Exploring Neighbourhood Interaction with Desktop GIS

Lena Hallin-Pihlatie
Finnish Environment Institute, P.O. Box 140 FI-00251 HELSINKI
lena.hallin-pihlatie@ymparisto.fi

INTRODUCTION

It is a recognized fact that in order to make good land-use models the processes underlying land-use change need to be sufficiently known. The land-use pattern that forms our urban landscape of today has emerged from a complex interaction between the human and the natural environment. A study made by Verburg (2004a) indicates that the historical land-use pattern can be explained by the conditions of soil and land form, while more recent land-use patterns and changes within these are seem to determined by accessibility, neighbourhood interaction and by spatial policies and plans (Verburg, 2004a).It seems as the proximity to urban land-use and the accessibility to existing networks steer the general location of present land-use change. A bottom-up driven process like this is referred to as a process of self-organisation, where the system self-organises its internal structure independent of external causes (Portugali, 1999). This development is to different degrees altered by spatial planning. Planning-induced constraints represent the top-down steered part of land-use change.

Cellular Automata (CA) models have been proven useful for modelling different kinds of bottom-up self-organising systems, including urban land-use. In CA models, the state of a cell is determined by a transition rule, where the prevailing state of a cell's neighbourhood at a time t returns a certain kind of cell state at time $t+1$ (O'Sullivan, 2000). Whereas in the traditional CA approach all actions of interest are local, taking place in the immediate vicinity of the cell, urban geographical phenomena is often affected by a broader neighbourhood. In an urban CA the neighbourhood is therefore usually defined as an extended neighbourhood (Batty 2005:73). For each land-use function, the transition rule is a weighed sum of distance functions calculated relative to other land-use functions and features. Ideally these transition rules represent the competition of human activities within the urban area in focus (Engelen, 2002). However presently, the choice of transition rule in urban CA models are only loosely based on empirical studies on land-use dynamics (Verburg, 2004a, Geertman, 2007, Malcewski, 2000), which is a serious drawback, since what actually has happened and happens within the neighbourhood is of central importance in the context of modelling urban growth.

The constraints on land-use imposed by land-use plans and spatial policies are often of a binary nature; either a certain type of building activity is allowed or prohibited, and therefore, are these constraints relatively easy to take into account into CA-based land-use models, if appropriate data is available. Taking into account the neighbourhood interaction and is far more complicated. In order to do so, we need to be able to answer questions such as: Which land-use repel and attract each other? How much? Does this neighbourhood interaction change over distance? Where is the interaction at its maximum and when does it even out? Is the neighbourhood interaction alike or is it different in different region and what happens to the neighbourhood interaction when we change the scale of observation? There are no universal answer to these questions and many authors have recently indicated that this is an urgent field of research (Hagoort, 2008, Hansen, 2008, Geertman, 2007, Verburg, 2004a). This paper is an attempt to take one step ahead to answer these questions.

COMPONENTS OF NEIGHBOURHOOD INTERACTION

Tobler's so called first law of geography “Everything is related to everything else, but near things are more related than distant things” plays a significant role for understanding importance of spatial interaction in urban land-use dynamics (Barredo, 2003). Tobler's law stresses that the neighbourhood of a feature, also beyond the most adjacent space, can influence the feature as a function of distance.

Distance decay functions that are based on empirical studies describe the influence a factor has on another factor over distance. The functions can be integrated in land-use models as so called transition rules. Transition rules that are based on neighbourhood interaction are referred to as neighbourhood rules. Seven generalised types of neighbourhood rule shapes have been identified (Figure 1). The seventh, not shown in the figure, represents a situation where no spatial relationship between two land-use classes can been identified.

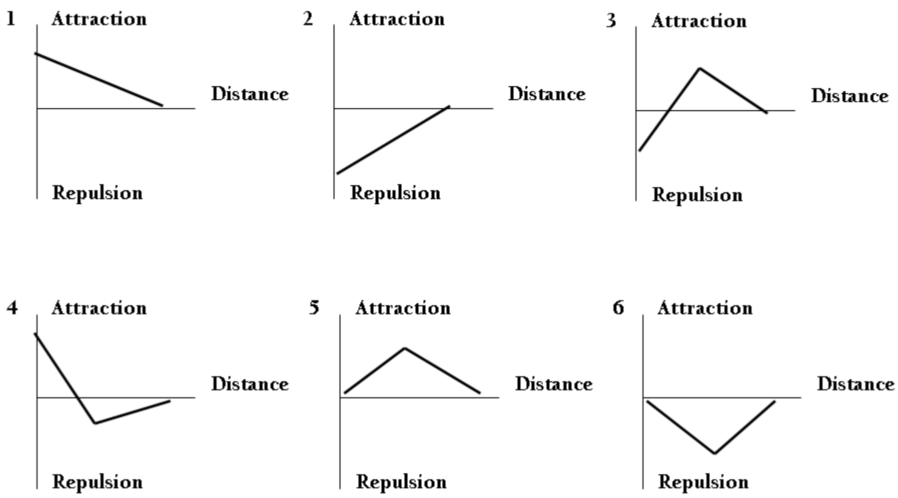

Figure 1: A neighbourhood rule shape typology modified after Haagort (2008).

One way to quantify neighbourhood rules is to analyse changes in land-use with so-called spatial metrics. Spatial or landscape metrics have been used for a couple of decades in landscape ecology to describe spatial characteristics, such as the size, shape, number, kind, and configuration of landscape structure, and more recently also outside that field (Geertman, 2007). Spatial metrics cannot explain the causes of observed land-use patterns and processes, but they can give scientific indications of causal relationships taking place (Geertman, 2007) and this is what is needed in order to justify the use of neighbourhood rules in urban CA.

Several authors have found a spatial metric – the so-called mean enrichment factor – appropriate for quantifying and analysing neighbourhood characteristics and for generating neighbourhood rules (Verburg, 2004a, Geertman, 2007, Hansen, 2008). The enrichment factor characterises the over- or under-representation of different land-use types in a neighbourhood of a specific raster cell. To measure this over- or under-representation, it compares the amount of cells of a particular land-use type in a pre-defined neighbourhood of a specific location in relation to the amount of cells of that land-use type in the study area in total. When the proportion of a land-use type in a neighbourhood equals the average, the neighbourhood has an enrichment factor of 1 for that land-use type. If the neighbourhood of a specific location (cell) consists of 20% residential areas, whereas the proportion of residential areas in the study area as a whole in total is 5%, we can characterise the neighbourhood by an enrichment factor of 4 for residential areas. Contrary an under-representation of a certain land-use type in the neighbourhood will result in an enrichment factor between 0 and 1.

The equations for the enrichment factor are specified in Verburg (2004a) as the following:

$$F_{i,k,d} = \frac{n_{k,d,i} / n_{d,i}}{N_k / N}, \tag{2.1}$$

where $F_{i,k,d}$ characterises the enrichment of neighbourhood *d* of location *i* with land-use type *k*. The shape of the neighbourhood and distance of the neighbourhood from the central grid cell *i* are identified by *d*. This grid cell *i* in a series of enrichment factors for the different land use types (*k*). The procedure is repeated for different neighbourhoods located at different distances (*d*) from the grid cell to study the influence of distance on the relation between land-use types.

The average neighbourhood factor for a particular land-use type $\overline{F}_{l,k,d}$ is calculated by taking the mean of the enrichment factors for all raster cells belonging to a certain land-use type *l*, following:

$$\overline{F}_{l,k,d} = \frac{1}{N} \sum_{i \in L} F_{i,k,d} \tag{2.2}$$

where *L* is the set of all locations with land-use type *l*, and N_l is the total number of cells belonging land-use class *l* in the study area.

To conclude, it seems according to literature, that if we can make a tool that derives enrichment factors, we can empirically capture and quantify the neighbourhood interaction that has taken place in our study area. The calculated results can help to arrive at theoretically and empirically more justifiable neighbourhood rules, which in their turn can help to improve urban CA models.

METHOD

We developed a programme for quantifying the neighbourhood interaction in our study area using the enrichment factor. The programme was implemented according to a conceptual model (Figure 2) in Python GNU programming environment using the functionalities of ArcGIS Spatial Analyst extension. The programme starts by finding the land use cells that has changed between "Land Use t_0" and "Land Use t_1" (Calculate Changes). The "Changed Land Use" is then classified according to the type of change we are interested to look deeper into (Reclassify). In the case of urban sprawl and expansion, it is most interesting to analyse the cells where new urban cells, for example "New Residential" cells, have emerged.

In the following, we enter a loop, where the amount of a particular cell-type within a user-specified neighbourhood is calculated (Calculate Focal Sum) with the FocalSum function of ArcGIS. In our case we are looking at the amount of residential cells in the neighbourhood of a new residential cell, i.e. "Residential Residential". We are not interested in the FocalSum values of all cells, but only in the values of those cells that are new residential cells in "Land Use t_1". The FocalSum values for the new residential cells are calculated based on the amount of residential cells within a pre-defined neighbourhood based on "Land Use t_0"; in this way we get a measure on the attracting or repelling effect of the land-use classes of t_0.

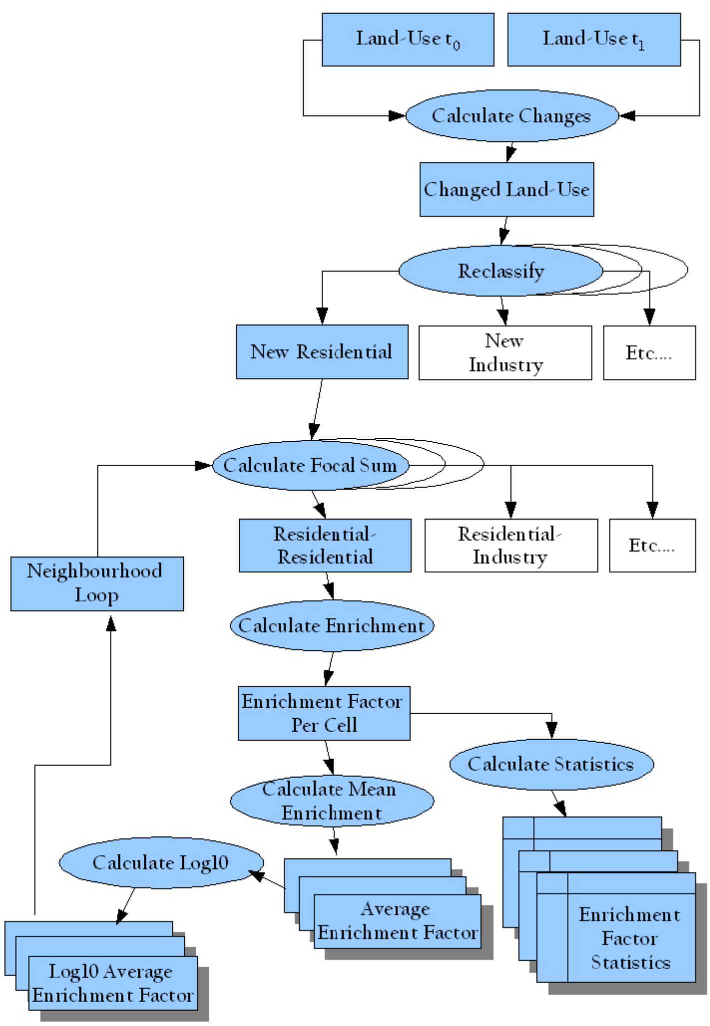

Figure 2: A conceptual model of the programme structure.

Next, the neighbourhood-based enrichment factor is calculated on a cell-to-cell basis (Calculate Enrichment). Based on the neighbourhood enrichment data a general statistics table, "Enrichment Factor Statistics" (Calculate Statistics) and a mean enrichment factor raster, "Average Enrichment

Factor, are produced (Calculate Mean Enrichment) using ArcGIS zonal functions. The local base 10 logarithm is then calculated from the "Average Enrichment Factor" (Calculate log10). The programme loops trough all user-specified neighbourhoods and make the described output tables and raster images for all neighbourhoods.

STUDY AREA AND DATA

The programme was run to get an indication of the neighbourhood interaction taking place in our study area in South-Western Finland. The study area covered eleven municipalities; Turku, Kaarina, Lieto, Masku, Merimasku, Naantali, Parainen, Piikkiö, Raisio, Rusko, Rymättylä (Figure 3), in all an area of 1913 square kilometres.

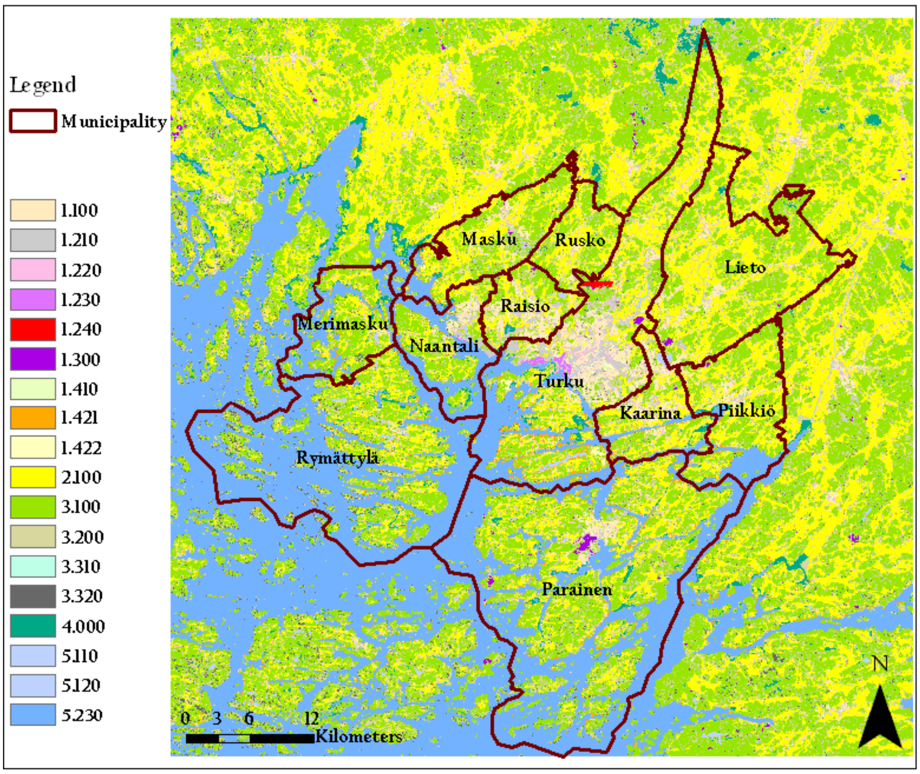

Figure 3: A map of the study area in South-Western Finland. The codes of the land-use classes are explained in table 1.

Data sets representing the land-use in year 1990 and in year 2000 were available. These land-use data sets have been produced according to the principles and classification of the European CORINE Land Cover project. CORINE LC is produced differently for Finland than in most other European countries. The production is based on a combined method consisting of automated interpretation of satellite images and the integration of existing digital map data. In addition to the European CORINE data set a national, less generalized data set have been made and it is the national version of CORINE

2000 that is used in this project. The national data set have a spatial resolution of 25 x 25 m pixels in comparison with the European vector data, where the minimum mapping unit for most features is 25 hectares (CLC, 2000).

Land-Use class in CORINE	Used land-use class	Description
1110 and 1120	1100	Residential
1210	1210	Industry and service
1220	1220	Traffic areas
1230	1230	Port areas
1240	1240	Airports
1310 - 1330	1300	Extraction and dump sites
1410	1410	Green urban areas
1421	1421	Summer cottages
1422	1422	Sports facilities
2110 - 2430	2100	Arable and semi-natural areas
3111-3133	3100	Forest
3241 -3246	3200	Grasslands and shrubs
3310	3310	Beaches, dunes and sand plains
3320	3320	Bare rock
4111-4212	4000	Mire
5110	5110	Lake
5120	5120	River
5230	5230	Sea

Table 1: The land-use classes used versus the land-use classes of the national CORINE.

CORINE is a land-cover data, not an urban land-use data set, and data that represent the urban land-use should preferably include only relevant land-use classes to explain the processes and spatial interaction of interest. For instance, it is not relevant in respect to urban land-use modelling purposes and for this project, to know the exact type of forest. Rather it is relevant to know if there is forest or not, to find out if the proximity of forests have an attractive or repulsive effect on urban land-use. A need for reclassification was also supported by the indications that imply that the data quality improves by using a broader thematic classification (CLC, 2000). A reclassification of the original land-use data was made, with ArcGIS *Reclass by ASCII file* function, to respond to these issues. The used classification is presented in table 1.

During the time period, the residential area has been (1100) the most expansive of the active class. In ten years the amount of residential cells has increased from 186322 pixels to 248893 pixels in our 25 m pixel data. The expansion has taken place on the expense of the arable and semi-natural areas, the forest and the grasslands and shrubs classes.

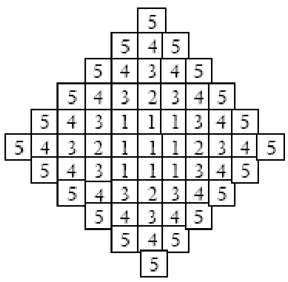

Figure 4: The configuration and extent of the neighbourhoods used in this study.

We calculated the enrichment factor for new residential areas, both in relation to the effect of existing residential cells and in relation to existing summer cottages. Five neighbourhoods were used (Figure 4). The neighbourhoods are based on aggregation, so that the first neighbourhood (N1) contains the cells with number 1, the second neighbourhood (N2) contains the cells of number 1 and number 2, and so on.

RESULTS AND DISCUSSION

All results presented in earlier studies focus on the average enrichment factor and its base 10 logarithm. However, since a cell-to-cell based enrichment factor raster is generated using this programme, we take a look at these raster images first (Table 2). According to our results some summer cottages are very over-representated in the proximity of some of the new residential cells. The highest over-representation values of the summer cottages are over three times higher than the over-representation values of residential cells. This is related to the relatively low share of the summer cottages, 1.71%, in the whole study area. The share of residential cells is 6.09 %. These values influences the calculations, so that a neighbourhood that include summer cottages get a higher value of over-representation, than if the same neighbourhood would contain the same amount of residential cells. This sensitivity to the total share of a land-use class is a drawback of the enrichment factor that is good to be aware of.

Neighbourhood	N1	N2	N3	N4	N5
Enrichment Factor: Residential-Residential	0-14.59	0-15.16	0-15.77	0-16.02	0-16.15
Enrichment Factor: Residential-Summer Cottage	0-51.98	0-49.52	0-46.83	0-42.79	0-43.14

Table 2: The variation of the enrichment factor on a cell-to-cell basis.

The mean or average enrichment factor is lower for summer cottages than for residential areas for the zone of new residential cells (Table 3). There are relatively few high enrichment factor values in the cell-to-cell based data set. The broader variation of enrichment factor values can also be seen as higher standard deviation (std) values for summer cottage areas. Despite the higher range of values and higher variation of enrichment values, the average enrichment factor seems to describe the overall situation well: in the neighbourhoods of new residential cells there where an over-representation of residential cells in 1990. On the contrary, summer cottages where under-represented or as present as in the study area on average in the proximity of cells that have turned into residential cells. From these results we can conclude that the residential cells are attracting new residential cells with a bigger effect than summer cottage cells. The figures of table 3 also seem logical in that sense that it is more likely that a cell will turn into a residential cell more than 50 meters away from existing residential or summer cottage cells, than closer . It is not needed to build buildings as close as 25 meters to each other, in a country like Finland, where there is enough space for everyone.

Neighbourhood	N1	N2	N3	N4	N5
Average Enrichment Factor(std): Residential-Residential	2,37879 (3.55)	2,50902 (3.37)	2,69285 (3.15)	2,73605 (3.00)	2,7045 (2.87)
Average Enrichment Factor(std): Residential-Summer cottage	0,867761 (4.36)	0,979208 (4.31)	1,16746 (4.24)	1,28651 (4.10)	1,34008 (3.91)

Table 3: The average enrichment factor for new residential cells in relation to existing residential and summer cottage cells in year 1990. The standard division is also included.

We had, however, expected that the mean enrichment and the base 10 log values to have higher values. Verburg (2004b) obtained the mean enrichment value of 7.9 in the Moore neighbourhood of residential cells in his study area when using land-use data from year 1989 and 1996. However, compared with the Netherlands, the building activities in Finland are in general more dispersed and less controlled, explaining at least part of the difference. There are also another area-specific reason for the lower values. When we compared the land-use datasets visually, we could observe that new residential buildings are generally built in the vicinity of existing ones. However, if there are only a few residential cells in the neighbourhood, as is the situation in many cases on the countryside in the outskirts of highly urbanized areas and where urban sprawl has taken place within the last ten years (Figure 5), the result will not indicate over-representation, which evens out the overall average enrichment factor.

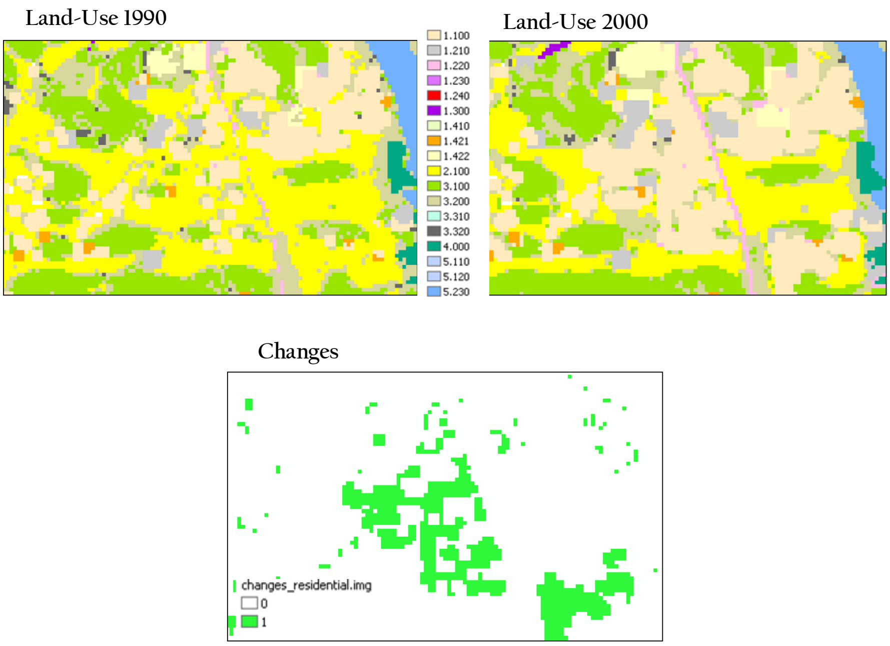

Figure 5: The location of the new residential cells (Changes) in relation to land-use in 1990 and 2000.

In some Northern European studies, the base 10 logarithm average enrichment values have been found to be as high as 0.9 to 1.3 (Verburg, 2004b, Hansen, 2008). In a study carried out by Geertman (2007) the over-representation near new residential areas for the period 1986-1993 were found to have values between 0.7 and 0.3, depending on the characteristics of the area observed. The latter value is similar to the values obtained in this analysis (Table 4). There have been no results published on the mean enrichment factor for summer cottage areas, so we cannot make any comparison with results of others regarding them.

It is also likely that the values of the enrichment factors have been evened out due to the broad temporal resolution of our data. Our land-use data has a time gap of 10 years; it may in many cases seem that the cells surrounding new residential cells are arable land or forest cells, even though in reality the process of urbanisation has continued in a certain direction cell by cell (Figure 5); that is the neighbours of an emerged residential cell more likely to be residential cells in reality than is the case in our data. To capture these yearly, subsequent changes, would require data of a better temporal resolution (Hansen 2008), so to some extent the low mean enrichment values of residential cells were also expected as a result of the land-use data used.

Neighbourhood	N1	N2	N3	N4	N5
LogAverage Enrichment Factor: Residential-Residential	0,376356	0,399505	0,430212	0,437124	0,43087
LogAverage Enrichment Factor: Residential-Summer cottage	-0,0616	-0,00913	0,067241	0,190941	0,12713

Table 4: The base 10 log average enrichment factor of new residential cells in relation to residential cells and summer cottage cells in the 1990 data.

The neighbourhood configuration also plays a role in comparisons to the results of. For instance, in the study carried out by Verburg (2004b), square-shaped basic Moore neighbourhoods and extended Moore neighbourhoods were used. In this way, a lot more cells where included in the calculations than in our study. For example for the fifth neighbourhood of an extended Moore neighbourhood contains 121 cells, while our corresponding neighbourhood includes only 61 cells.

The data resolution used in this study also explains part of the deviation from previous results. The use of larger 500 meter cells in comparison with 25 meter cells in this study, also explain part of the difference. In our study no single standing houses were eliminated through aggregation or resampling techniques. In comparison, the data used in the Dutch study, was aggregated from 25 m to 500 m cells according to the majority rule, diminishing the role of dispersedly situated settlements.

CONCLUSION AND PROSPECTS

In this paper, we have recognized that neighbourhood interaction plays a central in land-use change and that we can improve the transition rules of CA-based land-use models by quantifying the rules of prevailing neighbourhood interaction. Literature indicates that a spatial metric, the mean enrichment factor, is an appropriate measure of neighbourhood interaction, and that was used to quantify neighbourhood interaction. A desktop based programme was developed to carry out the calculations.

Our results show that there are less residential areas within the neighbourhoods of new residential areas in Finland than in other parts of Northern Europe. This indicates that the neighbourhood interaction is comparatively low; existing residential areas does not seem to attract new residential in the same manner as in other Northern European countries. In subsequent work it would be interesting to further look into how big a share of the comparatively low enrichment values of the study area are due to data, and how big a share depends on the tradition to build scarcely in Finland. To study this, we would need to make run our script using land-use data of subsequent years and also of a broader scale, to find out how sensitive our present results are with change in resolution and in order to make our results more comparative with scales used by others.

According to literature (Verburg, 2004b), it is possible to use the calculated enrichment factors to define neighbourhood rules. Before going into that procedure, it would be good to test the sensitivity of our results to changes in resolution and neighbourhood configuration. Furthermore, we need to run our programme to find out the neighbourhood interaction of between other classes of our land-use data to get the whole picture of the attracting and repelling effects taking place within our study area.

BIBLIOGRAPHY

Barredo, J.L., Kasanko, M., McCormick, N. and Lavalle, C., 2003. Modelling dynamic spatial processes: Simulation of urban future scenarios through cellular automata. *Landscape and Urban Planning*, vol 64:145-160. p 145.

Batty, Michael, 2005. *Cities and Complexity. Understanding cities with Cellular Automata, Agent-Based Models, and Fractals*. The MIT Press, Cambridge, Massachussets, London, England. pp preface, 68, 73.

CLC, 2000. *CLC2000-Finland*. Final Report. Finnish Environment Institute (SYKE), Geoinformatics and Land Use Division (GEO), May 2005. pp 3, 41, 43. Retrieved from http://www.ymparisto.fi/download.asp?contentid=38725&lan=en on May 20th 2008.

CLC, 1990. *Production of Land Cover and Change Information for ENVIFACILITATE-project.* Geoinformatics and Land Use Division, Finnish Environment Institute.

Engelen, G., White, R. and Uljee, I., 2002. *The MURBANDY aqnd MOLAND models for Dublin.* Final Report. Research Institute for Knowledge Systems BV. pp 5-9, 23-25. Retrieved on May 20th 2008 from http://www.geo.ucl.ac.be/LUCC/MODLUC_Course/PDF/G.%20Engelen.pdf

Geertman, S., Hagoort, M. and Ottens, H., 2007. Spatial-temporal specific neighbourhood rules for cellular automata land-use modelling. International *Journal of Geographical Information Science* vol *2*1(5):547–568. Taylor & Francis. pp 549-552, 554, 559

Hagoort, M., Geetman, S.,Ottens, H., 2008. Spatial externlities, neighbourhood rules and CA land-use modelling. *Ann Reg Sci* vol 42:39-56. Springer. pp 40-45.

Hansen, H. S., 2008. Quantifying and Analysising Neighbourhood Characteristics Supporting Urban Land-Use Modelling. In: *Lecture Notes in Geoinformation and Cartography.* The European Information Society. Springer. pp 293-298.

Malcewski, J., 2000. On the Use of Weighted Linear Combination Method in GIS: Common and Best Practice Approaches. *Transactions in GIS*, vol 4(1): 5-22. Blackwell Publishers. pp 16-19, 21. 71 72

O'Sullivan, David & Torrens, Paul M. (2000). Cellular Models of Urban Systems. *Working Paper Series*. Paper 22. Centre for Advanced Spatial Analysis, University College London. pp 1-2.

Portugali, Juval, 1999. Self-Organization and the City. Springer. pp 49

Verburg, P.H., van Ritsema, E. J., de Nijs, T. C. M., Dijst, M. J. and Schot, P., 2004a. Determinants of land-use change patterns in the Netherlands. *Environment and Planning B: Planning and Design* vol 31:125-150. Great Britain. pp 125-126, 146.

Verburg, P. H., de Nijs T. C. M., van Ritsema, E. J., Visser, H. and de Jong K., 2004b. A method to analyse neighbourhood characteristics of land use patterns. Computers, *Environment and Urban Systems* 28:667-690. pp 668-672, 674, 676, 677-680, 685, 687.