, A.A., Hesje, P.J., Nelson, H.E., Lawseth, A.E., and English, A. (2012). Farmland conversion to non-agricultural uses in the US and Canada: current impacts and concerns for the future. *Int. J. Agric. Sustain.,* 10(1), 8-24. http:// dx.doi.org/10.1080/14735903.2012.649588

Feng, Y., Liu, Y., Tong, X., Liu, M. and Deng, S., 2011, Modeling dynamic urban growth using cellular automata and particle swarm optimization rules, Landscape and Urban Planning, 102(3): 188-196.

Ginkel, H. (2008). Urban Future. *Nature,* 456, 32-33. http://dx. doi.org/10.1038/twas08.32a

Girard, L.F., Cerreta, M., de Toro, P., Forte, F. (2007). The human sustainable city: Values, approaches and evaluative tools, *in Sustainable Urban Development 2: The Environmental Assessment Methods,* Deakin, M., Mitchell, G., Nijkamp, P., and Vreeker, R. (Eds.) , London, Routledge, pp. 65-93.

Han, J., Hayashi, Y., and Cao, X. (2009). Application of an integrated system dynamics and cellular automata model for urban growth assessment: A case study of Shanghai, China. *Landscape Urban Plann.,* 91(3), 133-141. http://dx.doi.org/10.1016/j.landurbplan. 2008.12.002

Hemson Consulting (2013). Greater Golden Horseshoe Growth Forecasts to 2041-Technical Report. http://www.hemson.com/.

Hernandez, T., and Jones, K. (2005). Downtowns in transition: Emerging business improvement area strategies. *Int. J. Retail Distrib. Manage.,* 33(11), 789-805. http://dx.doi.org/10.1108/095 90550510629392

Hu, Z., and Lo, C. (2007). Modelling urban growth in Atlanta using logistic regression. *Comput. Environ. Urban Syst.,* 31(6), 667-688. http://dx.doi.org/10.1016/j.compenvurbsys.2006.11.001

Johnson, M. (2001). Environmental impacts of urban sprawl: A survey of the literature and proposed research agenda. *Environ. Plann. A,* 33, 717-735. http://dx.doi.org/10.1068/a3327

Jokar Arsanjani, J.J., Helbich, M., and Vaz, E. (2013a). Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. *Cities,* 32, 33-42. http://dx.doi.org/10.1016/j.cities.2013.01.005

Jokar Arsanjani, J., Helbich, M., Kainz, W., Darvishi Boloorani, A. (2013b). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *Int. J. Appl. Earth Obs. Geoinf.,* 21, 265-275. http://dx.doi.org/10.1016/j.jag.2011.12.014

Koomen, E., Stillwell, J., Bakema A., Scholten, H.J. (2007). *Modelling Land-Use Change: Progress and Applications,* Springer, Dordrecht, NL.

Krakover, S. (1985). Spatio-Temporal Structure of Population. Growth in Urban Regions: The Cases of. Tel-Aviv and Haifa, Israel. *Urban Stud.,* 22(4), 317-328. http://dx.doi.org/10.1080/00420988520080551

Lambin, E.F, Turner, B.L., Geist, H.J., Agbola, S.M., Angelsen, A., Bruce, J.W., Coomes, O.T., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skanes, H., Steffen, W.S., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C., and Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environ. Change,* 11, 261-269. http://dx.doi.org/10.1016/S0959-3780(01)00007-3

Landis, L.R., and Koch, G.G. (1977). The measurement of obser- ver agreement for categorical data. *Biometrics,* 33(1), 159-174. http://dx.doi.org/10.2307/2529310

Li, T., Shilling, F., Thorne, J., Li, F., Schott, H., Boynton, R., and Berry, A.M. (2010). Fragmentation of China's landscape by roads and urban areas. *Landscape Ecol.,* 25, 839-853. http://dx.doi.org/10.1007/s10980-010-9461-6

Macdonald, S., and Keil, R. (2012). The Ontario Greenbelt: Shif- ting the Scales of the Sustainability Fix? *Prof. Geogr.,* 64(1), 125-145. http://dx.doi.org/10.1080/00330124.2011.586874

Moore, E.G., Rosenberg, M.W. (1995) Modelling migration flows of immigrant groups in Canada. *Environ. Plann. A,* 27(5), 699-714. http://dx.doi.org/10.1068/a270699

Muller, M.R., and Middleton, J. (1994). A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecol.,* 9(2), 151-157.

Nazzal, Y., Rosen, M., A., and Al-Rawabdeh, A.M. (2013). Assessment of metal pollution in urban road dusts from selected highways of the Greater Toronto Area in Canada. *Environ. Monit. Assess.,* 185(2), 1847-1858. http://dx.doi.org/10.1007/s10661-012-2672-3

Newbold, K.B., and Scott, D. (2013). Migration, commuting distance, and urban sustainability in Ontario's Greater Golden Horseshoe: Implications of the Greenbelt and Places to Grow legislation. *Can. Geogr.,* 57(4), 474-487. http://dx.doi.org/10.1111/j.1541-0064.2013.12044.x

Nijkamp, P., and Kourtit, K. (2013). The "New urban Europe": global challenges and local responses in the urban century. *Eur. Plann. Stud.,* 21(3), 291-315. http://dx.doi.org/10.1080/09654313.2012.716243

Patino, J.E., and Duque, J.C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Compu. Environ. Urban Syst.,* 37, 1-17. http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.003

Pontius, R.G., and Schneider, L.C. (2001). Land-cover change model validation by a ROC method. *Agric. Ecosyst. Environ.,* 85(1-3), 239-248. http://dx.doi.org/10.1016/S0167-8809(01) 00187-6

Pontius, R.G., Jr. Shusas, E., and McEachern, M. (2004). Detecting important categorical land changes while accounting for persistence. *Agric. Ecosyst. Environ.,* 101(2-3), 251-268. http://dx.doi.org/10.1016/j.agee.2003.09.008

Razin, E., and Rosentraub, M. (2000). Are Fragmentation and Sprawl Interlinked? North American Evidence. *Urban Aff. Rev.,* 35(6), 821-836. http://dx.doi.org/10.1177/107808700221 84697

Rienow, A., and Goetzke, R. (2014). Supporting SLEUTH-Enhancing a cellular automaton with support vector machines for urban growth modeling. *Comput. Environ. Urban Syst.,* http://dx.doi.org/10.1016/j.compenvurbsys.2014.05.001

Sahely, H.R., Dudding, S., and Kennedy, C.A. (2003). Estimating the urban metabolism of Canadian cities: Greater Toronto Area case study. *Can. J. Civ. Eng.,* 30(2), 468-483. http://dx.doi.org/10.1139/l02-105

Schneider, A., and Woodcock, C.E. (2008). Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sense data, pattern metrics and census information. *Urban Stud.,* 45, 659-692. http://dx.doi.org/10.1177/0042098007087340

Sexton, J.O., Urban, D.L, Donohue, M.J., and Song, C.H. (2013a). Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sens. Environ.,* 128(21), 246-258. http://dx.doi.org/10.1016/j.rse.20 12.10.010

Sexton, J.O., Song, X.P., Huang, C.Q., Channan, S., Baker, M.E., and Townshend, J.R. (2013b). Urban growth of the Washington, D.C.-Baltimore, MD metropolitan region from 1984 to 2010 by annual estimates of impervious cover. *Remote Sens. Environ.,* 129, 42-53. http://dx.doi.org/10.1016/j.rse.2012.10. 025

Silva, E.A., and Clarke, K.C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Compu. Environ. Urban Syst.,* 26(2), 525-552. http://dx.doi.org/10.1016/S0198-9715(01)00014-X

Statistics Canada, 2011, The Canadian Population in 2011: Population Counts and Growth. Ottawa, ON: Statistics Canada. http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-310-x/98-310-x2011001-eng.pdf

Sunde, M.G., He, H.S., Zhou, B., Hubbart, B., and Spicci, A. (2014). Imperviousness Change Analysis Tool (I-CAT) for simulating pixel-level urban growth. *Landscape Urban Plann.,* 124, 104-108. http://dx.doi.org/10.1016/j.landurbplan.2014.01. 007

Syphard, A.D., Clarke, K.C., and Franklin, J. (2005). Using a cellular automaton model to forecast the effects of urban growth on habitat pattern in southern California. *Ecol. Complex.,* 2(2), 185-203. http://dx.doi.org/10.1016/j.ecocom.2004.11.003

United Nations. (2012). *World urbanization prospects,* The 2011 Revision, New York.

Vaz, E.N., Nijkamp, P., Painho, M., and Caetano, M. (2012). A multi-scenario forecast of urban change: a study on urban growth in the Algarve. *Landscape Urban Plann.,* 104, 201-211. http://dx.doi.org/10.1016/j.landurbplan.2011.10.007

Vaz, E., and Bowman, L. (2013). An application for regional coastal erosion processes in urban areas: a case study of the golden horseshoe in Canada. *Land,* 2(4), 595-608. http://dx.doi.org/10.3390/land2040595

Vliet, J., White, R., and Dragicevic, S. (2009). Modeling urban growth using a variable grid cellular automaton. *Compu. Environ. Urban Syst.,* 33(1), 35-43. http://dx.doi.org/10.1016/j.compenvurbsys. 2008.06.006

Voorhees, E.M. (1986). Implementing agglomerative hierarchic clustering algorithms for use in document retrieval. I*nf. Process. Manage.,* 22(6), 465-476. http://dx.doi.org/10.1016/0306-4573(86)90097-X

White, R., and Engelen, G. (1997). The use of constrained cellular automata for high resolution modelling of urban land-use dynamics. Environ. Plann. A, 24, 323-343. http://dx.doi.org/10.1068/b240323