Journal of Urban and Environmental Engineering, v.12, n.2, p.277-286

ISSN 1982-3932
doi: 10.4090/juee.2018.v12n2.277286

Journal of Urban and Environmental Engineering

www.journal-uee.org

MODELING SPATIAL DYNAMICS APPLIED TO THE ASSESSMENT OF URBAN GROWTH IN THE VICINITY OF PIPELINES IN RIO DE JANEIRO

Marcelo Elias Delaneze[1*] and Paulina Setti Riedel[2]

[1]*Universidade Estadual Paulista – UNESP, Post-Graduation in Geoscience and Environmental Studies, Av. 24A, 1515, 13.506-900, Rio Claro, São Paulo, Brazil, mdelaneze@yahoo.com.br.*

[2]*Universidade Estadual Paulista – UNESP, Department of Applied Geology, Av. 24A, 1515, 13.506-900, Rio Claro, São Paulo, Brazil, psriedel@rc.unesp.br*

Received 12 January 2018; received in revised form 16 June 2018; accepted 18 July 2018

Abstract: The use of mathematical models in environmental studies allows significant scientific contributions to the physical planning of an area, since they help to understand the impact of changes in land cover and predict future trends of changes in ecosystems. The study area includes portions of the municipalities of Duque de Caxias, Nova Iguaçu, and Belford Roxo, located in the state of Rio de Janeiro, Brazil, and aims to carry out experiments in land cover changes in the medium-term (10 years) within the catchment area of ORBEL pipelines. The modeling performed in this study was developed using the software Dinamica EGO. For model calibration, we used the weights of evidence method. Positive values foster certain transitions, while negative values indicate low probability of transition. Model validation was executed by means of the fuzzy similarity method using exponential decay. The results show that for the study period (1998-2010) the rate of change for the transition of the class 'others to woody vegetation' is much higher than for the other class transitions observed in the same periods. One annual land cover scenario was generated for the year 2020. The scenario indicated the growth of suburbs near the ring road around Rio de Janeiro, especially in the municipality of Duque de Caxias.

Keywords: Pipeline safety; statistical modeling; landscape dynamics; DINAMICA; oil pipelines

[*] Correspondence to: Marcelo Elias Delaneze. E-mail: mdelaneze@yahoo.com.br

INTRODUCTION

In 2016, the Brazilian pipeline infrastructure was composed of 610 ducts intended to carry petroleum, derivatives, natural gas and other products. The combined length of these ducts was 19 717 km, divided into 14 256 km for transport, and 5461 km for transfer. Extending 11 696 kilometers, 110 ducts were to carry natural gas. For derivatives, there were 429 ducts, totaling 5962 km. The other 32 ducts, 1985 km long, were intended to carry petroleum. The remaining 77 km, consisting of 39 ducts, were reserved to carry other products such as ethanol and solvents (ANP, 2017).

These ducts are essential for the energy supply chain in Brazil and are a vital connection between refineries, power plants and distribution centers.

Pipelines are an efficient and safe vehicle for the transportation of volatile or flammable substances that have the potential to cause damage to the neighboring population and the environment. The variety of terrain that the pipeline network goes through, such as densely populated areas, mountain regions and river crossings, make it more vulnerable to accidents. Therefore, obtaining information on the pipeline network status represents a fundamental role in the operation of pipelines, not only for the purpose of maintaining operational efficiency, but also to minimize the risks associated with possible accidents, like the disruption of the duct due to landslides or collapsing mass blocks, causing harm to humans and the environment.

Besides the risks associated with the physical environment, which could lead to disruption of the ducts, there is also the risk associated with human use and occupation, both resulting from urban and rural use, which implies buildings, streets, people and machine traffic around these installations.

Although pipelines are cited in literature as one of the safest means of carrying hazardous substances (Choi *et al.* 2003; Papadakis, 2000), considering the accidents associated with road or rail transportation, pipeline failures can occur and sometimes cause catastrophic consequences, therefore, the elaboration of studies regarding the prevention of accidents has become increasingly necessary.

The growing demand for energy and growth of metropolitan areas has resulted in more people living near pipelines. Thus, public safety in areas that cross over pipelines is a problem, and it may become more frequent, with population growth and the expansion of pipeline networks in the country, which is still relatively small, taking into account the size of the Brazilian territory.

Several factors are associated with urban growth: social, economic, transportation, communication, migration and public policies, among others. The study of urban growth factors and their driving forces is a complex activity, and requires sophisticated methods and tools (Aljoufie *et al.*, 2013).

In this context, modeling, especially if done in a spatially explicit form, is an important technique for the generation of alternative scenarios of future changes in land cover (Perez-Vega *et al.*, 2012) and therefore, will be an important contribution to the study of changes in the pipeline surroundings, which can cause substantial pressure on them.

Despite the flaws in spatial modeling in the past (Lee, 1973), in recent decades, increases in computational capacity, the increased availability of spatial data, and the need for innovative planning tools to help support decision making meant that there was a restoration of this type of modeling (Brail and Klosterman, 2001; Geertman & Stillwell, 2002).

These models include the development of new computational methods, including multivariate and logistic regression, exploratory spatial data analysis, spatial regression and hybrid geospatial approaches (Zeng *et al.*, 2015), micro-simulation, e.g. based on agents and cellular automata, which shows the potential for representing and simulating the complexity of the processes involved in the dynamic space and land use and land cover changes.

The dynamic simulation model of land cover used in this study was implemented in the software Dinamica EGO. Dinamica EGO is a model for simulation explicitly for spatial landscape dynamics, which is based on the paradigm of cellular automata. It brings together the functions of transition based on multi-scale neighborhoods, the concept of phases using a stochastic process of simulation in multi-steps, spatial feedback from calculations of dynamic variables, a component that directs the expansion of road networks, and application of logistic regression (Soares-Filho *et al.*, 2001, 2002) or weights of evidence (Almeida *et al.*, 2003) to calculate the probability maps of transition using the information stored in a GIS (Soares-Filho *et al.*, 2003).

The purpose of this paper is to simulate urban expansion, creating a scenario of land cover for the year 2020, for the selected range of ORBEL pipelines using weights of evidence to calibrate the model and cellular automata.

The study area and the ORBEL oil pipeline

The state of Rio de Janeiro is located in the southeastern region of Brazil and is the largest oil producer in the country, where the Campos Basin is located, one of the regions of greater oil production from deep waters in the world.

The study area is approximately 97 km^2, encompassing portions of the municipalities of Nova

Iguaçu, Belford Roxo and Duque de Caxias, located in the Metropolitan Region of Rio de Janeiro (**Fig. 1**). The

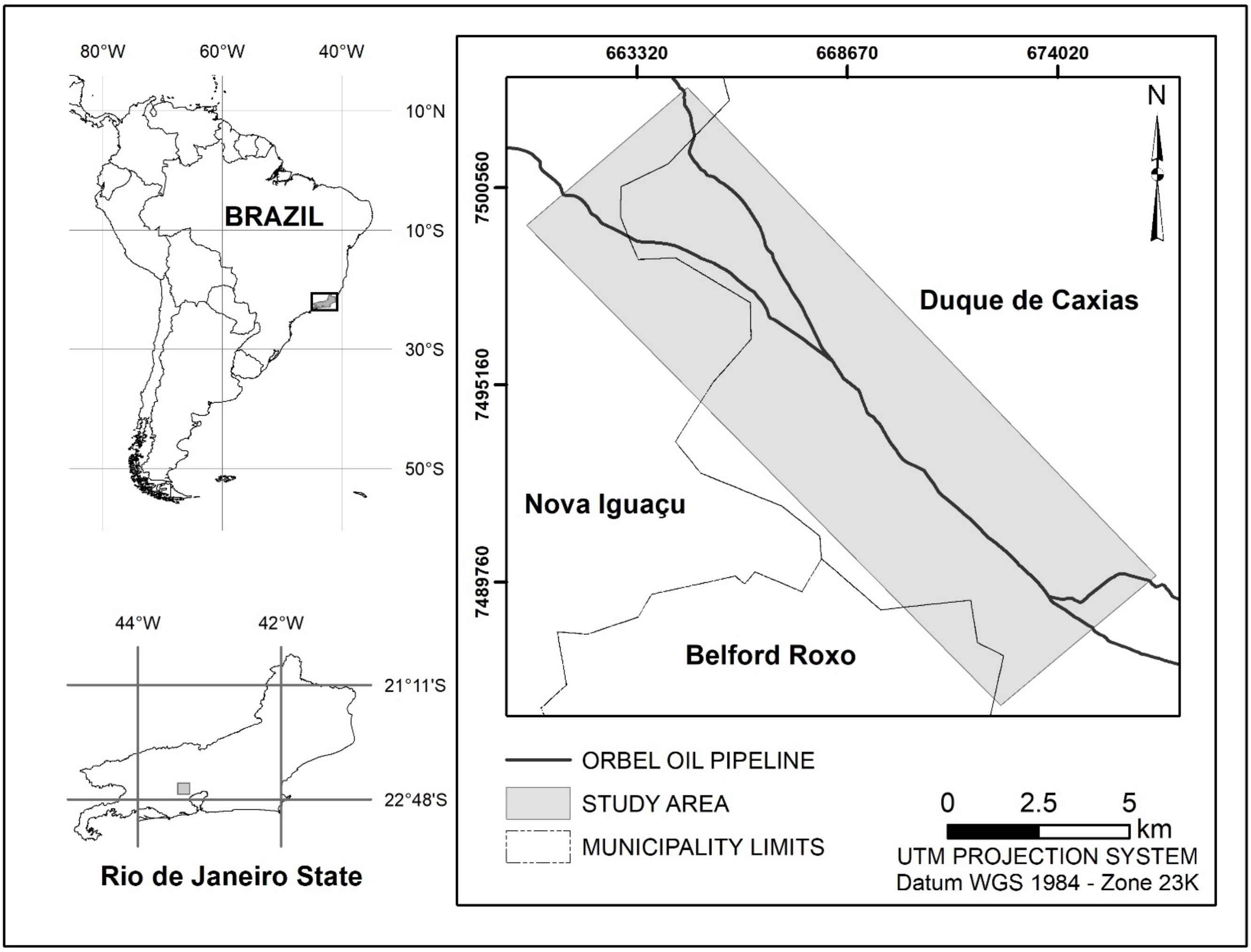

Fig. 1 Study area.

municipality of Belford Roxo, although not crossed by a pipeline, was also considered in this article because we consider both areas of influence on the pipeline: the area of indirect influence, which is real or potentially subject to indirect operational impact of the pipeline, including a 5 km strip of land on each side of the duct, and also the area of direct influence, which is that subject to direct operational impacts impact of the pipeline and includes a 400 m strip of land (or right of way) on each side of the duct.

The municipalities in the study area are among the most populous in Brazil and present socioeconomic dynamics accelerated by the presence of large industrial parks and proximity to the city of Rio de Janeiro.

The topography of the area varies between 1 and 326 m, with slopes ranging from mild to sharp (0° - 78°). In lithological terms, there is an occurrence of Fluviolacustrine deposits, Cretaceous/Tertiary alkaline rocks of the Tinguá Alkaline Suite and granitoids of the Serra dos Órgãos Unit and Santo Aleixo Unit.

The ORBEL pipeline first began operating in 1968, linking the Terminal Campos Elíseos or the refinery Duque de Caxias (REDUC) in Rio de Janeiro, to the Gabriel Passos refinery (REGAP) in the city of Betim - Minas Gerais state. This duct transports clear petroleum derivatives (gasoline, diesel oil, naphtha, cracked naphtha, and recycled light oil) with a capacity of 6,000 m^3/day. The ORBEL 1 pipeline has a diameter of 18 inches and a depth of approximately 1 - 1.5 m, it is 362 km long, crossing a total of 24 municipalities. In 2010, according to the IBGE (2010) the total population of these municipalities was 3 006 866 habitants, 55.8% in the state of Rio de Janeiro and 44.2% in the state of Minas Gerais.

Weights of evidence modeling

Calculation of transition rates

For the determination of the global rates of land cover change, throughout the whole simulation period, we use a cross-tabulation, whose output is called in Dinamica EGO unitary transition matrix. The transition matrix, which corresponds to the Markov chain, describes a system that changes in discrete intervals of time in which the value of any variable in a given period of time is the sum of the fixed percentages of values of all variables in the preceding time step (JRC & ESA, 1994).

Since the matrix elements are not negative, and the sum of elements in each line is equal to 1, each matrix element is called a probability vector, and the matrix P is a stochastic matrix or probability, according to Equation 1 (Judge & Swanson, 1962).

$$\begin{bmatrix} 1 \\ 2 \\ . \\ j \end{bmatrix}_{t=v} = \begin{bmatrix} P_{11} & P_{12} & P_{1.} & P_{1j} \\ P_{21} & P_{22} & P_{2.} & P_{2j} \\ P_{31} & P_{32} & P_{33} & P_{3j} \\ P_{j1} & P_{j2} & P_{j.} & P_{jj} \end{bmatrix}^{v} * \begin{bmatrix} 1 \\ 2 \\ . \\ j \end{bmatrix}_{t=o} \quad (1)$$

where the columns represent the probabilities of a determined state i remain in the same state or move to state j during the time interval of $t \rightarrow t + v$. Using the coverage maps from the start and end time of each period, matrixes were calculated in Dinamica EGO, of global and annual transition, and for the generation of this last one, the method proposed by Bell and Hinojosa (1977) was used, based on eigenvalues and eigenvectors of the global matrix, according to **Eq. 2**.

$$MT_{annual} = H.V^{\frac{1}{n}}.H^{-1} \quad (2)$$

where MT_{annual} is an annual matrix of land cover transitions, H are the eigenvectors of the global matrix of transitions, V are the eigenvalues of the global matrix of transitions, n is the annual number of steps within the total period examined, H^{-1} is the inverse matrix of eigenvectors of the global matrix of transitions.

Model Calibration

The calibration of the model aims at selecting the best set of variables that attempts to explain the changes in land cover that occurred in a given period. For this purpose, we use the method of weights of evidence, based on the theorem of conditional probability of Bayes (Bonham-Carter, 1994). This method uses the Bayesian probability model and is used to assess the mineral potential (Bonham-Carter *et al.*, 1988, 1989, Ford and Blenkinsop, 2008; Benomar *et al.* 2009; He *et al.*, 2010). For changes in land cover, this method assesses the probability of an occurring event, such as the alteration of a class (eg. cover of 'woody vegetation from others'), since there is evidence (e.g. slope) that this has already occurred.

In weights of evidence, the effect of a spatial variable in a transition is calculated independently from a combined solution. The weights of evidence represent each influence of a variable on the spatial probability of a transition i ⇒ j (Soares-Filho *et al.*, 2009) and are calculated according to **Eqs 3–4**.

$$O\{D/B\} = O\{D\}\frac{P\{B/D\}}{P\{B/\overline{D}\}} \quad (3)$$

$$\log\{D/B\} = \log\{D\} + W^{+} \quad (4)$$

since $O\{D\}$ and $O\{D/B\}$ are the odds, respectively, to occur prior to the event ***D*** and ***D*** occur as a spatial pattern ***B***. W^{+} is the positive weight of evidence of occurrence to event ***D*** as a spatial pattern ***B*** (Bonham-Carter, 1994; Soares-Filho *et al.*, 2003). The posterior probability of a transition $i \Rightarrow j$, given a set of spatial data ($B, C, D \ldots N$), is expressed as **Eq. 5**.

$$P\{i \Rightarrow j \mid B \cap C \cap D \ldots \cap N\} = \frac{O_{ij} e^{\sum W_{N^+}}}{1 + O_{ij} e^{\sum W_{N^+}}} \quad (5)$$

While the negative weight of evidence (W^{-}) is calculated according to **Eqs. 6–7**.

$$O\{D/B\} = O\{D\}\frac{P\{\overline{B}/D\}}{P\{\overline{B}/\overline{D}\}} \quad (6)$$

$$\log\{D/B\} = \log\{D\} + W^{-} \quad (7)$$

To test the significance of the association between the occurrences of transitions, related to land cover change, each one of the defined intervals for the variables, the measure of contrast between the positive and negative values of the weights of evidence, was used, calculated by Dinamica, whose formula is given by:

$$C = W^{+} - W^{-} \quad (8)$$

This measure is considered statistically significant, with a confidence interval of 95% if $|C| > 1{,}96\ s(C)$, being that the variance of the contrast, $S^2\ (C)$, is determined by (Goodacre *et al.*, 1993):

$$S^2 \quad (C) = \frac{1}{area(B \cap D)} + \frac{1}{area(B \cap \overline{D})} + \frac{1}{area(\overline{B} \cap D)} + \frac{1}{area(\overline{B} \cap \overline{D})} \tag{9}$$

Weights of evidence are applied in Dinamica EGO after the categorization of the maps of continuous variables, like the maps of distances. In this process, the preservation of the data structure is essential. For this purpose, the Dinamica EGO uses a method adapted from Agterberg & Bonham-Carter (1990) and Goodacre *et al.* (1993), in which intervals are calculated according to the data structure. Afterwards, intervals of categorization are applied, for the calculation of the weights of evidence. Positive values are obtained for variables that favor determined transition, while negative values indicate low probability of transition, and values close to zero are obtained for variables that have no effect on the transition.

Among the advantages of using the method of weights of evidence, Bonham-Carter (1994) points out:

1. Is not restricted by assumptions of parametric statistical methods, which are often violated by the data space.
2. It is a simple and objective method, since it uses only the result of the cross-tabulation between map changes from a cross of multitemporal maps - and the maps of proximal variables to feed formulas implemented in electronic spreadsheets.
3. The effect of each variable can be calculated independently from a joint solution, with the assumption that only the input variables are spatially independent.

The main disadvantage of this method is that it assumes that the input maps are conditionally independent between them.

Test of conditional independence

The weights of evidence method require(s) that the input maps are spatially independent. For this, is applied a group of statistical tests to estimate this assumption, like the index of Cramer (V), Joint Information Uncertainty (U), entropy measures and the chi-squared test. According to Bonham-Carter (1994), values less than 0.50 suggest less association than more. In pair values of correlated maps that present results of more than 0.50, should be chosen a variable and eliminated, considering:

1) The variable to be eliminated must be that which presents lower explanatory power for the phenomenon of land cover change, i.e.: the one that is less correlated with the event.

2) When both variables are important to model the phenomenon, both can be combined in a single layer, using a Boolean operation. Thus, there is no loss of information.

In this paper, we used the values of Cramer's test and the Joint Information Uncertainty to estimate the independence between the variables. These indexes have values ranging from zero (0) to one (1), in that the closer the values are to zero, the lower the degree of dependence between variables.

The Cramer Index is then defined by Bonham-Carter (1994) as:

$$V = \sqrt{\frac{X^2}{T.. \quad M}} \tag{11}$$

where X^2 is the chi-squared statistic of the contingency table relating two variables; $T..$ corresponding to the sum total of the values from the contingency table, M is the minimum of (n-1, m-1), where n is the number of lines, and m the number of columns.

The Joint Information Uncertainty belongs to the class of entropy measures, which are also based on the matrix of cross-tabulation, but can also be used for the measurement of associations. Assuming that the values T_{ij} are transformed in area proportions, p, by dividing each area element by the sum total $T..$ .Therefore, $p_{ij} = T_{ij}/T..$, and the marginal proportions are defined as $p_{i.} = T_{i.}/T..$ and as $p_{.j} = T_{.j}/T..$. Measures of entropy, or ‘statistical information’, are defined using the proportions of area as estimates of probabilities. Proportions do not have a dimension, which makes the entropy indexes have the advantage over the chi-squared, for not being affected by units of measure (Bonham-Carter, 1994).

Assuming that a matrix of area proportions for the maps A and B have been determined from T, then the entropy of A and B are defined as:

$$H(B) = -\sum_{i=1}^{n} p_{i.} \ln p_{i.} \tag{12}$$

$$H(A) = -\sum_{j=1}^{m} p_{.j} \ln p_{.j} \tag{13}$$

where *ln* is the natural logarithm. The joint entropy of combination, *H (A, B)* is:

$$H(A, B) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \ln p_{ij} \tag{14}$$

Thus, the Joint Information Uncertainty of *A* and *B*, *U (A, B*) can be used as a measure of association and is defined by:

$$U(A,B)=2\ [\frac{H(A)+H(B)-H(A,B)}{H(A)+H(B)}] \quad (15)$$

which varies between 0 (zero) and 1 (one). When both maps are completely independent, then $H(A,B) = H(A) + H(B)$ and $U(A,B)$ is *0*, and when both maps are completely dependent, $H(A) = H(B) = H(A,B) = 1$, and $U(A,B)$ is 1.

The transitional functions of Dinamica EGO

In models using cellular automata, the influence of neighboring cells on the transition probability is a key issue. The Dinamica divides the transition function on neighboring cells in two processes, denominated expander and patcher.

The expander function is responsible for the expansion or contraction of spots previously existing in a certain class of land cover, while the patcher function is intended exclusively to the formation of new spots via a mechanism formation of seeds, i.e.: the expander function performs transitions from one state *i* to one state *j* only in the vicinity adjacent to cells with the state *j*, and the function patcher makes transitions from one state *i* to state *j* only in adjacent neighborhoods of cells in a different state of *j* (Soares-Filho *et al.*, 2002).

These processes need to be defined for the parameters of the average size of the spot and variance of the size (given in hectare), beyond the isometry which varies between zero and two and refers to the aggregation of the spots.

Validation test

Validation tests can be understood as procedures to verify that models either reflect or don't reflect the reality in the desired degree (Batty, 1976). The method employed in this paper is based on the concept of fuzziness of location, in which the representation of a cell is influenced by itself, and to a lesser extent, by neighboring cells (Hagen, 2003). Regardless of the uncertainty of the category or cellular state, the neighborhood's vector "fuzzy neighborhood" may represent the uncertainty of location.

This evaluation method is considered a flexible concordance as it is not based on pixel by pixel, but the adjustment for multiple resolutions, in which values tend to be higher when compared to indexes of rigid compliance. The larger the sample window, the higher the index tends to be.

Thus, from a certain resolution (typically above 11 or 13 pixels) it is common that saturation occurs, giving the same inefficiency to assess the adjustment between the real map and the simulated map (Costanza, 1989).

In the method of validation by an exponential decay a fuzzy vector is associated to each cell on the map. This vector has as many elements as categories (classes of land cover) of the maps, adopting 1 for the category $= i$, and $2^{-d/2}$ for categories different than *i*, where (*d*) is the unitary distance between cells. When the class is not found in the neighborhood's window, the value 0 (zero) is used. The vector of fuzzy neighborhood (*Vnbhood*) for each cell is given by:

$$V_{nbhood} = \begin{pmatrix} \mu_{nbhood\ 1} \\ \mu_{nbhood\ 2} \\ . \\ . \\ . \\ \mu_{nbhoodC} \end{pmatrix} \quad (16)$$

$$\mu_{nbhood_i} = \left| \mu_{nbhood_{i,1}} * m_1, \mu_{crisp_{i,2}} * m_2, \ldots \right. \quad (17)$$

where $\mu\ nbhood_i$ represents relevance to the category *i* in a neighborhood of *R* cells (typically $R = n^2$); $\mu\ crisp_{ij}$ is the relevance of the category *i* to the neighboring cell *j*, assuming the vector *crisp* 1 for *i* and 0 for different categories of *i* ($i \subset C$); m_j is the relevance based on distance of neighboring cell *j*, where m refers to a decay function of distance, for example, an exponential decay ($m = 2^{-d/2}$), where d is the unitary distance between cells, measure from centroid to centroid.

The most appropriate choice in decay function and window size of the sampling depends on the uncertainty of the data and the tolerance threshold of spatial error. Although there is no consensus on which setting defines a threshold of acceptance or rejection of the model, it is accepted that a model gives good results when its setting is higher than obtained from a comparison between the initial and final maps (Hagen, 2003).

Input data

The dynamic modeling for land cover developed in this paper, was performed using two land cover maps, 1998 (initial landscape), and 2010 (final landscape), both derived from visual interpretation of Landsat 5 TM images.

The land cover maps were chosen in the period selected because of the great changes that are occurring in the study area, such as the construction of the Rio de Janeiro Metropolitan Arc (RJMA), Brazilian economic growth, and the World Cup 2014, which increased investments in Brazil, therefore, this period analyzed better characterizes future growth trends.

The Rio de Janeiro Metropolitan Arc is a great road construction project, which connects the Rio de Janeiro Petrochemical Complex (COMPERJ) and the Atlantic Steel Company (CSA) to the Port of Itaguaí, located in Sepetiba Bay, 80 km from the city of Rio de Janeiro, and major federal highways in the state to reroute the flow of vehicles away from the city of Rio de Janeiro.

We considered three classes of land cover: (1) woody vegetation, (2) constructed areas and (3) others (involving the entire non-urbanized area, water bodies, exposed soil and other less representative classes). The generalization of land cover classes was needed due to the resolution of the images used, and the complexity of the model tends to increase when greater numbers of distinct activity are considered (Batty *et al.*, 1999).

Eight spatial variables were used to explain the changes in land coverage in the analyzed period (1998-2010): (i) elevation, (ii) slope, (iii) distance to roads, (iv) protected areas, (v) distance to rivers, (vi) geology, (vii) mining areas, and (viii) distance to RJMA. **Table 1** shows the source of data and the process performed to obtain it.

The variables mined area and distance to RJMA began to cause an impact on the area in 2001 and 2009, respectively. Therefore, as the input data in the simulation should correspond to the initial time that the variable starts to influence the area, the Dinamica EGO has an operator called 'select' for inserting these variables in the iteration corresponding to the years they began to cause impact, and also adding the weights of evidence.

These variables were chosen according to the knowledge of the changes occurring in the area and the data availability.

Analysis and results

Transition rates

Three transitions occurred in the study period, as shown in **Table 2**. The transition 'woody vegetation to others' shows that a significant percentage of vegetation decreased at a rate of 2.52% per year during the period 1998-2010. The rate of urban growth is small compared with the other changes in the area.

Simulation period 1998-2010

As discussed in Section 3.3, in this paper Cramer's coefficient and Joint Information Uncertainty were used to verify the independence between pairs of variables. The results obtained for the statistical tests are shown in **Tables 3–5**.

Table 1. Source of data used and process performed to obtain it

Layer	Scale	Date	Source	Processing
Elevation	1:10.000	1975	FUNDREM[a]	DEM
Slope	1:10.000	1975	FUNDREM	DEM
Distance to roads	1:10.000	1975	FUNDREM	Euclidian distance
Distance to rivers	1:10.000	1975	FUNDREM	Euclidian distance
Protected areas	1:100.000	2000	EM[b]	-
Geology	1:500.000	2000	CPRM[c]	-
Mining areas	1:150.000	2001	TM Landsat 5	Visual interpretat ion
Distance to RJMA	1:250.000	2007	EIS[d]	Euclidian distance

[a] Foundation for the Development of the Metropolitan Region of Rio de Janeiro (FUNDREM); [b] Environmental Ministry (EM); [c] Mineral Resources Research Company (CPRM); [d] Environmental Impact Study (EIS).

Table 2. Annual transition of land cover

Land cover transitions	1998-2010 (%/year)
Woody vegetation to others	2.52
Others to woody vegetation	0.43
Others to constructed areas	0.05

Table 3. Values between pairs of variables for the transition 'others to constructed areas'

U/V	e	DRi	DR	DM	S	G
E	-	0,0757	0,0908	0,2028	0,2426	0,3990
DRi	0,0188	-	0,1236	0,0840	0,1950	0,5662
DR	0,0256	0,0494	-	0,1573	0,1523	0,4812
DM	0,1585	0,0215	0,0857	-	0,1029	0,3445
S	0,1461	0,0955	0,0800	0,0299	-	0,5330
G	0,2392	0,2170	0,2019	0,1182	0,4272	-

E = Elevation, G = Geology, DR = Dist. to rivers, Dist. to roads, DM = Dist. to RJMA, and S = Slope.

Cramer's coefficient (*V*)
Joint Information Uncertainty (*U*)

Table 4. Values between pairs of variables for the transition 'others to woody vegetation'

U/V	E	P	DRi	DR	S
E	-	0.2359	0.0904	0.0987	0.2706
P	0.0274	-	0.0239	0.2121	0.0969
DRi	0.0172	0.0008	-	0.0510	0.1110
DR	0.0265	0.0150	0.0025	-	0.0869
S	0.1634	0.0053	0.0170	0.0135	-

E = Elevation, P = Protected areas, DR = Dist. to rivers, Dist. to roads, and S = Slope.

Cramer's coefficient (*V*)
Joint Information Uncertainty (*U*)

Table 5. Values between pairs of variables for the transition 'woody vegetation to others'

U/V	E	M	P	DRi	DR	DM	S
E	-	0.1671	0.2360	0.0747	0.1039	0.0948	0.2708
M	0.0166	-	0.0208	0.4065	0.6285	0.0834	0.0327
P	0.0212	0.0040	-	0.0454	0.2126	0.0265	0.1032
DRi	0.0194	0.2472	0.0020	-	0.2473	0.0764	0.2252
DR	0.0243	0.1276	0.0140	0.1070	-	0.0955	0.1967
DM	0.0110	0.0136	0.0048	0.0057	0.0176	-	0.0464
S	0.1632	0.2479	0.0042	0.1250	0.0837	0.0043	-

E = Elevation, M = Mining areas, P = Protected areas, DR = Dist. to rivers, Dist. to roads, DM = Dist. to RJMA, and S = Slope.

Cramer's coefficient (*V*)

Joint Information Uncertainty (*U*)

The results obtained in tests of conditional independence shows that three pairs of variables had values above 0.50 for *V*. However, as the test *U* have the advantage of not being affected by the area, as happens with V and presented values below the accepted rate, all variables were retained. The values obtained for the weights of evidence for the variable slope for the transition 'others to woody vegetation' are in **Fig. 2**.

The graph in **Fig. 2** shows that the last ranges of the variable *slope* are of steeper area and thus costlier for human occupation, and therefore more prone to the regeneration of trees. The simulation tests were conducted with the land cover map of the initial land cover (1998), the set of spatial variables, the weights of evidence and the transition matrix. Also in this step, the parameters of the patcher and expander functions should be defined for the medium size, variance and isometry of the spots to be formed or expanded/contracted. In this paper, only the expander function was used due to the changes in landscape which occur exclusively through the expansion of previously existing spots. The parameters used to generate simulated landscape are inserted into Dinamica EGO in hectare (ha), the values used are shown in **Table 6**.

The relative values for average size and variance of the spots were obtained through the model Calc_Mean_Patch_Sizes_And_Standard_Deviations.Ego available from version 1.7.8 software, and adjustments were made to these parameters through visual analysis so the simulated map was as close to the actual final map as much as possible. The simulation test is shown in **Fig. 3**.

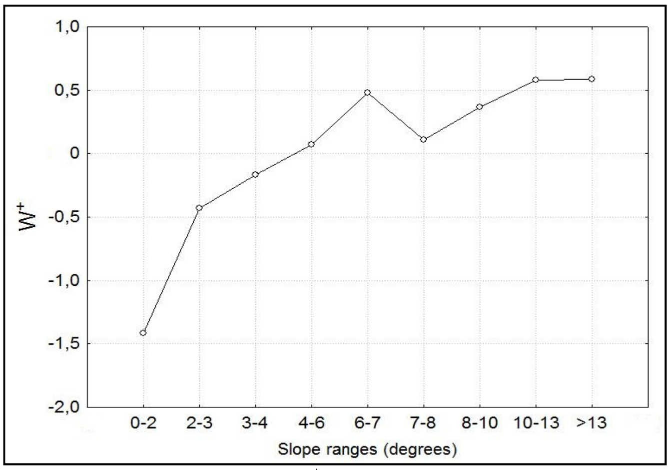

Fig. 2 Weights of evidence (W^+) values obtained in the simulation considering the variable slope in relation to the transition 'others to woody vegetation'.

Table 6. Parameters used in expander to generate landscape simulated during the period of 1998-2010

	Mean Patch Size (ha)	Patch Size Variance (ha)	Patch Isometry
Woody vegetation to others	0.8	0.5	1.0
Others to woody vegetation	0.7	0.5	1.0
Others to constructed areas	0.2	2.0	1.0

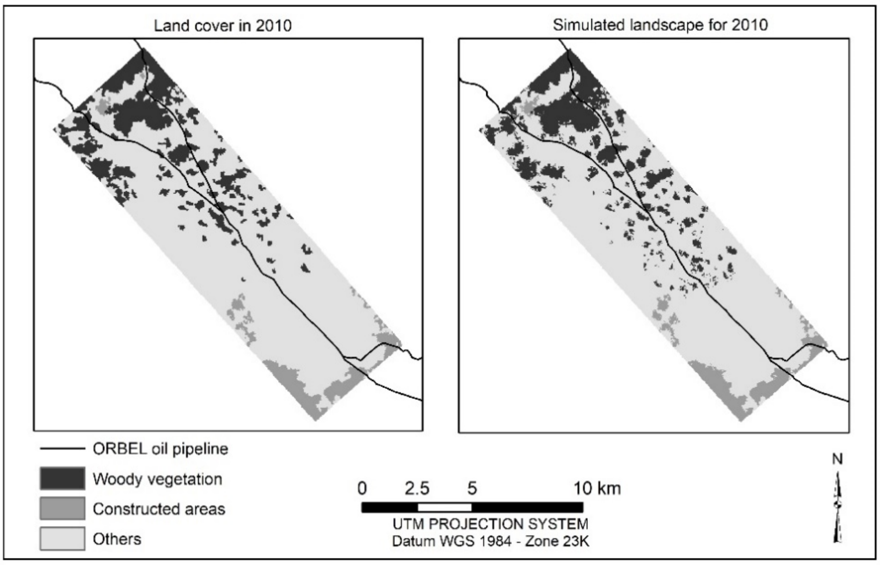

Fig. 3 Real and simulated landscape for the year of 2010.

The fragmentation of the woody vegetation class prevented better results for the simulation. The model validation in the second period of analysis was performed by the method of fuzzy similarity by exponential decay. The values are shown in **Fig. 4**.

Generation of scenarios

The construction of scenarios is a tool that can be used to understand and anticipate changes and thus improve decision-making. The scenarios are not predictions in the strict sense, but offer a different vision of different future alternatives, possible and imagined, in which decisions will be made (Chermack, 2007).

With the validated model, an annual land cover scenario was generated for the year 2020 called 'projective', which is based on the continuation of current trends into the future and, therefore, based on the same reference parameters for the expander and the same transition matrix for the period 1998-2010. Based on the outcome generated, changes were calculated for each class of land cover, according to **Table 7**.

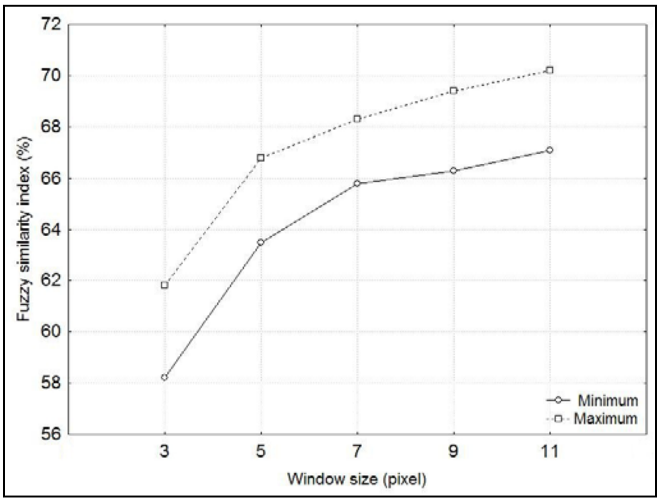

Fig. 4 Fuzzy similarity index based on multiple window sizes for the simulation of the landscape in 2010.

Table 7. Total annual area for the class of land cover for the scenario projective

Simulated maps "projective scenario"	Woody vegetation (km²)	Constructed area (km²)	Others (km²)
2010	16.67	6.17	74.28
2011	16.54	6.17	74.40
2012	16.44	6.20	74.46
2013	16.35	6.25	74.52
2014	16.26	6.29	74.56
2015	16.18	6.33	74.60
2016	16.10	6.37	74.64
2017	16.01	6.42	74.68
2018	15.94	6.46	74.72
2019	15.86	6.50	74.75
2020	15.78	6.55	74.78

The **Table 7** was calculated using a map of the actual coverage of the year 2010, obtained through classification of Landsat imagery, and annual maps of the generated prognosis.

The analysis of the table permits evaluating the source and destination of land cover classes. There was an increase of 0.38 km^2 in constructed area in 10 years, corresponding to a total increase of 6.0%, while the woody vegetation decreased, 0.89 km^2 totaling a loss of 5.47% over the same period.

It is also possible to evaluate that in the first two years that RJMA (2012/2013) will be in operation, there will be the highest rate of growth in the constructed area class, showing an increase of 0.81% during that period. This major growth in the early period of RJMA operation can be explained by the increase of road connections and regular transport lines, which are currently lacking in the area.

The modeling results indicate the growth of suburbs close to AMRJ. The neighborhoods located south of the study area, belonging to the municipalities of Belford Roxo and Duque de Caxias, which already suffer from a lack of infrastructure relative to the lack of paved streets, unregulated land occupation, and flooding during the rainy season.

The planning of the pipelines is done taking into consideration the crossing, at its greatest extent, in rural areas. However, population growth and economic causes make these areas previously unoccupied come to be used for urban use. This population growth near the range of pipes increases the risk of damage to the pipeline and the population living or working near these areas.

Considering the direct area of influence of 400 m on each side of the duct, special attention should be given to districts within the municipality of Duque de Caxias that are crossed by the pipeline. The expansion of these areas must be accompanied by the public power who should also consider the possibility of expropriation of property and relocation of these people due to the proximity of the pipeline.

CONCLUSIONS

The reduced size of the study area and the resolution of images used to obtain the maps of land cover is a factor that should be taken into account in modeling, since more accurate results could be obtained if satellite images or aerial photographs with a higher resolution were available, and also if the model was extrapolated for the total municipality area, and thus make it possible to consider a greater number of variables.

The model developed in this article helped to understand the phenomena involved in the changes of land cover through the results obtained using statistical methods. The Dinamica EGO proved to be a useful and flexible tool, capable of creating different models to represent reality.

The safety of the pipeline and the population living near it, can present the following considerations:

- The establishment of standards for the use and land cover in areas close to the range of pipelines by the public in conjunction with the company responsible for operating the pipeline, through management plans, can guide urban development near these areas. In this sense, the dynamic modeling contributes to the planning of the areas near the pipelines, providing the main trends of occupation.
- The creation, maintenance and availability of a database through the internet, with information relating to pipelines in operation, transported products, and the types of land use and land cover near the pipeline would also be an important tool for risk management and decision making in case of accidents.

Acknowledgements: The authors acknowledge the financial support granted by Cenpes/Petrobras and CNPq.

REFERENCES

Aljoufie, M., Brussel, M., Zuidgeest, M., van Maarseveen, M., 2013. Urban growth and transport infrastructure interaction in Jeddah between 1980 and 2007. *Int. J. Applied Earth Observ. Geoinf.,* **21** 493–505.

Almeida, C.M., Batty, M., Monteiro, A.M.V., Camara, G., Soares-Filho, B.S., Cerqueira, G.C., Pennachin, C.L., 2003. Stochastic cellular automata modeling of urban land use dynamics: empirical development and estimation. *Comp. Environ. Urban Sys.,* **27**(5) 481-509.

ANP, 2017. *Oil, natural gas and biofuels.* Statistical Yearbook 2017. ANP, Rio de Janeiro.

Batty, M. (1976). *Urban modelling: algorithms, calibrations, predictions.* Cambridge University Press.

Batty, M., Xie, Y., Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Comp. Environ. Urban Sys.,* **23**(3) 205-233.

Bell, E.J., Hinojosa, R.C. (1977). Markov Analysis of Land-Use Change - Continuous Time and Stationary Processes. *Socio-Economic Plann. Sci.,* **11**(1) 13-17.

Bonham-Carter, G., Agterberg, F., Wright, D. (1988). Integration of geological datasets for gold exploration in Nova Scotia. *Photogr. Engin. Remote Sens.,* **54**(11) 1585-1592.

Bonham-Carter, G.F.; Agterberg, F.P., Wright, D.F. (1989). *Weights of evidence modeling: A new approach to mapping mineral potential.* In: Statistical Applications in the Earth Sciences, Agterberg, F.P. & Bonham-Carter. G.F. (Ed.), 171-183, Geological Survey of Canada 98, ISBN 0660135922, Canadian Government Publishing Centre

Bonham-Carter, G.F. (1994). *Geographic information systems for geoscientists: modelling with GIS.* Pergamon, New York.

Benomar, T.B., Hu, G., Bian, F. (2009). A predictive GIS model for potential mapping of copper, lead, and zinc in Langping area, China. *Geo-Spatial Inf. Sci.,* **12**(4) 243-250.

Chermack, T.J., 2007. Disciplined imagination: Building scenarios and building theories. *Futures* **39**(1) 1-15.

Choi, J., Goo, B., Kim, J., Kim, Y., Kim, W., 2003. Development of limit load solutions for corroded gas pipelines. *Int. J. Pressure Vessels Piping* **80**(2) 121-128.

Costanza, R. (1989). Model Goodness of Fit - a Multiple Resolution Procedure. *Ecol. Model.,* **47**(3-4) 199-215.

Ford, A., Blenkinsop, T.G. (2008). Combining fractal analysis of mineral deposit clustering with weights of evidence to evaluate patterns of mineralization: Application to copper deposits of the Mount Isa Inlier, NW Queensland, Australia. *Ore Geology Rev.,* **33**(3-4) 435-450.

Geertman, S., Stillwell, J.C.H. (2002). *Planning support systems in practice.* Springer Verlag.

Goodacre, A.K., Bonham-Carter, G.F., Agterberg, F.P., Wright, D.F. (1993). A Statistical-Analysis of the Spatial Association of Seismicity with Drainage Patterns and Magnetic-Anomalies in Western Quebec . *Tectonophysics* **225**(4) 551-551.

Hagen, A. (2003). Fuzzy set approach to assessing similarity of categorical maps. *Int. J. Geogr. Inf. Sci.,* **17**(3) 235-249.

He, B., Chen, C., Liu, Y., 2010. Mineral Potential Mapping for Cu-Pb-Zn Deposits in the East Kunlun Region, Qinghai Province, China, Integrating Multi-source Geology Spatial Data Sets and Extended Weights-of-Evidence Modeling. *Gisci. Remote Sens.,* **47**(4) 514-540.

IBGE - Brazilian Institute of Geography and Statistics (2010). *Censo Demográfico 2010.* IBGE, Rio de Janeiro.

JRC - Joint Research Centre, European Commission/Institute for Remote Sensing Applications (1994). *Modelling deforestation processes – a review.* Research Report No.1. Luxembourg: ECSC-EC-EAEC.

Judge, G.G., Swanson, E. R. (1962). Markov chains: Basic concepts and suggested uses in agricultural economics. *Austr. J. Agric. Economics* **6**(2) 49-61.

Lee, D.B. (1973). Requiem for Large-Scale Models. *J. Amer. Inst. Planners* **39**(3) 163-178.

Papadakis, G.A. (2000). Assessment of requirements on safety management systems in EU regulations for the control of major hazard pipelines. *J. Hazard. Mat.,* **78**(1-3) 63-89.

Pérez-Vega, A., Mas, J.-F., Ligmann-Zielinska, A. (2012). Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest. *Environm. Model. Soft.,* **29**(1) 11-23.

Soares-Filho, B.S., Assunção, R.M., Pantuzzo, A.E. (2001). Modeling the spatial transition probabilities of landscape dynamics in an amazonian colonization frontier. *Bioscience* **51**(12), 1059-1067.

Soares-Filho, B.S., Cerqueira, G.C., Pennachin, C.L. (2002). DINAMICA - a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecol. Model.,* **154**(3) 217-235.

Soares-Filho, B.S., Rodrigues, H.O., Costa, W.L. (2009). *Modeling environmental dynamics with Dinamica EGO.* CSR-UFMG, Belo Horizonte.

Zeng, C., Liu, Y, Stein, A., Jiao, L., 2015. Characterization and spatial modeling of urban sprawl in the Wuhan Metropolitan Area, China. *Int. J. Applied Earth Observ. Geoinf.,* **34**. 10-24.