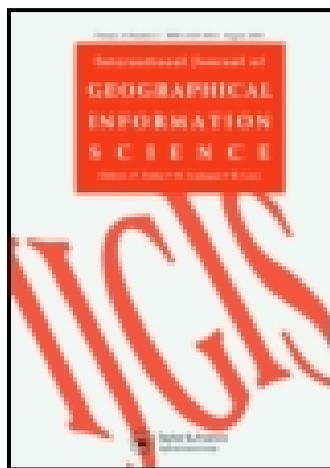


This article was downloaded by: [Gordon Green]

On: 17 July 2015, At: 06:54

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London, SW1P 1WG



International Journal of Geographical Information Science

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tgis20>

Modelling forest canopy trends with on-demand spatial simulation

Gordon M. Green^a & Sean C. Ahearn^a

^a Center for the Advanced Research of Spatial Information, Department of Geography, Hunter College, City University of New York, New York, NY, USA

Published online: 17 Jul 2015.



CrossMark

[Click for updates](#)

To cite this article: Gordon M. Green & Sean C. Ahearn (2015): Modelling forest canopy trends with on-demand spatial simulation, International Journal of Geographical Information Science, DOI: [10.1080/13658816.2015.1066791](https://doi.org/10.1080/13658816.2015.1066791)

To link to this article: <http://dx.doi.org/10.1080/13658816.2015.1066791>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Modelling forest canopy trends with on-demand spatial simulation

Gordon M. Green* and Sean C. Ahearn

*Center for the Advanced Research of Spatial Information, Department of Geography,
Hunter College, City University of New York, New York, NY, USA*

(Received 3 November 2014; accepted 11 June 2015)

Understanding trends in forest canopy cover at local, national, and global scales is important for many applications, including policymaking related to forest carbon sequestration. Globally consistent land-cover data sets derived from MODerate-resolution Imaging Spectroradiometer (MODIS) are now available for a period of more than 10 years, long enough to detect trends both in deforestation and in afforestation. However, methods of modelling land-cover change normally require specialized software and expertise, limiting the availability of this information. This barrier to access can be eliminated through the use of web services that construct models on demand based on user-specified regions of interest, so that parameters are inferred from, and relevant to, local conditions. In this paper we present a proof-of-concept system for building and running spatial Markov chain models of forest-cover change on demand, and demonstrate how the on-demand approach may be implemented for similar applications.

Keywords: web services; spatial Markov; land cover; MODIS

Introduction

The contribution of land-cover change to current anthropogenic climate forcing has been estimated at 17% (EPA 2013), and afforestation and deforestation are key components of this change. The satellite-based record of land-cover products derived from the MODerate-resolution Imaging Spectroradiometer (MODIS) is now available for a period of over a decade, long enough to begin to capture and model changes in forest cover. Rates of land-cover change, however, are highly influenced by policy decisions implemented within areas that may or may not correspond to well-defined administrative or grid-cell boundaries. Currently, models of land-cover change are usually constructed for specific regions of interest, requiring model developers to download and process data for each particular location, even for relatively simple modelling methods. Constructing models on demand presents a more streamlined alternative: land-cover trends can be derived for any given area on the fly, using data that has been processed for this purpose. Only a region of interest, selected interactively, is required. This makes it possible to quickly examine land-cover trends for arbitrary regions of interest. The purpose of this paper is to demonstrate the feasibility of one such approach by describing a proof-of-concept implementation, which is available at <http://www.carsilab.org/forest>.

*Corresponding author. Email: ggreen1@gradcenter.cuny.edu

Modelling land-cover change

Changes in land cover over time have long been modelled using Markov chains (e.g. Balzter *et al.* 1998, and Balzter 2000, or see reviews by Baker 1989, and Perry and Enright 2006), wherein a matrix of transition probabilities between different discrete states at discrete time intervals is derived from observations, then used to predict the distribution of those states in the future. Forest succession models, in which successional stages are represented by discrete states with probabilities of transitions between them, have a long history (e.g. Horn *et al.* 1975 and Leps 1987), and spatial Markov models have been used to model succession (e.g. Van Tongeren and Prentice 1986, Liu *et al.* 2008) as well as deforestation (e.g. Soares-Filho *et al.* 2002, and Moreno *et al.* 2007).

Land-cover data sets derived from remote sensing (see Giri 2012, for a review of the state of the art) have been used to create landscape change models (e.g. Eastman 2012). With MODIS data acquisition in its second decade, MODIS products now make it possible to capture and model deforestation and afforestation (e.g. Aide *et al.* 2013). Compared to in situ measurements, MODIS-based estimates of forest-cover change have been shown to be less accurate than those based on Landsat or higher-resolution imagery (Morton *et al.* 2005), partly due to errors of omission at the sub-pixel scale. Nevertheless, the high spectral and temporal resolution of the MODIS sensor makes possible cloud-free and consistent global products that are particularly useful for multi-temporal applications.

Land-cover change analysis is usually performed offline as a one-time or periodic task. What is novel in the method presented here is that the analysis happens at the time the request is made, allowing landscape trends to be estimated and visualized quickly for any given location, without the need for custom analyses or large data downloads. Additionally, the on-demand approach makes it possible for the model to be invoked for any polygonal region of interest such as an ecoregion, sub-national boundary, or national park. This ensures that the results are representative of conditions within that specific region of interest due to the localization of spatial Markov parameterization.

Spatio-temporal Markov chains

The Markov condition assumes that, given a system with a chain of n states, the state of the system $n + 1$ can be modelled as a function of the system in state n alone, without direct influence from $n - 1$ other states. In the context of modelling land cover over time, the chain is a series of discrete time steps, and the state is a set of mutually exclusive land-cover classes. Given the state of the landscape at time t_0 and at time t_1 , a matrix of transition probabilities from each cover type x_i to each other cover type x_j is described in Equation (1) (adapted from Balzter 2000).

$$P = \begin{matrix} P(x_{0_1}|x_{0_0}) & P(x_{0_1}|x_{0_0}) & \dots & P(x_{0_1}|x_{k_0}) \\ P(x_{1_1}|x_{0_0}) & P(x_{1_1}|x_{1_0}) & \dots & P(x_{1_1}|x_{k_0}) \\ \dots & \dots & \dots & \dots \\ P(x_{k_1}|x_{0_0}) & P(x_{k_1}|x_{1_0}) & & P(x_{k_1}|x_{k_0}) \end{matrix} \quad (1)$$

This allows us to express the probability of land-cover types at time $t + 1$ using Equation (2).

$$p(t + 1) = p(t) \cdot P \quad (2)$$

The matrix P can be derived from discrete data by calculating the frequency of transitions in two source data sets at t_0 and t_1 . Given a large enough number of time steps, a homogenous Markov chain, wherein P remains constant, converges to a steady state. In the case of land-cover modelling, this convergence may not represent a likely future state of the landscape, as exogenous factors not captured in the transitions from t_0 to t_1 are likely to occur over time, changing P .

While this conceptual model can capture non-spatial probabilities, land-cover change is a spatial process, and the probability of change in a given location is affected by its proximity to other landscape elements and processes. The Markov chain approach can be expanded to handle spatial conditions by adding another dimension to the vector of random variables. Equation (3) shows how each element in the matrix P is augmented by a second vector of possible spatial contexts s , so that the land-cover state at time $t + 1$ is dependent on the probability of both the land-cover state and the spatial context at time t , where the range of i and j is the number of land-cover types, and the range of k is the number of spatial context types.

$$P = [P(x_{0_{i+1}}|x_{0_i}, s_{0_i}), \dots, P(x_{i+1}|x_j, s_{k_i})] \quad (3)$$

Cellular automata models of landscape change, descended from the original Game of Life (Conway 1976), expand spatial realism similarly, by including ancillary data such as slope and elevation within the representation of the spatial context of each cell. This kind of spatially enhanced Markov model has been used to model fire spread (e.g. Clarke *et al.* 1994), urbanization (e.g. Clarke and Leonard 1998, Batty *et al.* 1999, Goodchild and Haining 2004, Almeida *et al.* 2005), and other highly spatial phenomena.

Markov chain models on demand

The transition probabilities between different landscape states at different discrete times can be inferred from observed data by counting the occurrence of each possible transition type between two observations. In the case of spatial Markov models based on MODIS tree-cover data, the state of the landscape at t_0 and t_1 can be represented by two gridded data sets. To extract transition probabilities between forested states, these data sets need to represent canopy cover as discrete values, such that transitions between states can be counted. By preprocessing global MODIS product data to meet the requirements of the spatial Markov approach, and making them available via web mapping services, the state of the grid cells falling within any given area at t_0 and t_1 is readily accessible. By extracting the transition probabilities for arbitrary regions of interest (i.e. 'a given area'), the on-demand system ensures that those probabilities are relevant to local conditions.

One problem faced by applications that process raster data is that the data size and computational burden can increase dramatically as the covered area increases. This is a particular issue for interactive applications, which need to return results quickly. Web-based mapping services reduce data transfer time by limiting the resolution of images used to represent larger areas, thereby increasing the area represented by each pixel. We maintain reasonable performance in the on-demand system using the rescaled output of the web map services as input to the model, such that the maximum amount of data processed in any request remains constant.

Source data

The infrastructure supporting the on-demand spatial Markov models includes several gridded data sets that were preprocessed and published as web mapping services. We chose the per cent tree layer within the MODIS version 5.1 Vegetation Continuous Fields product (Townshend *et al.* 2011a) as the basis for this application because it is sensitive to sub-pixel variability (it expresses tree cover as a continuous percentage rather than discrete binary value); it has been used to characterize deforestation (e.g. Chagnon and Bras 2005); it is now available for more than 10 years; and it characterizes per cent tree cover independent of categorical land-cover type, which can help account for tree coverage, for example, in areas that might be classified as urban or agricultural.

Methodology

Data preparation

We downloaded the MOD44B 5.1 data set from 2000 to 2013 from the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center in Sioux Falls, South Dakota, and extracted the per cent tree, quality, and cloud layers. We then re-projected and resampled the data to 1 km resolution using nearest-neighbour resampling to retain the categorical water and fill pixel values, and combined the resulting rasters into global annual mosaics. To counteract the effect of noise, low-quality pixels, and transient pixel values, we combined three-year mosaics into three global rasters: 2000–2002, 2005–2007, and 2010–2012. The three three-year averages are referred to as 2001, 2006, and 2011, respectively, in the discussion below.

Each pixel was assigned the average of the three annual values. Pixels with fill, no data, or water values in one or two of each three-year periods were assigned to the average of the remaining valid pixels. Pixels without values between 0 and 100 in all three years were set to a value of 255. The proof-of-concept system was implemented using ArcGIS Server 10.1, and raster data was accessed by the simulation software using the exportMap web service. Pixel values were 8-bit unsigned integers output without contrast stretching so that the original pixel value ranges were retained. Bilinear interpolation of map layers was enabled to limit distortion due to on-the-fly resampling at different zoom levels. The user interface was developed using the ArcGIS JavaScript API, which invokes web services developed in C#, which in turn query the multi-year MOD44B web map services.

Web service algorithm

The processing steps implemented in the proof-of-concept web service application are the following:

- (1) Calculate the bounding box of the geometry found in the incoming HTTP GET request.
- (2) Issue a map service call to download rasters representing t_0 and t_1 , limited by the dimensions of the bounding box and the maximum number of pixels that can be processed in a single request.
- (3) If the requested geometry is a polygon, set the values in the t_0 to t_1 rasters outside of this polygon to a mask value.

- (4) Apply binning to the remaining pixels so that each pixel contains a discrete tree-cover class label.
- (5) Extract the transition probabilities from the t_0 to t_1 rasters for each spatial context type, by counting the number of cases of each possible transition between classes.
- (6) For each iteration:
 - (a) Simulate the state of each pixel at t_2 by stochastically sampling from the possible new values using the extracted probability distribution associated with each pixel in t_1 , based on its tree-cover class and spatial context.
 - (b) For validation, compare this with the actual t_2 state if it is available, using the kappa index of agreement, described below.
 - (c) Copy the state at t_2 to t_1 .
 - (d) Recalculate the spatial context for each pixel in t_1 .
 - (e) Repeat for a fixed number of iterations, or until the distribution of states stabilizes.

Spatial Markov implementation

The spatial component of the model is defined within each transition type, as an additional dimension in the transition matrix. We defined the spatial context as the weighted proportion of cells adjacent to the current cell with a different tree-cover class than the current cell, using an eight-cell neighbourhood, with corner cells given a lower weight (0.5 for each corner cell vs. 1.0 for rook-adjacent cells). The proportion values, ranging from zero to one, are divided into four equal-sized bins, each indicating a level of divergence between the current cell and its neighbours. To reduce noise, cases where no cells differ from the current cell are left unchanged. In step 4, the MOD44B per cent tree value, which ranges from 0 to approximately 80, is binned into four classes: no data, 0–9%, 10–49%, and 50% and above. These classes are labelled as Class A, Class B, Class C, and Class D, respectively, in the discussion below. No data values, which include fill, water, and no data as described above, are assigned to Class A.

Each transition probability shown in Table 1 is extracted directly from the data by counting the number of cells with each state t_0 in the columns, and the t_1 state in the rows. By normalizing the number to the column total, the cell values represent the conditional probability that a cell with the value given in a column at time t will take on the value in the corresponding row in time $t + 1$. For example, column A_0 represents the probability that a cell given value A at time t will take on the values A, B, C, or D in time $t + 1$.

To estimate the spatial component of the transition, we expand the transition matrix by including the distribution among possible spatial contexts within each possible transition

Table 1. Transition probabilities where A, B, C, and D are the four tree-cover classes, and the subscripts indicate the time step.

	A_0	B_0	C_0	D_0
A_1	$P(A_1 A_0)$	$P(A_1 B_0)$	$P(A_1 C_0)$	$P(A_1 D_0)$
B_1	$P(B_1 A_0)$	$P(B_1 B_0)$	$P(B_1 C_0)$	$P(B_1 D_0)$
C_1	$P(C_1 A_0)$	$P(C_1 B_0)$	$P(C_1 C_0)$	$P(C_1 D_0)$
D_1	$P(D_1 A_0)$	$P(D_1 B_0)$	$P(D_1 C_0)$	$P(D_1 D_0)$

Table 2. Spatial context probabilities for a single tree-cover class, A_0 , where A_1 , B_1 , C_1 , and D_1 are the four land-cover classes at time $t + 1$, and S_{0-3} are the four spatial context types.

Tree-cover class A_0				
	S_0	S_1	S_2	S_3
A_1	$P(A_1 A_0, S_0)$	$P(A_1 A_0, S_1)$	$P(A_1 A_0, S_2)$	$P(A_1 A_0, S_3)$
B_1	$P(B_1 A_0, S_0)$	$P(B_1 A_0, S_1)$	$P(B_1 A_0, S_2)$	$P(B_1 A_0, S_3)$
C_1	$P(C_1 A_0, S_0)$	$P(C_1 A_0, S_1)$	$P(C_1 A_0, S_2)$	$P(C_1 A_0, S_3)$
D_1	$P(D_1 A_0, S_0)$	$P(D_1 A_0, S_1)$	$P(D_1 A_0, S_2)$	$P(D_1 A_0, S_3)$

between states, as shown in Table 2. As in the non-spatial transition matrix, each spatial context column is normalized to sum to one, so each cell value represents the transition probability conditioned on both land-cover class and spatial context.

As in the non-spatial transition matrix, the probabilities are derived directly from the data within the region of interest. Each cell has one of four tree-cover types, and one of four spatial context types, resulting in 16 possible states in each time step. For each state, transition probabilities are estimated for each of the four possible land-cover types in the next time step.

To give a specific example, Figure 1 shows a sample area in central Brazil (64°W 10°S to 62°W 8°S), with simplified tree-cover classes, in the years 2001 and 2011, containing 57,600 cells. Figure 2 shows the transition matrix derived from the area shown in Figure 1, representing the conditional probability of each transition given each spatial context and cover type.

To accommodate different resolutions, the system operates on the values as returned by the underlying web map service. It extracts the bounding rectangle from the submitted region of interest, and sets the maximum number of pixels in either dimension for the bounding rectangle. If this maximum is less than the number of pixels that would be returned at full resolution, the other dimension is reduced proportionally, so the original aspect ratio is retained. The resulting raster dimensions are then included in the request to the map server, and the pixel values for each time span are acquired. Results are expressed as a proportion of the total selected area.

Accuracy assessment

To evaluate the results, we set t_0 to 2001 and t_1 to 2006, extracted the transition probabilities, then generated an output raster for 2011 and compared it to the actual 2011 data. Following the methods described in Congalton and Green (2009), we evaluated the results using categorical error matrices, from which were derived overall accuracy, per-class user's and producer's accuracy, and kappa indexes. The kappa index of agreement takes into account the probability of agreement by chance, and was calculated as follows (adapted from Congalton and Green 2009):

$$\hat{K} = \frac{n \sum_{i=1}^k n_{ii} - n \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} \quad (5)$$

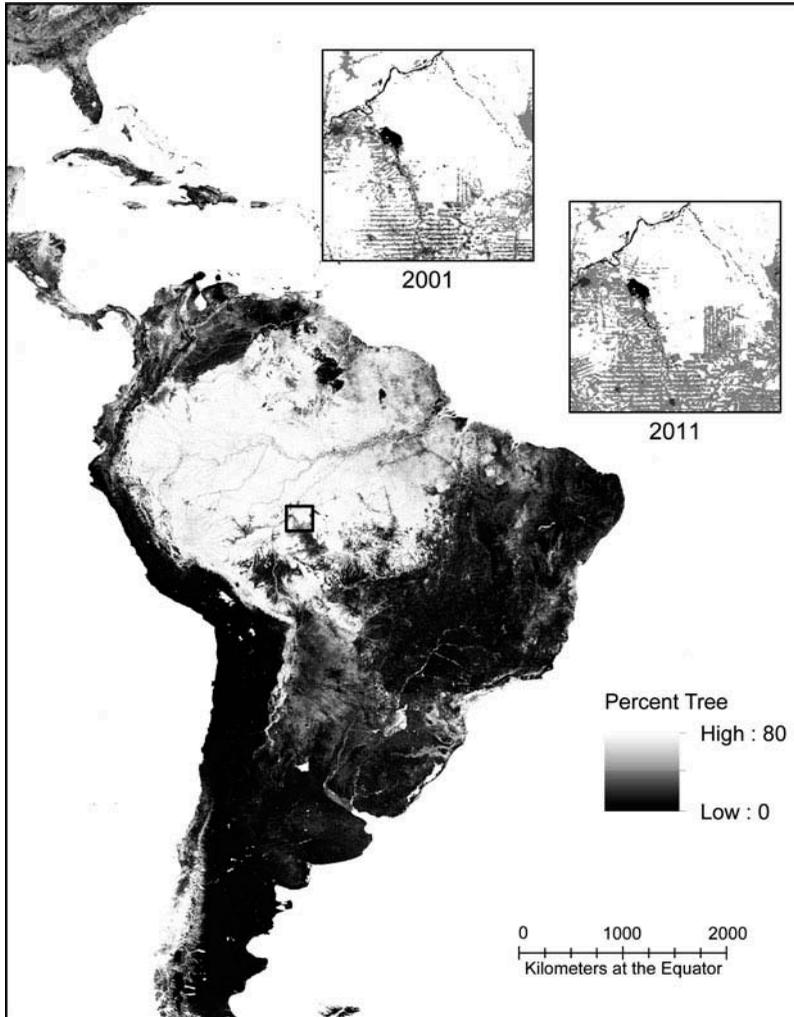


Figure 1. Sample rasters for t_0 and t_1 (2001 and 2011) extracted from the MOD44 B product shown in the background image. Key for inset images: black: water, no data, and fill values; dark grey: <10% per cent tree cover; light grey: 10–49% tree cover; white: 50% tree cover and above.

where \hat{K} is an estimate of kappa, n is the number of observations, n_{ii} are the diagonal (correct) values of the error matrix, n_{i+} are the row totals (estimated classes), n_{+i} are the column totals (reference classes), and k is the number of classes (4). When the system is run normally, it sets t_0 to 2001 and t_1 to 2011, and runs are limited to five 10-year iterations.

Correctly distinguishing between inter-annual variability in satellite-based land-cover products and actual deforestation and afforestation remains an ongoing research problem, and there are significant issues in accurately inferring land-cover change from MODIS products (e.g. Song *et al.* 2014). The focus of this present work is the proof-of-concept for on-demand spatial modelling, and we perform preliminary model

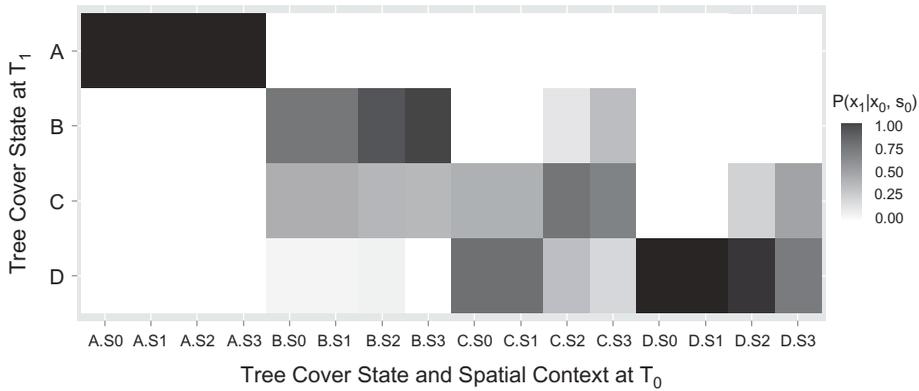


Figure 2. Transition probabilities extracted from the two sample rasters in Figure 1. The x-axis shows the combination of spatial context S_0 and tree-cover state X_0 at t_0 . The y-axis shows the estimated state X_1 at t_1 . The shading represents the conditional probability of X_1 given both X_0 and S_0 . A, B, C, and D represent the four tree-cover classes.

validation as part of this description, apart from validation against ground truth. Additional limitations include the accuracy of the underlying MOD44B product: the 5.1 (2011) version includes improvements over the earlier version (see the MOD44B 5.1 Users Guide, Townshend *et al.* 2011b), and the initial accuracy was assessed with a root mean squared error (RMSE) of approximately 10% at two sets of validation sites. Also, the binned representation of tree cover adopted by the model makes it less sensitive to sub-pixel variability, and the simplified representation of spatial context also limits the representation of other factors that can affect transition probabilities, such as proximity to roads, which is strongly correlated with deforestation (see Eastman 2012).

Sample areas

To evaluate the models generated by the on-demand process, we chose 10 sample areas from diverse biomes, as shown in Figure 3 and Table 3. We refer to them by number in the subsequent figures and discussion.

Sites 2 and 3 were used as calibration sites, and the parameters of the model that visibly affected the results, and that affected the accuracy statistics, were adjusted. The accuracy statistics described above were calculated by comparing the 2011 data to the results from running the model for one time step from the 2001 data, resulting in reference and simulated rasters that were compared. The model was calibrated by selecting the measure of spatial context (the proportion of cells within the window that have a different land-cover type than the current cell); the number of categories used to characterize the spatial context (four); and the treatment of waterbodies and barren areas (these were assumed to be invariant).

Results and discussion

Using the validation methods described above, generating a model from 2001 and 2006 and using it to predict 2011, then comparing the 2011 results to the actual 2011 data, for

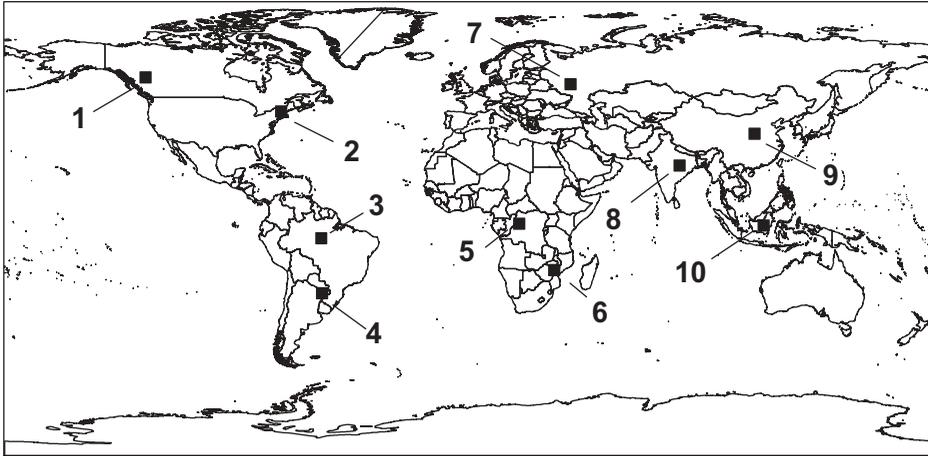


Figure 3. Sample sites used to evaluate spatial Markov models.

Table 3. Sample sites used to evaluate spatial Markov models.

Site	Description	Coordinates
1	British Columbia, Canada	123°W 52°N 121°W 54°N
2	New England, USA	74°W 42°N to 72°W 44°N
3	Amazon Basin, Central Brazil	64°W 10°S 62°W 8°S
4	Northern Argentina	62°W 27°S 60°W 25°S
5	Gabon, Central Africa	13°E 0 15°E 2°S
6	East South Africa	29°E 25°S 31°E 23°S
7	Eastern Russia	33°E 56°N 35°E 58°N
8	Eastern India	83°E 20°N 85°E 22°N
9	Central China	106°E 33°N 108°E 35°N
10	Borneo	112°E 0 114°E 2°N

the sample sites, the non-spatial results show an R^2 of 0.98 and an RMSE of 2625 cells, or 18.2% of the mean pixel count for each land-cover class and site, as shown in the scatterplot in Figure 4.

The high value for the coefficient of determination here does not necessarily reflect accuracy in the model. Most per cent tree classes remain unchanged, and a simple assumption of a constant state also results in a high R^2 value. A more appropriate metric is the kappa index of agreement, which is shown in Figure 5 for each site. A positive value indicates agreement above chance, with values closer to 1 indicating more complete non-chance agreement, 1.0 being perfect agreement, and -1.0 perfect disagreement.

Table 4 shows the error matrix for all sites taken together, indicating an overall agreement of 82%, and a per-class agreement varying between 73% and 88%. Class A, which encompasses water, fill, and no data values, shows full agreement because the data summarization methods exclude those pixel changes from the input rasters in the current implementation of the system.

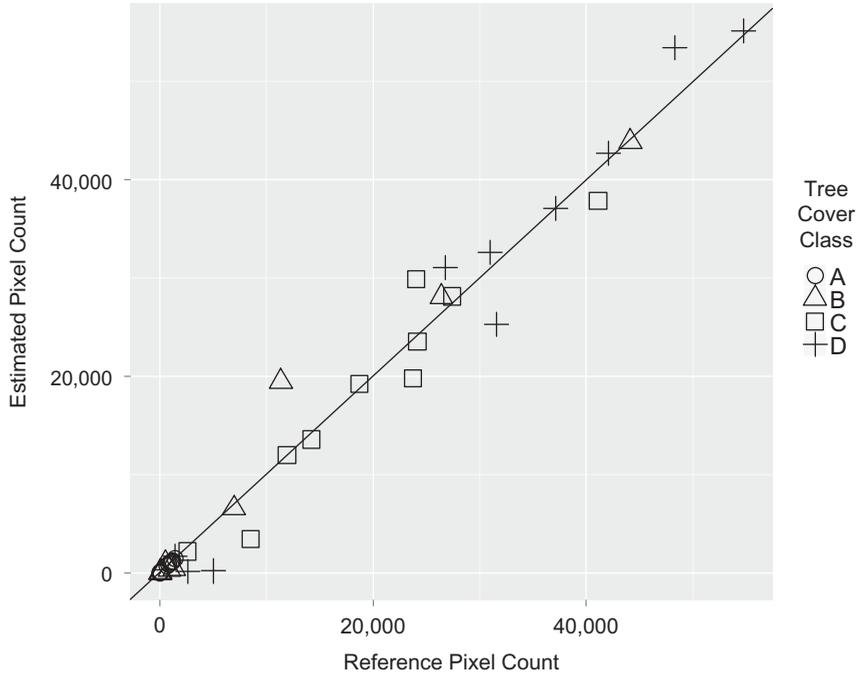


Figure 4. Projected versus actual pixel counts per class and site ($n = 40$), with $R^2 = 0.98$ and RMSE = 2625 cells, or 18.2% of the mean reference value.

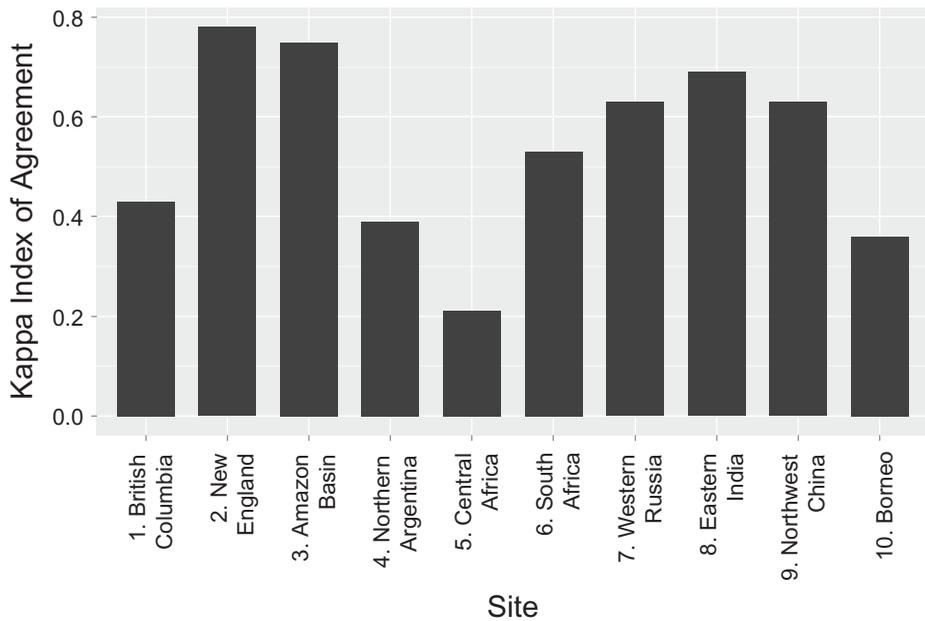


Figure 5. Kappa index calculated for each site.

Table 4. Error matrix for all sites combined.

		2011 actual				
2011 predicted	N/A	<10%	10–49%	≥50%		
N/A	6,707	0	0	0	1	
<10%	0	77,778	13,622	515	0.85	
10–49%	0	21,065	142,955	32,397	0.73	
≥50%	0	1,125	33,072	246,524	0.88	
	1	0.78	0.75	0.88	0.82	

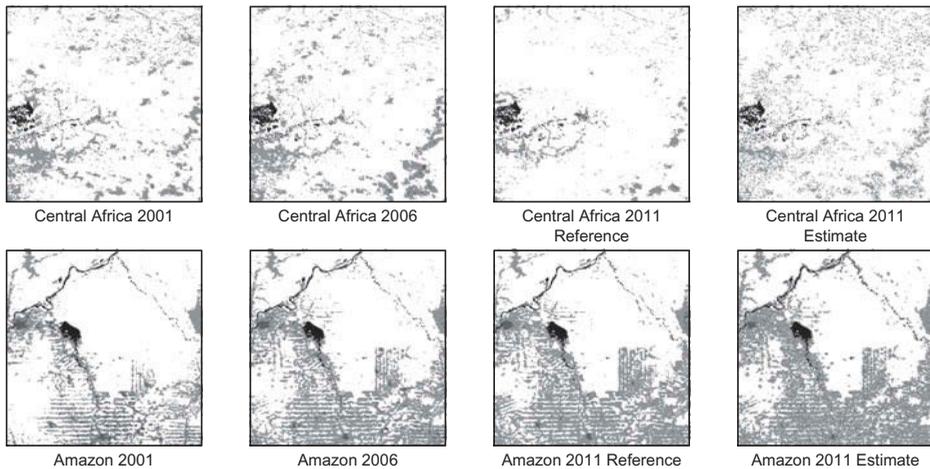


Figure 6. Input and output rasters for the Central Africa and Amazon Basin sites. Key: black: water, no data, and fill values; dark grey: <10% per cent tree cover; light grey: 10–49% tree cover; white: 50% tree cover and above.

Looking more closely at the per-site kappa values, we can see that Central Africa and Borneo constitute the low range of values, while the Amazon Basin and North Eastern USA show higher kappa values. Figure 6 shows the t_0 and t_1 rasters for the Central Africa and Amazon Basin samples. Both the Central Africa and Borneo inputs show diffuse patterns of forest-cover change. This is consistent with the results reported by Hansen *et al.* (2013), who found a higher degree of small-scale fragmentation in Central Africa versus sites in South America. The New England and Amazon Basin inputs show more spatial coherence. The spatial Markov process relies on the spatial adjacency of landscape change, and the wide range in kappa values is likely a symptom of the varying spatial scale and pattern of the underlying causes of disturbance.

Computationally, performance is affected strongly by the maximum raster size permitted by the system. For example, limiting the dimensions of the results to a maximum of 200 pixels in either direction resulted in maximum processing time for a rectangular region of interest of approximately 5 seconds for five time steps, while setting the maximum to 400 increased the processing time to approximately 30 seconds. When the region of interest is polygonal, the processing time increases with the complexity of the polygon, and a large polygon can take several minutes to process.

Conclusion

In this paper we demonstrated a method of creating and running spatial Markov models on demand, to infer approximate trends in land cover and their spatial arrangement over time. This approach eliminates a number of practical barriers to the detection of land-use change for any location on Earth. First, instead of requiring time-consuming data downloads, re-projection, and other data management tasks, the on-demand approach estimates the trends and simulates future land-cover distributions on the fly. Second, the techniques employed have been long-time components of the landscape modelling toolbox; making them available online may help make them more readily accessible to a broader range of end-users.

When the system is used to predict forested areas for the temporal range within which reference data are available, the coefficient of determination is ~ 0.98 , and the RMSE is 18.2% of the mean reference pixel count, across 10 sample sites and four tree-cover classes. The spatial prediction shows moderate per-pixel accuracy across most of the sites, with lower accuracy among sites with less spatially coherent patterns of change. Over longer model timeframes, accuracy is likely to decrease as additional sources of disturbance and change emerge that are not captured in the initial time steps used to create each model.

We expect the accuracy of these results to increase in future iterations of the application. The current availability of global road network data (e.g. Open Street Map, openstreetmap.org) suggests that the accuracy of the system could be improved by incorporating distance to roads among its set of spatial predictors, using on-demand access to road network data. As more years of MODIS land-cover data accumulate, we also expect the value and reliability of the predictions to increase as well. Additional enhancements include optimizing the simplification of land-cover classes based on the classes present in each input area. Similarly, optimizing the definition of each spatial and temporal neighbourhood to account for the differing rates and patterns of change in different regions of interest and at different spatial scales should also improve the relevance and accuracy of the results.

Acknowledgements

The authors would like to thank the two generous anonymous reviewers whose suggested revisions greatly improved the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Aide, T.M., *et al.*, 2013. Deforestation and reforestation of Latin America and the Caribbean (2001–2010). *Biotropica*, 45 (2), 262–271. doi:[10.1111/j.1744-7429.2012.00908.x](https://doi.org/10.1111/j.1744-7429.2012.00908.x)
- Almeida, D., *et al.*, 2005. GIS and remote sensing as tools for the simulation of urban land use change. *International Journal of Remote Sensing*, 26 (4), 759–774. doi:[10.1080/01431160512331316865](https://doi.org/10.1080/01431160512331316865)
- Baker, W.L., 1989. A review of models of landscape change. *Landscape Ecology*, 2 (2), 111–133. doi:[10.1007/BF00137155](https://doi.org/10.1007/BF00137155)
- Balzter, H., 2000. Markov chain models for vegetation dynamics. *Ecological Modelling*, 126 (2–3), 139–154. doi:[10.1016/S0304-3800\(00\)00262-3](https://doi.org/10.1016/S0304-3800(00)00262-3)
- Balzter, H., Braun, P.W., and Köhler, W., 1998. Cellular automata models for vegetation dynamics. *Ecological Modelling*, 107 (2–3), 113–125. doi:[10.1016/S0304-3800\(97\)00202-0](https://doi.org/10.1016/S0304-3800(97)00202-0)

- Batty, M., Xiem, Y., and Sun, Z., 1999. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems*, 23 (3), 205–233. doi:10.1016/S0198-9715(99)00015-0
- Chagnon, F.J.F. and Bras, R.L., 2005. Contemporary climate change in the Amazon. *Geophysical Research Letters*, 32, 13. doi:10.1029/2005GL022722
- Clarke, K.C., Brass, J.A., and Riggan, P.J., 1994. A cellular automaton model of wildfire propagation and extinction. *Photogrammetric Engineering and Remote Sensing*, 60 (11), 1355–1367.
- Clarke, K.C. and Leonard, J.G., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12 (7), 699–714. doi:10.1080/136588198241617
- Congalton, R. and Green, K., 2009. *Assessing the accuracy of remotely sensed data: principles and practices*. 2nd ed. Boca Raton, FL: CRC Press.
- Conway, J., 1976. *On numbers and games*, Vol. 6. London: Academic Press.
- Eastman, J.R., 2012. *IDRISI Selva*. Worcester, MA: Clark University.
- EPA, 2013. *Global Greenhouse Gas Emissions* [online]. Available from: <http://www.epa.gov/climatechange/ghgemissions/global.html> [Accessed July 2014].
- Giri, C., ed., 2012. *Remote sensing of land use and land cover; principles and applications*. Boca Raton, FL: CRC Press.
- Goodchild, M.F. and Haining, R.P., 2004. GIS and spatial data analysis: converging perspectives. *Papers in Regional Science*, 83 (1), 363–385. doi:10.1007/s10110-003-0190-y
- Hansen, M., Potapov, P., and Turubanova, S., 2013. Use of coarse-resolution imagery to identify hot spots of forest loss at the global scale. In: F. Achard and M. Hansen, eds. *Global forest monitoring from earth observation*. Boca Raton, FL: CRC Press.
- Horn, H.S., Cody, M.L., and Diamond, J.M., 1975. Markovian properties of forest succession. In: M.L. Cody and J.M. Diamond, eds. *Ecology and evolution of communities*. Cambridge, MA: Harvard University Press.
- Leps, J., 1987. Vegetation dynamics in early old field succession: a quantitative approach. *Vegetatio*, 72 (2), 95–102.
- Liu, D., et al., 2008. Using local transition probability models in Markov random fields for forest change detection. *Remote Sensing of Environment*, 112 (5), 2222–2231. doi:10.1016/j.rse.2007.10.002
- Moreno, N.R., et al., 2007. Biocomplexity of deforestation in the Caparo tropical forest reserve in Venezuela: an integrated multi-agent and cellular automata model. *Environmental Modelling and Software*, 22 (5), 664–673. doi:10.1016/j.envsoft.2005.12.022
- Morton, D.C., et al., 2005. Rapid assessment of annual deforestation in the Brazilian Amazon using MODIS data. *Earth Interactions*, 9 (8), 1–22. doi:10.1175/EI139.1
- Perry, G.L. and Enright, N.J., 2006. Spatial modelling of vegetation change in dynamic landscapes: a review of methods and applications. *Progress in Physical Geography*, 30 (1), 47–72. doi:10.1191/0309133306pp469ra
- Soares-Filho, B.S., Cerqueira, G.C., and Pennachin, C.L., 2002. DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154 (3), 217–235. doi:10.1016/S0304-3800(02)00059-5
- Song, X.P., et al., 2014. Annual detection of forest cover loss using time series satellite measurements of percent tree cover. *Remote Sensing*, 6 (9), 8878–8903. doi:10.3390/rs6098878
- Townshend, J.R. et al., 2011a. *Vegetation continuous fields MOD44B, 2001 percent tree cover, collection 5*, University of Maryland, College Park, MD, 2000 to 2013. digital data. Available from: lpdaac.usgs.gov [Accessed 3 December 2015].
- Townshend, J.R. et al., 2011b. *User guide for the MODIS vegetation continuous fields product, collection 5 version 1*, University of Maryland, College Park, MD.
- Van Tongeren, O. and Prentice, I.C., 1986. A spatial simulation model for vegetation dynamics. *Vegetatio*, 65, 163–173. doi:10.1007/BF00044816