Land use change modelling using cellular automata

Ane Dicks Herrán, Quincy van Iersel and Robert Locher

Utrecht University

February 2023

1. Introduction

Land use change is one of the research subjects of global environmental change and can be advantageous at ensuring future sustainable land development (Guan et al., 2011). The aim of this paper is investigating the land change of residential areas in Randstad. The interest in this lies in the lack of housing the Netherlands has been experiencing (Plasschaert, C. & Venema, A., 2021). By simulating the growth of residential areas, potential new residential areas in the real world can be identified. To further investigate this problem, the two questions that are addressed in this study are (i) Is the growth of a residential area dependent on the size of its city? and (ii) Are there going to be more residential areas in the upcoming years?

One of the methods that can be used simulating land use change is cellular automata (CA). CA has been proved to be an outstanding method in predicting future land use change through spatial simulation and modelling process. Therefore, in order to address the research question, three CA models are proposed to simulate the growth of residential areas in Randstad since the year 2000. The paper starts discussing the underlying concepts of CA for forecasting land use change. Next, the used datasets are described and the three CA models used for this case-study are introduced in Section 3. The results obtained from the models are presented and discussed in Section 4. Finally, the conclusions, remaining problems and possible improvements of the model are presented in Section 5.

2. Cellular automata

CA is a dynamic model that can be used to simulate future land use changes. In CA, a landscape is divided into a grid of cells, each of which can be in one of a set of possible states. Examples of state of cells in spatial data analysis are agriculture land, forest land, urban land, other land, water, etc. The model simulates the state change of a cell based on the state of its neighbouring cells and based on a set of transition rules that define each cell's transition (Wu, 2002). *Figure 1* shows an example of two of the most commonly defined neighbourhood configurations, namely the Moore neighbourhood and the Von Neumann neighbourhood. Moreover, it also shows some of the different neighbourhood sizes that can be used for CA.

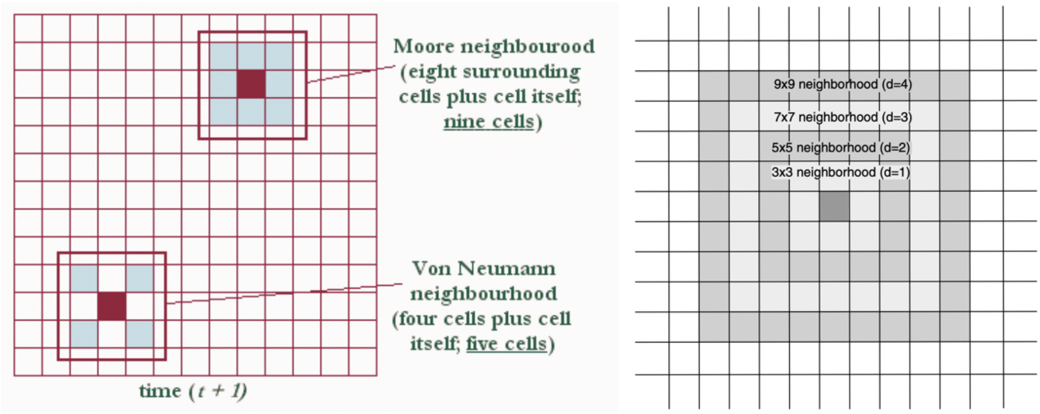

Figure 1: Two-dimensional CA with distinct neighbourhood sizes and distinct neighbourhood configurations.

Moreover, transition rules are said to take a key role because they specify the behaviour of cells between time-step evolutions, deciding the future states of cells. Therefore, it is important to give attentiveness to the selected rules to obtain realistic simulations of land use as, on the contrary, it could lead to a biased model where the transition rules are defined in such a way that they implicitly connote a pre-defined aim of how the model would develop (Clarke, 2014).

Neighbourhood characteristics

Neighbourhood characteristics define the relationships between locations (i.e. cells). A way to characterise the neighbourhood of a location in a land use map is the enrichment factor (F), which represents for each cell the occurrence of a land use type in the neighbourhood of a location relative to the occurrence of this land use type in the whole study area. The F is shown in *formula (1)*.

$$F_{i,k,d} = \frac{n_{k,d,i}/n_{d,i}}{N_k/N} \quad (1)$$

where i is the cell location, k is the land use type and d is the neighbourhood. $n_{k,d,i}$ is the number of cells of land use type k in the neighbourhood d of cell i, $n_{d,i}$ is the total number of cells in the neighbourhood d, N_k is the number of cells with land use type k in the whole raster and N is the number of cells in the whole raster (Verburg et al., 2004).

Logistic regression model

The analysis of the enrichment factor alone is not enough to explain to what extent the spatial pattern can be explained by the neighbourhood characteristics. Therefore, a logistic regression can be used to relate the location of changed land use with the calculated enrichment factors, as shown in *formula (2)*

$$\mathrm{Log}\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_{\text{grass},d=1}F_{i,\text{grass},d=1} + \beta_{\text{forest},d=1}F_{i,\text{forest},d=1} + \beta_{\text{arable},d=1}F_{i,\text{arable},d=2} + \cdots + \beta_{k,d}F_{i,k,d} \quad (2)$$

where P_i is the logistic probability of conversion of a grid cell, $F_{i,k,d}$ are the enrichment factors for each cell location i of the neighbourhood d with land use k, and $\beta_{k,d}$ are the coefficients which can be estimated with Machine Learning applications (Verburg et al., 2004).

Conversion probabilities

In order to obtain the total probability of a cell to transition at a specific timestep, the resulting probability from *formula (2)* can be used and expanded with the following *formula (3)*, as suggested by *Liao et al., (2016)*:

$$P_{ij}^{t} = (P_l)_{ij} \cdot (P_\Omega)_{ij} \cdot con\,() \cdot P_r \quad (3)$$

where P_{ij}^{t} is the total probability of a cell i to transition at moment t, $(P_l)_{ij}$ is the logistic probability of a cell, $(P_\Omega)_{ij}$ is the development density of a cell, *con()* is a local constant and P_r is a stochastic disturbance factor (between 0 and 1).
This conversion probability can be used as a transition rule to simulate land use change.

3. Methodology

3.1 Study area overview

The dataset '*Ontwikkeling bodemgebruik in Nederland 1996-2000'* was retrieved from Centraal Bureau voor de Statistiek (CBS). The study area is located in Randstad, the Netherlands. The rasterized map containing the land use of Randstad in the year 2000 is shown in *Figure 2*. The map has a scale of 1: 10 000, which is equal to a 5-metre resolution.

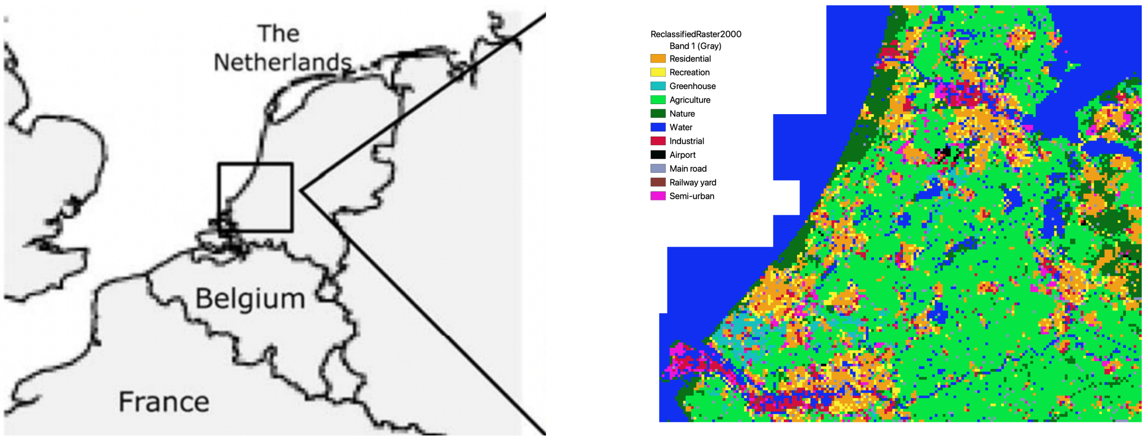

Figure 2: Location of Randstad region and land use in year 2000

Table 1 Randstad map legend

Legend							
	Residential		Agriculture		Industrial		Train tracks
	Recreation		Forest/ Nature		Airport		Semi-urban
	Greenhouse		Water		Main road		

The classification in the original raw map consisted of 9 main groups, which were subdivided into a number of separate categories. In total, there were 33 classifications. In order to account for model complexity and time constraints, the classification was summarised into 11 main groups, without any subgroups. The coloured classification used for the map is shown in *Table 1*. Moreover, *Table 2* shows the main land use types and the subgroups assigned to each group.

Table 2 : Land use classification used in the analysis

Land use type	Contents
Residential	Residential, shops, public facility, cultural facility
Recreation	Parks, sports facilities, allotment, day recreation area, stay tourism area
Greenhouse	Greenhouse

Agriculture	Agriculture sites
Forest/ Nature	Forest, dry land, wet land
Water	Ijsselmeer, closed estuary, Rhine and Meuse, Borderlake, reservoirs, recreational waters, mineral extractional waters, Sludge area, other inside area, Weddensea, Eems, Dollard, Easterscheld, Westerscheld, North sea
Industrial/Commercial	Industrial site
Airport	Airport
Main road	Main road
Train tracks	Railway yards
Semi-urban	Dump, wreck storage, cemetery, mineral extraction site, building site, other semi-paved area

Moreover, land use data in Randstad for the year 2017 was also retrieved from CBS to compare with the map for the year 2000 and the obtained results. This map was first rasterized and then converted in the same raster format as the map for the year 2000. The same classification and colour-coding for the map were assigned, as shown in *Figure 3*.

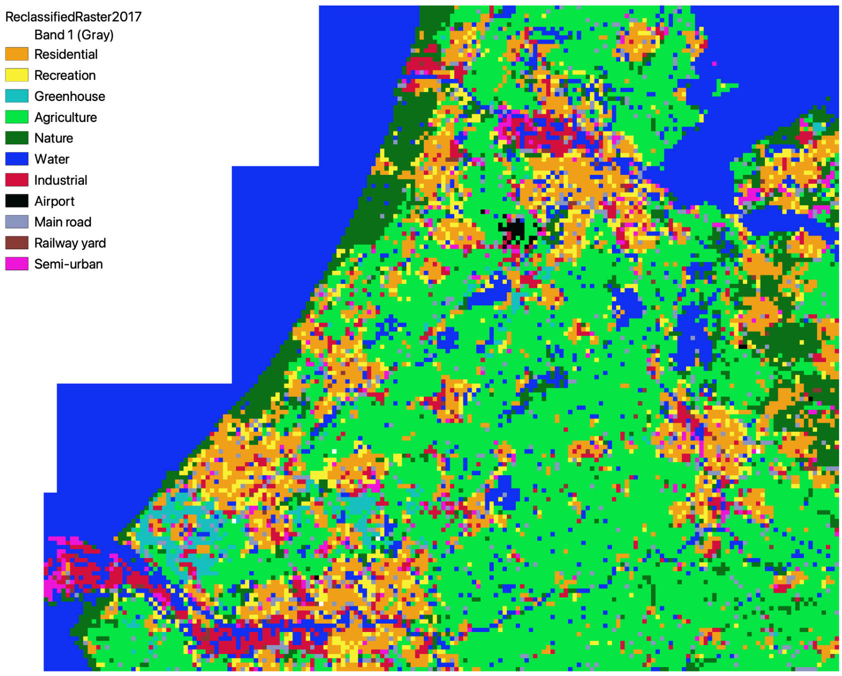

Figure 3: Land use of Randstad in year 2017

3.2 Proposed CA models

3.2.1 Model 1 - PCRaster model with pre-defined logistic regression coefficients

The CA model 1 used in this study was implemented in PCRaster. The code and the category table used can be found at https://github.com/anedicksh/CA-landusechange-Randstad.

For the CA model, the categories that are shown in *Table 2* were used as initialising scalars. Next, the model calculated the enrichment factors for neighbourhoods of size 1 and 3 (3x3, 7x7). These factors were used as coefficients together with the state of the cell itself. Each coefficient had its own weight.

Next, a new value per cell was calculated based on the implemented function which imposed a residential cell to stay residential. On the contrary, if the cell was non-residential , it would acquire its previously assigned land type.

The attempt to simulate Model 1 was by using the enrichment factors calculated in PCRaster and the coefficients shown in *Table 3*, which were proposed by Verburg et al. (2014). These coefficients were chosen as the study conducted by *Verburg et al.* (2014) also consisted of modelling land use change in Randstad between 1989 and 1996. These coefficients were implemented in the logistic regression model, *formula (2)*, to explain the spatial distribution of new residential areas by the enrichment of the neighbourhood.

Table 3: Logistic regression coefficients

Variable	Estimated coefficient	Odds ratio
Constant	−5.537*	–
$F_{\text{residential},i,1}$	0.364*	1.438
$F_{\text{industrial},i,1}$	0.023*	1.023
$F_{\text{forest/nature},i,1}$	−0.201*	0.818
$F_{\text{residential},i,3}$	0.090*	1.094
$F_{\text{industrial},i,3}$	0.028*	1.029
−2 Log likelihood	9478	
χ^2	2594*	
ROC	0.91	

*Significant at $p < 0.01$.

3.2.2 Model 2 - MOLUSCE QGIS plug-in

The land use change was also predicted using MOLUSCE plug-in implemented in QGIS software, resulting in Model 2. The plug-in required as input data two maps (land use map for the year 2000 and year 2017) and the spatial variables, which corresponded to the close and distant enrichment factors for every land use type and for three different sizes of neighbourhood (3x3, 5x5 and 7x7). Based on these enrichment factors and the land use change between the year 2000 and 2017, the plug-in would predict land use in the year selected by the user. Transition probabilities were also obtained with MOLUSCE.

3.2.3 Model 3 - PCRaster with transition probabilities

Model 3 did not make use of the logistic regression model. Instead, it made use of the transition probabilities obtained with MOLUSCE and the close neighbourhood enrichment factors. First, the model determined whether a cell would transition by chance and, if this was the case, it checked whether the cell had a close neighbourhood enrichment factor above 1. If these two requirements were met, the cell would transition to another land type.

4. Results and discussion

4.1 CA model in PCRaster

The land use prediction for Randstad in PCRaster did not yield realistic results. It is suspected that it was because the coefficients proposed by *Verburg et al.* were obtained using a different map of Randstad, namely for the year 1989. Therefore, for the different timestep and map format used to obtain those coefficients, it probably did not apply to the map used for this case-study for the year 2000. In addition, when running the simulation in PCRaster, land would stop changing earlier than expected because only residential areas were being updated, instead of updating all land types. Therefore, a different approach was taken by implementing the MOLUSCE plug-in in QGIS to see whether more substantial results could be obtained.

4.2 CA model in MOLUSCE QGIS plug-in

This model was expected to give interpretable coefficients for the logistic regression model that could then be implemented back in the PCRaster model. However, this was not the case. Instead, the plug-in displayed the coefficients for the logistic regression function for each conversion possibility, instead of for an overall regression model. In addition, these coefficients could not be implemented into the PCRaster model and a functional transition rule could not be obtained from these results.

Moreover, transition probabilities for every combination of land use were also obtained with MOLUSCE, which are shown in the *Appendix*. The columns correspond to the land types used for the initial map (year 2000) and the rows correspond to the land types for the predicted map (year 2017). These probabilities can be interpreted as follows: For example, the transition probability for residential to stay residential in the year 2000 until 2017, is 91.7%.

Finally, the resulting land use prediction with MOLUSCE for the year 2017 can be seen in *Figure 4*. If *Figure 3* is compared with *Figure 4*, it can be seen that, on average, it gives similar results in terms of the growth of residential areas.

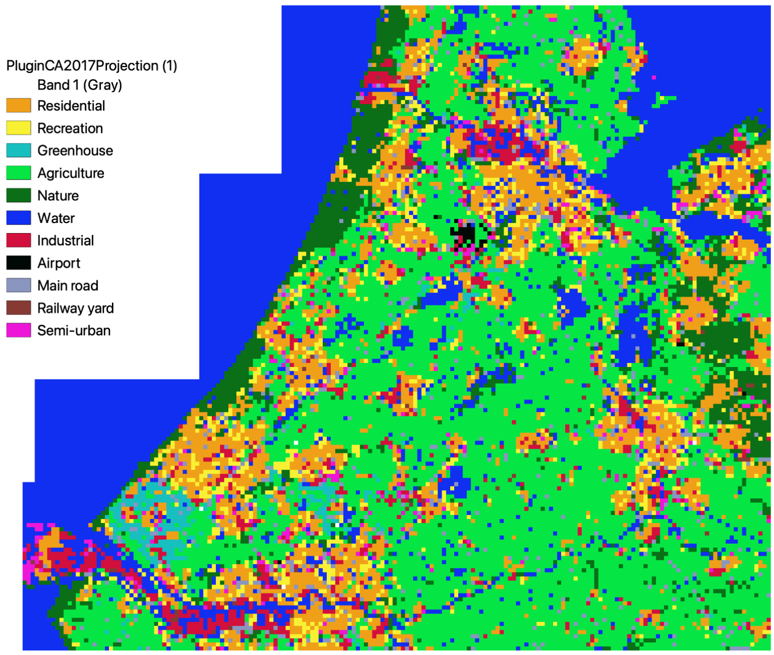

Figure 4: Predicted land use in 2017 according to MOLUSCE

However, it was concluded that this model would not be an adequate choice to use for predicting land use change in Randstad. This is because the plug-in lacked the functionality to simulate more than one timstep accurately.

4.3 Model 3 - PCRaster with transition probabilities

The predicted land use for the year 2017 using model 3 is shown in *Figure 5*. The colour-coding in this map is different to the maps presented until now. *Table 4* shows the colour coding used for *Figure 5*. From Figure 5, it is concluded that this model yielded visually the best results among the three models. However, it is also the most biased model as the approach taken was not accurately justified as to why it was done in such a way.

Table 4: Figure 5 map legend

Legend			
Residential	Agriculture	Industrial	Train tracks
Recreation	Forest/ Nature	Airport	Semi-urban
Greenhouse	Water	Main road	

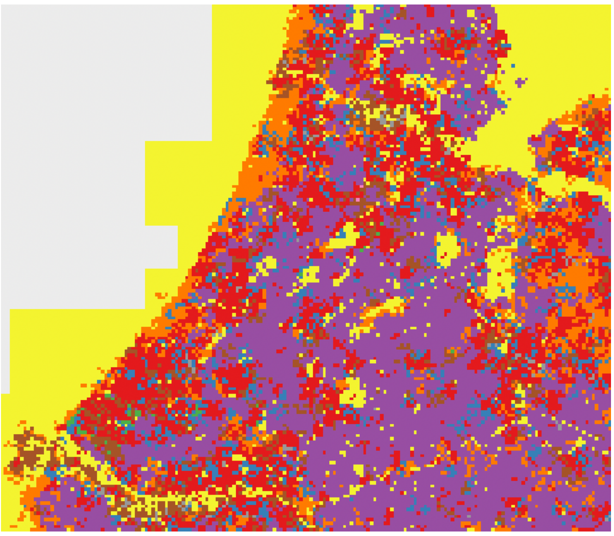

Figure 5: Predicted land use in 2017 according to Model 3

To explain how this model works, the conversion from recreation to residential land is taken as an example: this transition has a transition probability 0.05, so there is 5% chance for a recreation cell to transition into residential. However, if only the transition probabilities are taken into account, random recreation cells also transition into residential. Therefore, the close neighbourhood enrichment factors for residential were taken into account and those had to be above 1 in order to allow for that specific cell to transition into residential.

The way this model is implemented only allows land types to expand and not to appear in a location where that land type does not appear. For example, airports will only expand with this formula, not allowing new airports at random points because there has to be an airport in the neighbourhood in order for a cell to transition into an airport land type. To improve this model, a better transition rule should have been used.

5. Conclusion

In this research the land use change for residential areas was simulated by using three different CA models. Unfortunately, due to the high amount of bias from all the three models it was impossible to answer the research question as to whether bigger cities are more prone to expand to residential areas. Out of the three models, model 1 was the least biased because it had the most appropriate methods according to previous studies (Clarke, 2014) (Verburg et al., 2004). The second model was biased because it was trained on the same data as it was validated on. Moreover, the third model was too biased to answer the proposed research question because the close enrichment factor above 1 means that residential areas will only grow in places where the residential cells per neighbourhood is larger than the average of the map.

Some of the limitations experienced in this case-study were the use of pre-defined coefficients for model 1 which did not work, brute forcing coefficients until something accurate was obtained and the lack of useful data to accurately model the land use change. Therefore, for future research it is recommended to obtain more datasets throughout years instead of one single year. Another recommendation is to look into different machine learning techniques where you can properly implement the coefficients from PCRaster.

References

Clarke, K. C. (2014). Cellular automata and agent-based models. . *Handbook of regional science*, 1217-1233.

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., & Hokao, K. (2011). Modeling urban land use change by the integration of cellular automata and Markov model. *Journal of Ecological Modelling, vol. 222*, 3781-3772.

Plasschaert, C. & Venema, A. (2021). Hoe groot is het huizentekort en wat kunnen we eraan doen? *nu.nl* https://www.nu.nl/economie/6158671/hoe-groot-is-het-huizentekort-en-wat-kunnen-we-eraan-doen.html

Liao, J., Tang, L. Shao, G., Su, X., Chen, D., & Xu, T. (2016). Incorporation of extended neigborhood mechanisms and its impact on urban land-use cellular automata simulations. *Environmental Modelling & Software, 75*, 163-175

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

Wu, F. (2002). Calibration of stochastic cellular automata: the application to rural-urban land conversions. *International Journal of Geographical Information Science. vol. 16, iss. 8*, 795-818.

Appendix

Table 1: Transition probabilities obtained with MOLUSCE plug-in. Numbering dictionary is as followed: 1- Residential, 2- Recreation, 3- Greenhouse, 4- Agriculture, 5- Nature, 6- Water, 7- Industrial, 8- Airport, 9- Main road, 10- Railway yard, 11- Semi-urban.

		New map (year 2017)										
		1	2	3	4	5	6	7	8	9	10	11
Old map (year 2000)	1	0.917	0.026	0.000	0.008	0.004	0.006	0.023	0.000	0.008	0.000	0.005
	2	0.059	0.805	0.000	0.026	0.042	0.012	0.012	0.000	0.017	0.000	0.024
	3	0.063	0.007	0.653	0.199	0.000	0.002	0.035	0.000	0.009	0.000	0.030
	4	0.019	0.015	0.006	0.903	0.015	0.009	0.009	0.000	0.006	0.000	0.013
	5	0.015	0.021	0.000	0.048	0.873	0.012	0.007	0.000	0.008	0.001	0.009
	6	0.001	0.001	0.000	0.002	0.005	0.980	0.002	0.000	0.000	0.000	0.004
	7	0.057	0.007	0.000	0.006	0.001	0.006	0.882	0.002	0.007	0.003	0.024
	8	0.000	0.000	0.000	0.088	0.058	0.000	0.058	0.705	0.000	0.000	0.088
	9	0.038	0.014	0.000	0.056	0.008	0.004	0.004	0.001	0.849	0.002	0.017
	10	0.026	0.043	0.000	0.061	0.000	0.008	0.061	0.000	0.052	0.745	0.000
	11	0.217	0.131	0.007	0.081	0.035	0.055	0.197	0.003	0.048	0.010	0.210