# ANALYSIS OF THE FACTORS THAT CONDITION THE EXPANSION OF A COASTAL VILLAGE THROUGH CELLULAR AUTOMATA SIMULATION

García, A.M.

Santé, I.

Crecente, R.

López, F.J.

*University of Santiago de Compostela*

## Abstract

Cellular automata, conceived by the mathematician Stanislaw Ulam and the physicist John von Neumann, have proven to be good tools for implementing models to predict and analyze land use change and urban growth dynamics. The implementation of such tools requires knowing the singularities of the processes to be simulated. Most of the models implemented so far have been concerned with the expansion of big cities but little has been done in relation to small urban areas. In this paper, we analyze the variables that govern urban growth in Ribadeo, a town of about 9000 people in northeastern Galicia, Spain, using logistic regressions and a simple cellular automata model that will allow us to study the factors that must be considered when performing simulations with these tools in small urban areas.

***Keywords:*** *Cellular automata, urban growth, variables.*

## 1. Introduction

### 1.1. Urban simulation background

In the late 1950s and early 1960s, a stage of sociopolitical development arose, which resulted in the growth of big cities. Such urban growth led to the appearance of new problems in urban areas, mainly related with transport. The need to scientifically analyze these new problems led to the development of the first models for the study of traffic. Later, land use allocation models were developed that were eventually coupled with the former, constituting integrated models.

In the 1970s, new growth dynamics emerged due to the rapid development of urban areas. These new processes were more complex than the processes that inspired the creation of integrated models. To make the study of such a complexity more accessible, the models were divided into subsystems, avoiding the establishment of too many relations between them. This led to a reduction in the capability of the models to simulate responses to changes in the dynamics. Models became complex, they needed a great amount of input data and provided a very general view that was not really useful to planners. All this led to the need to use new modeling methods that could overcome such limitations.

With the development of computer science in the 1970s and the consequent increase in the computing power of computers, using new methods became possible. In 1970, Waldo Tobler developed a new population growth simulation model based on cellular automata (CA). With the advances in computers in the 1980s, the interest in the study of cellular automata increased and Helen Couclelis established the theoretical basis for the application of such paradigm to the simulation of urban growth (Berling-Wolff and Wu, 2004).

### 1.2. Cellular automata

Cellular automata were developed by the mathematician Stanislaw Ulam and the physicist John von Neumann to study self-reproduction and the modeling of biological processes. In the 1970s, the English mathematician John Hotton Conway created “The game of life”, the first application of a cellular automaton aimed at reproducing the growth of an animal population. The model was composed of a grid in which each cell could adopt two states; “alive” or “dead”, according to some transition rules that determined that:

- If a grid cell was dead and surrounded by 3 living cells, the cell was born.
- If a live cell was surrounded by 2 or 3 live cells, the cell kept living.
- If a live cell was surrounded by less than 2 live cells, the cell died of loneliness and if it was surrounded by more than 3 live cells it died of competence.

These rules were applied at discrete time intervals, such that at each iteration of the model all the cells were updated simultaneously.

The game of life is a formal cellular automaton. Therefore, CA are composed of:

- A grid or raster space.
- A group of states that define the cells of the grid.
- A definition of the neighborhood of each cell.
- A series of transition rules that define the state of each cell based on the states of the neighboring cells.

If the game of life is iterated several times, complex patterns emerge. This complexity arising from underlying simplicity makes CA a powerful tool for the simulation of urban growth because of their emergent behavior, given that CA can exhibit complex global dynamics arising from simple rules on a local scale.

Cellular automata are able to reproduce growth patterns with fractal geometry, which can be found in most cities. Because CA work on a grid, they facilitate the use of raster data obtained with a geographical information system (GIS) or remote sensing techniques.

The simplicity of cellular automata and their ability to simulate urban growth allow us to easily analyze the processes simulated by them. For this reason, CA are useful tools for research. A number of authors (Sui and Zeng, 2001; Wu and Martin, 2002; Cheng and Masser, 2004; Aguilera, 2006) have studied the processes that control the growth of large cities using cellular automata. Almost all the existing studies tackle the growth of large urban areas, but only a few models have been developed to study the dynamics of small towns. In this work, CA are used to analyze urban growth in Ribadeo, a coastal municipality on the border between Galicia and Asturias, Northwest Spain.

Earlier in this section, we defined a formal CA using the Conway's game of life and its basic rules as an example. Formal cellular automata do not have the abilities required to correctly simulate urban growth processes. Consequently, some relaxations to the formal rule should be adopted in order to obtain more accurate results (Verburg et al., 2004; White, 1997). To design a model that can produce accurate results, the variables and processes that influence urban growth processes and interactions must be known.

In this paper, we have performed an analysis based on the research by Sui and Zeng (2001), Wu and Martin (2002), Cheng and Masser (2004) and Aguilera (2006) in order to identify the variables that may condition urban development using logistic regression techniques. Such an analysis allowed us to select the most important variables to include in a CA based model and to determine its capacity to simulate urban growth. Results were analyzed and some conclusions were drawn from the results.

### 1.3. Methods

Ribadeo is a municipality of 9619 people located in the province of Lugo, Spain, on the border between Galicia and Asturias. The urban core of Ribadeo constitutes the administrative and service center of the area.

In 1987, the *Puente de los Santos* Bridge was built over river Eo estuary, communicating the north of the municipality with Asturias. This infrastructure brought added value to Ribadeo and the village became the service center of the Asturian municipalities on the other shore of the estuary. The construction of the bridge fostered the growth of population and commerce. New infrastructures like the Transcantabrian motorway will attract more growth to the area.

To perform the present analysis, photointerpretation techniques were used to obtain land use maps of Ribadeo and its four surrounding parishes for the years 1995 and 2003. These data were converted to raster format with a pixel size of 25x25 m.

Later, we considered the variables that could be used according to the available data. After having analyzed other models, we decided to consider the following variables:

- Accessibility (distance to roads).
- Height.
- Distance to forested land.
- Distance to main urban core.
- Distance to peripheral urban cores.
- Distance to the shore.
- Distance to main roads.
- Distance to secondary roads.
- Distance to other roads.
- Distance to railways.
- Distance to railway station.

- Distance to port.
- Distance to residential land uses.
- Distance to commercial land uses.
- Distance to industrial land uses.
- Distance to parks and green areas.
- Distance to institutional land uses (educational centers, cultural centers, administrative buildings, public sport facilities, welfare and health centers and churchs).
- Orientation.
- Slope.
- 300 m radius neighborhood. To calculate the neighborhood, we have considered a square region around each central cell with a cell radius equivalent to a distance of 300m measured between the edge and the center of the area. In this area, we applied equation (1), where *I* is 0 if the cell is non-urban land and 1 if the cell is urban land; *w* is a coefficient that weights the value according to the distance of each cell in the neighborhood to the central cell, computed as follows: the neighborhood window is divided into several concentric squares, separated by a cell, *w* will have the same value in each square and it will decrease with the increase in distance to the central cell. The decrease in the value of coefficient *w* will be related to a line whose slope has the same value as the fractal dimension[2] of Ribadeo village.

$$N = \sum w \times I \qquad (1)$$

- Zoning (the current urban planning instrument in the municipality is the Local Development Plan 1977, which establishes only the zoning of Ribadeo urban core).
- Shape index. The shape index is a spatial index that defines a relation between the perimeter and the area of a patch on a map (equation 2). In the present study, the areas of built land were defined and the value of the shape index of each urban patch was assigned to the nearest non-urban pixels.

$$IF = perimeter / 4\sqrt{area} \qquad (2)$$

- Area of the cadastral parcels.
- Shape index (equation 2) of the cadastral parcels.

Following the method used by Aguilera-Benavente (2006), we used Idrisi software to estimate logistic regressions for each variable, using a binary map that considered the built up area between the years 1995 and 2003 as independent variable. Roads were not considered in these maps. A mask was used to exclude the following elements from the regressions: built up cells in 1995, water surfaces and the 100m coastal fringe protected by the Spanish Coastal Act. Such elements were excluded from regressions because the variables present in those zones

---

[2] The fractal dimension is determined considering ring areas concentric to the urban core center, so that the ratio that relates built up area and total area is calculated in each ring. Then a line is fitted to the values of the ratios obtained. The slope of the line is the fractal dimension of the urban area considered.

did not influence urban development during the period considered. In order to reduce computation time and spatial dependence between variables, not all the cells were used in the regressions. Systematic sampling was used to choose the cells that would be considered in the regressions.

Logistic regression considers the probability of a cell being urban and determines the relation between such a probability and the variables considered by fitting the following equation (3):

$$P(y=1\,|\,X)=\frac{\exp(\sum BX)}{1+\exp(\sum BX)} \qquad (3)$$

where:

- $P$ is the probability of the dependent variable being 1 (urban).
- $X$ are the independent variables. $X=(x_0, x_1, x_2 ..., x_k);\ x_0 = 1$
- $B$ are the estimated parameters. $B=(b_0, b_1, b_2 ..., b_k)$.

By linearizing the equation is linearized, the following expression is obtained (equation 4):

$$Ln(P/(1-P)) = b_0 * x_0 + b_1 * x_1 + b_2 * x_2 + ... + b_k * x_k + error \qquad (4)$$

This equation can be fitted by using a linear regression.

The results were analyzed by using two parameters:

- Pseudo $R^2$ indicates the degree of fit of the regression. If the value was greater than 0.2 the fit was considered to be good.
- ROC considers the degree of relation between the independent variables and the dependent variable. If the value was greater than 0.5 it was considered that there was a dependence.

- Once the variables were analyzed, those with the highest ROC were chosen for use in a multiple logistic regression model, thus allowing us to know the weight of each variable in urban growth.

Then, we used the weights obtained and Cheng and Masser (2004) model as a reference, and applied the following transition rule in the cellular automaton (equation 5):

$$P=(\sum ri \times Xi + n \times N)\times v \qquad (5)$$

Where:

- *P* is the probability of change.
- *ri* is the coefficient of each variable *i* obtained by logistic regression.
- *Xi* is the value of variable *i*.
- *N* is the neighborhood.
- *n* is the neighborhood coefficient obtained by logistic regression.
- $v$ is the stochastic variable, which introduces randomnesss in the system and is calculated from equation (6):

$$RA = 1 + (-\ln \gamma)^{\alpha} \qquad (6)$$

Where γ is a random number between 0 and 1 and α is a coefficient that controls the degree of randomness introduced in the model. The value of α used in the model was equal to the fractal dimension of Ribadeo (Aguilera, 2006).

The urban area map for 1995 was used as a starting point, and the model was iterated a number of times equivalent to the number of years between 1995 and 2003. At each iteration, the urban area was determined by dividing the built up area in the considered period by the number of years. At each iteration, the cells with the greatest probability of urbanization changed to urban use until the previously calculated area was reached.

| Variable | ROC |
|---|---|
| 300m radius neighborhood | 0.8463 |
| Distance to institutional land uses | 0.8162 |
| Distance to main urban core | 0.8049 |
| Distance to commercial land uses | 0.7996 |
| Distance to forested areas | 0.7966 |
| Distance to main roads | 0.7857 |
| Distance to railway stations | 0.7827 |
| Distance to residential land uses | 0.7773 |
| Distance to port | 0.7734 |
| Distance to industrial land uses | 0.749 |
| Accessibility | 0.7461 |
| Height | 0.736 |
| Distance to other roads | 0.7356 |
| Distance to parks and green areas | 0.7336 |
| Distance to peripheral urban cores | 0.7233 |
| Zoning | 0.6759 |
| Distance to the shore | 0.6719 |
| Shape index | 0.6667 |
| Distance to railways | 0.6557 |
| Distance to secondary roads | 0.6106 |
| Area of cadastral parcels | 0.6057 |
| Shape index of cadastral parcels | 0.5864 |
| Orientation | 0.3793 |
| Slopes | 0.0401 |

Table 1. Variables chosen to perform the simulation.

The variables with the highest ROC were chosen (Table 2).

From among the variables included in Table 3, distance to main urban core, distance to railway station and distance to port were eliminated because of their correlation with distance to institutional land uses. All these elements are close to one another and near Ribadeo urban core. The variable 'distance to institutional land uses' was not removed because it showed the highest ROC. In addition, distance to commercial land uses, distance to industrial land uses and distance to residential land uses were eliminated because they were correlated with neighborhood, whereas distance to main roads was removed because it was related to accessibility.

A new logistic regression was calculated to obtain the relative weight of each selected variable and to determine its influence on land change from non-urban to urban (Table 3).

| Variables | Coefficient |
|---|---|
| Error parameter | -2.06422057 |
| Neighborhood | 2.15979110 |
| Accessibility | -24.39321381 |
| Height | -3.63615921 |
| Distance to forested areas | 1.34067087 |
| Distance to institutional land uses | -3.53720871 |

Table 2. Coefficients obtained for each variable by multiple logistic regression. Negative coefficients suggest that the higher the value of the variable, the lower its influence on urban growth, whereas positive values indicate the opposite relationship.

- ROC = 0.8760
- Pseudo $R^2$ = 0.2260

Because the value of pseudo $R^2$ is greater than 0.2, it can be considered that the fit is good. The value of ROC is also high.

In table 3, we can observe that distance to roads plays a key role in the urban growth process. The fact that neighborhood does not have as much weight in this model as in other revised models may be due to the sprawl of growth during the study period.

The coefficients of the variables were rounded and applied in equation 4. After 8 iterations, results were crossed with 2003 data and analyzed (figure 1).

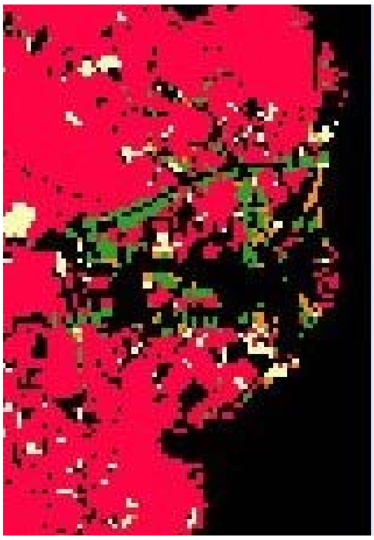

Figure 1. Comparison between real data and simulation results with a cell size of 30x30 m. Red cells: non-urban land, black cells: urban land in 1995, yellow cells: urban land in 2003 unpredicted by the model, green cells: wrong predictions, orange cells: correct predictions of the model.

With this model, we did not intend to make an accurate simulation of reality, but to observe the growth patterns obtained in order to be able to analyze the growth dynamics that generates such patterns and draw some conclusions about which elements should be considered when implementing models in similar areas. This is the reason why visual analysis of data is considered more important than statistical analysis.

The results suggest that the ring road north of Ribadeo attracts a lot of growth. However, in reality, this road gives access to *Puente de los Santos* and is fenced. Therefore, there is no direct access to it, such that it does not attract growth. For this reason, the model was run another time without considering the ring road in accessibility maps.

Removing the ring road modified one of the variables, such that its weight in the model varied. Consequently, the logistic regression was performed a second time to calculate the new coefficients. The following results were obtained (Table 4):

| Variables | Coefficient |
|---|---|
| Error term | -2.00074547 |
| Distance to institutional land uses | -4.35845586 |
| Height | -2.71871732 |
| Distance to forested areas | 2.35060711 |
| Accessibility | -29.20497587 |
| Neighborhood | 1.09300021 |

Table 3. Coefficients of each variable calculated by logistic regression, using the variables selected for the model. In this simulation, the ring road has not been considered.

After the model was run a second time and the results were crossed with the urban land map for the year 2003, we obtained the following results (fig. 2):

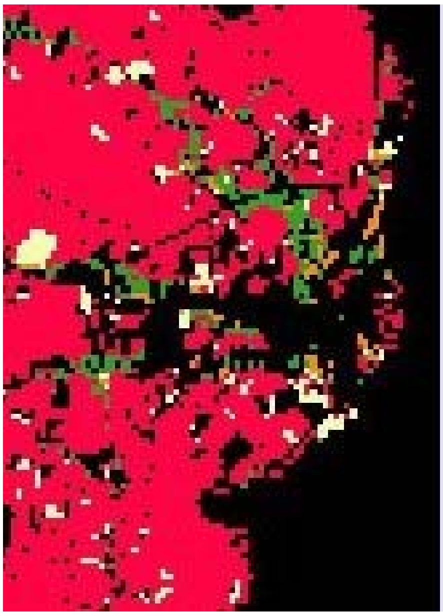

Figure 2. Comparison between real data and simulation results with a cell size of 30x30 m, not considering the ring road. Red cells: non-urban land, black cells: urban land in 1995, yellow cells: urban land in 2003 unpredicted by the model, green cells: wrong predictions, orange cells: correctly predicted cells.

The results of the simulation have improved. For example, the growth that appeared in the small villages along the provincial road heading northwest and inside the urban core is approximate to the actual core growth trend.

In addition, we can observe unpredicted growth to the southeast and southwest of the urban core. Maybe these areas were favored because their slopes are oriented to the estuary. In the first case, the location of the parcels near the urban center and the sea could have favored the urbanization of these parcels. In the second case, the parcels are near Ribadeo and they are well communicated with the village.

These factors should be considered in further analysis, including the visual basins and the proximity to the shore in the analyzed variables, despite the fact that the analyzed variables show a lower ROC than the variables used. The results of the model could also be improved by considering different land uses in the simulation in addition to urban and non-urban land. Thus, more detailed models could be implemented, which could better reproduce the dynamics and avoid the mistakes derived from assigning the same growth attraction power to all urban land uses or by considering that the variables influence the location of the different land uses in the same way.

## 2. Conclusions

The results of this work have allowed us to draw some conclusions about the future development and application of more suitable CA based models for the simulation of the singularities of the Galician urban system.

Logistic regressions show that one of the variables with the highest weight in the model is distance to roads. After having run the model, we can observe some distortions derived from considering that all roads attract urban growth in the same way. To improve the results, different road proximity variables based on the kind of road should be used. In the case of ring roads and highways, the distance to access points could be used as a variable.

Neighborhood is the variable that most influences urban growth, but its relative importance with respect to other variables inside the model is low. In the neighborhood used in this study, only the proximity to urban land has been considered and no difference has been made between the different land uses that constitute urban land. If we analyze variables such as proximity to industrial, commercial or residential areas, we observe that they are closely related to growth. However, such variables have not been used because we have assumed that they were represented in the neighborhood. Models will be more accurate if the influence of different land uses on the neighborhood is considered.

All these conclusions point to the need to develop more complex models that better simulate growth in small urban areas.

Because most of the CA models used for urban simulation have been applied to the analysis of large cities, they use a coarser spatial scale than the scale used in this work. Besides, a relatively small time scale has been considered because meaningful growth cannot be appreciated in an 8-year period. These two issues hamper the application of CA models to the simulation of growth in characteristic Galician urban areas.

To overcome this problem, wider temporal and spatial scales should be used when designing cellular automata for predicting the growth of small urban areas, such that the patterns are well defined and allow for simpler analysis and identification. Thus, we will be able to make simulations with simpler models that can be better analyzed.

However, we must not forget that by increasing the time scale we may cover different periods in which growth may depend on different variables and, consequently, the model should be calibrated for each of these periods. The same applies to the use of small scales; since the geographical area covered is greater, there will be more heterogeneity in the processes and variables that condition urban growth. For this reason, it is necessary to consider a balance, defining more or less homogeneous areas and considering temporal spaces where changes in urban trends have not occurred, such that the performance of the models is good and reliable predictions can be made.

## References

Aguilera Benavente F., 2006. Predicción del crecimiento urbano mediante sistemas de información geográfica y modelos basados en autómatas celulares, *Geofocus,* Vol. 6, pp. 81-112.

Berling-Wolff, S. and Wu J., 2004. Modeling urban landscape dynamics: A review, *Ecological Research,* Vol. 19**,** pp.119-129.

Cheng J. Q. and Masser I., 2004. Understanding spatial and temporal processes of urban growth: cellular automata modelling. *Environment and Planning B-Planning & Design,* Vol. 31, pp.167-194.

Sui, D. Z. and Zeng H., 2001. Modeling the dynamics of landscape structure in Asia's emerging desakota regions: a case study in Shenzhen. *Landscape and Urban Planning,* Vol. 53, pp.37-52.

Verburg P. H., Schot P., Dijst M. and Veldkamp A., 2004. Land use change modelling: current practice and research priorities. *Geojournal*, Vol.61, pp.209-324.

White R. Cities and cellular automata. *Discrete Dynamics in Nature and Society,* 1997, pp. 111-125.

Wu F. L. and Martin D., 2002. Urban expansion simulation of Southeast England using population surface modelling and cellular automata. *Environment and Planning A,* Vol. 34**,** pp.1855-1876.

**Correspondence** (for further information, please contact):

Andrés Manuel García Lamparte

Laboratorio do Território. Escola Politécnica Superior de Lugo. Pavillón II

Rúa Benigno Ledo s/n. Campus Universitario s/n. C.P. 27002 Lugo, Spain

Telephone number: +34 982 25 22 31 ext. 23292

Fax: +34 982 28 59 26

E-mail: andresmanuel.garcia@usc.es

URL: http://laborate.usc.es