Geocomputation techniques for spatial analysis: are they relevant to health data?

Técnicas de geocomputação para análise espacial: é o caso para dados de saúde?

Gilberto Câmara [1]
Antônio Miguel Vieira Monteiro [1]

[1] *Divisão de Processamento de Imagens, Instituto Nacional de Pesquisas Espaciais. C. P. 515, São José dos Campos, SP 12201-027, Brasil.*

Abstract *Geocomputation is an emerging field of research that advocates the use of computationally intensive techniques such as neural networks, heuristic search, and cellular automata for spatial data analysis. Since increasing amounts of health-related data are collected within a geographical frame of reference, geocomputational methods show increasing potential for health data analysis. This paper presents a brief survey of the geocomputational field, including some typical applications and references for further reading.*
Key words *Spatial Analysis; Geographical Methods; Computational Technics*

Resumo *A geocomputação é um campo de pesquisa emergente que propõe o uso de técnicas intensivas em computação, tais como redes neurais, busca heurística e autômatos celulares para análise de dados espaciais. Com o aumento do volume de dados de saúde coletados dentro de um referencial geográfico, os métodos geocomputacionais demonstram um potencial crescente para a análise desses mesmos dados. Os autores apresentam uma revisão breve do campo da geocomputação, apresentando algumas aplicações típicas e sugestões bibliográficas.*
Palavras-chave *Análise Espacial; Métodos Geográficos; Técnicas Computacional*

Introduction

In recent years, the use of computer-based techniques for spatial data analysis has grown into an important scientific field, combining techniques from geographic information systems and emerging areas such as neurocomputing, heuristic search, and cellular automata. In order to distinguish this new interdisciplinary area from the simple extension of statistical techniques to spatial data, Oppenshaw & Abrahart (1996) coined the term "geocomputation" to describe the use of computer-intensive methods for knowledge discovery in physical and human geography, especially those involving non-conventional data clustering and analysis techniques. More recently the term has been applied in a broader sense to include spatial data analysis, dynamic modeling, visualization, and space-time dynamics (Longley, 1998).

This paper is a brief survey of geocomputational techniques. The review should not be considered exhaustive, rather attempting to provide an overview of the concepts and motivation behind the term "geocomputation". Our prime motivation is to draw the attention of the public health community to the new analytical possibilities offered by geocomputational techniques. We hope this discussion will serve to broaden their perceptions of new possibilities in the spatial analysis of health data.

Motivations for research in geocomputation

Simply defined, geocomputation *"is the process of applying computing technology to geographical problems"*. According to Oppenshaw & Abrahart (1996:665), *"Many end-users merely want answers to fairly abstract questions such as 'Are there any patterns, where are they, and what do they look like?'"*. Although this definition is generic, it points to a number of motivating factors, like the emergence of computerized data-rich environments, affordable computational power, and spatial data analysis and mining techniques.

The first motivation (data-rich environments) has come about through the massive collection of socioeconomic, environmental, and health-related data, increasingly organized in computerized databases with geographical references such as census tracts or postal codes. Even in Brazil, a country with a limited tradition of public availability of geographical data, the 2000 Census is being described as the first such initiative where all data collection will be automated and georeferenced.

The second motivation (computational power) has materialized in two forms: the emergence of the Geographic Information Systems (GIS) technology and of a set of algorithmically-driven techniques such as neurocomputation, fuzzy logic, and cellular automata.

The third motivation (data analysis and mining techniques) has been heavily driven by the application of data analysis techniques to spatial statistics, a research topic of considerable importance in recent decades.

The broad nature of challenges and approaches to geocomputational research is perhaps best illustrated by four different yet complementary approaches: computer-intensive pattern search, exploratory spatial data analysis, artificial intelligence techniques, and dynamic modeling, as described in the following sections.

Focus 1 – Computer-intensive pattern search

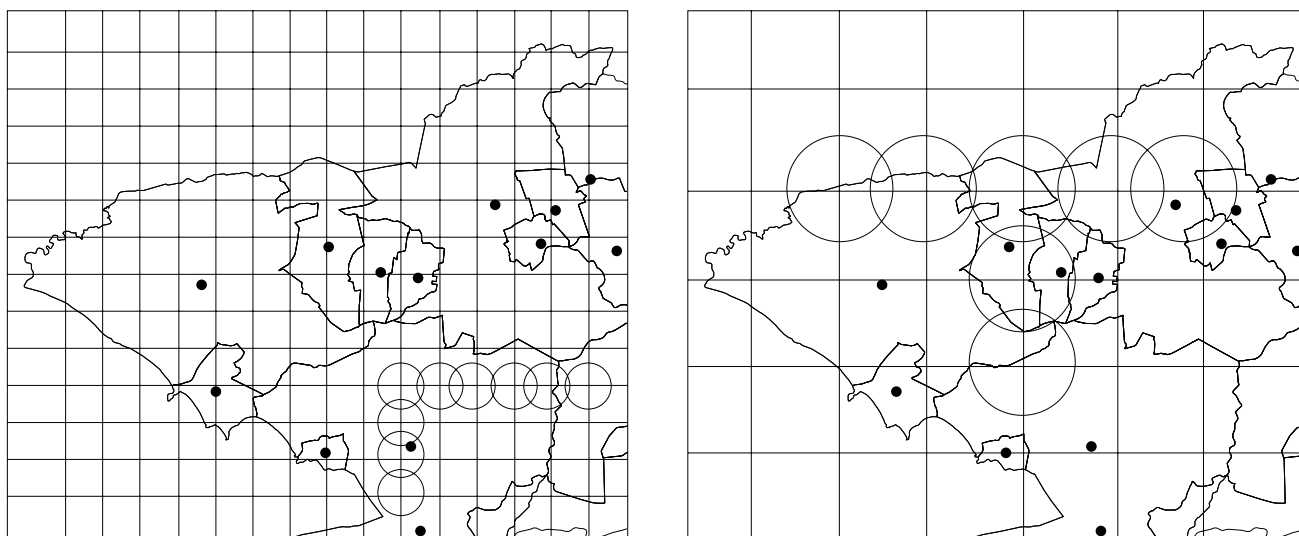### GAM – The Geographical Analysis Machine

One of the most typical examples of the computer-intensive approach to geocomputation is the Geographical Analysis Machine (GAM) developed by Stan Oppenshaw and co-workers at the Centre for Computational Geographics at the University of Leeds. For a recent survey of the GAM, see Oppenshaw (1998). The following description is largely based on Turton (1998).

GAM is basically a cluster finder for point or small-area data. Its purpose is to indicate evidence of localized geographical clustering in cases where statistical distribution of the phenomenon is not known in advance. For the GAM algorithm, a cluster is like a localized excess incidence rate that is unusual in that there is more of some variable than might be expected. Examples would include: a local excess disease rate, a crime hot spot, an unemployment black spot, unusually high positive residuals from a model, the distribution of a plant, surging glaciers, earthquake epicenters, patterns of fraud, etc (Figure 1).

The basic idea of the GAM is very simple. Within the study region containing a spatial point pattern, GAM works by examining a large number of circles of varying sizes that completely cover the region of interest. The circles overlap to a large degree to allow for edge effects and to provide a degree of sensitivity

Figure 1

Two steps in the GAM algorithm – left, initial step with smaller circles; right, later step with larger circles.

GAM = Geographical Analysis Machine.

analysis. Within each random circle, one counts the number of points and compares this observed value with an expected value based on an assumption about the process generating the point pattern (usually that it is random). Ideally, the population at risk should be used as the basis for generating the expected value, such as using a Poisson probability model with the observed mean and the population at risk within each circle. Once the statistical significance of the observed count within a circle has been examined, the circle is drawn on a map of the region if it contains a statistically significant cluster of points. The process is repeated many times until a map is produced containing a set of circles centered on parts of the region where interesting clusters of points appear to be located.

A GAM application to infant mortality in Rio de Janeiro

Oppenshaw (1998) makes a strong case for performance of the GAM algorithm to locate clusters of diseases, including a comparison with other cluster-finding techniques. To better assess and understand the potentials and limitations of GAM, Teruiya et al. (1999) conducted an investigation using data from the study *Spatial Analysis of Live-Born Profile and Socioeco-*

nomic Conditions in Rio de Janeiro, by D'Orsi & Carvalho (1998). This study assessed the spatial birth and socioeconomic patterns in subdivisions of the city of Rio de Janeiro, aiming to identify the main groups of infant morbidity and mortality risks and the selection of prime areas for preventive programs.

In order to apply the GAM algorithm, the values had to be converted from areal-related patterns to point variables. The authors selected some of the basic attributes used by D'Orsi & Carvalho and converted each area unit (corresponding to a city district) to a point location which received the value of the areal unit it represented, as illustrated by Figure 2.

The GAM algorithm was applied to the values for the live-born quality index for all neighborhoods of Rio. GAM found three clusters of high values for this index, located approximately in the Botafogo, Barra da Tijuca, and Ilha do Governador regions (Figure 3). The results were concentrated in what is perceived by the algorithm as "extreme" events of high values for the index, disregarding cases which are not "significant" enough. As a basis for comparison, the traditional cloropleth-map is shown in Figure 4, where the areal-based values are grouped by quintiles.

It should be noted that we have used the Rio de Janeiro birth patterns merely as an ex-

Figure 2

Location of Rio de Janeiro urban sub-divisions.
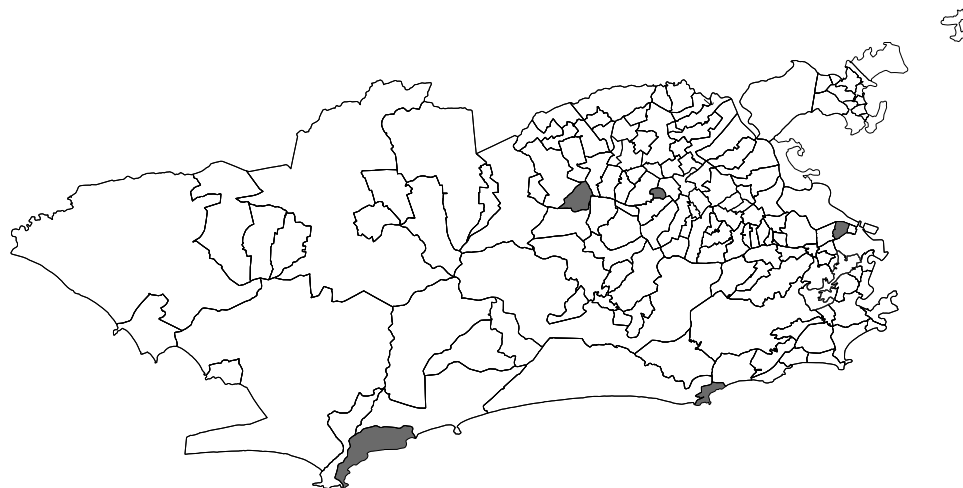
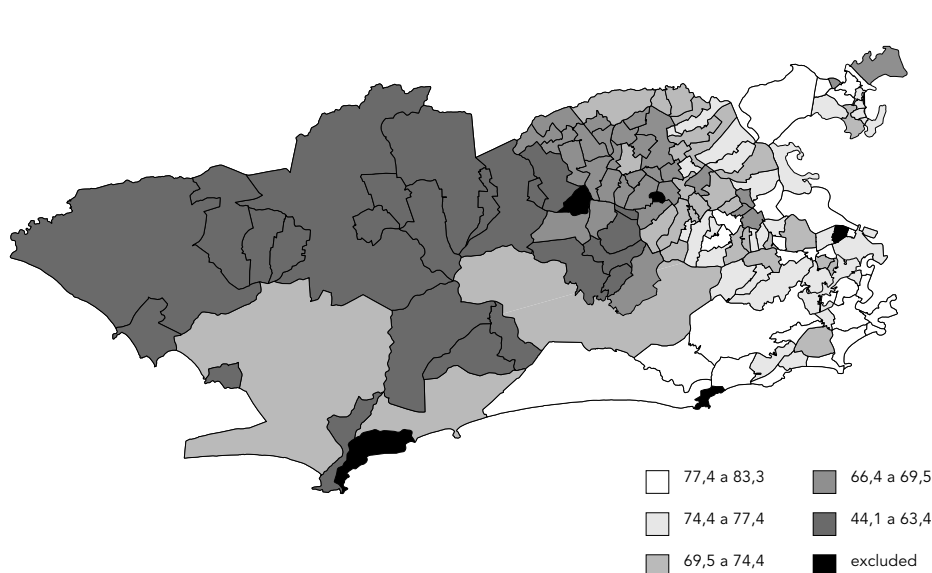Source: d'Orsi & Carvalho, 1997.

Figure 3

Clusters of high APGAR index values found by GAM.

GAM = Geographical Analysis Machine; APGAR = APGAR scoring for newborns.

Figure 4

Grouping of APGAR index values.

Source: d'Orsi & Carvalho, 1997.
APGAR = APGAR scoring for newborns.

ample to illustrate the computational behavior of the GAM technique. It is important to note that the algorithm was only searching for clusters of high values for the live-born quality index. Clusters of low values are disregarded by GAM, since the algorithm was originally conceived to find clusters of high disease incidence. We hope to motivate health researchers to apply the GAM techniques to problems closer to its original intended use, such as sets of epidemiological events.

Focus 2 – Exploratory spatial data analysis

Local spatial statistics

Statistical data analysis currently is the most consistent and established set of tools to analyze spatial data sets. Nevertheless, the application of statistical techniques to spatial data faces an important challenge, as expressed in Tobler's (1979) *First Law of Geography*: everything is related to everything else, but near things are more related than distant things. The quantitative expression of this principle is the effect of *spatial dependence*: the observed values will be spatially clustered, and the samples will not be independent. This phenomenon, al-

so termed *spatial autocorrelation*, has long been recognized as an intrinsic feature of spatial data, and measures such as the *Moran coefficient* and the *semi-variogram plot* have been used to assess the global association of the data set (Bailey & Gattrel, 1995).

Most spatial data sets, especially those obtained from geo-demographic and health surveys, not only possess global spatial autocorrelation, but also exhibit significant patterns of spatial instability, which is related to regional differentiations within the observational space. As stated by Anselin (1995), *"the degree of nonstationarity in large spatial data sets is likely to be such that several regimes of spatial association would be present"*.

In order to assess the degree of spatial instability, various *local spatial statistics indicators* have been proposed, such as the Moran Local I_i (Anselin, 1995), the Moran scatterplot (Anselin, 1996) and the G_i and G_i^* statistics (Ord & Getis, 1995). For a recent review, see Getis & Ord (1996). Although local spatial statistics can be seen as a branch of spatial statistics, they have been highly praised by geocomputational proponent Stan Oppenshaw: *"it is absolutely fundamental that we can develop tools able to detect, by any feasible means, patterns and localized association that exist* within the map*"* (Oppenshaw & Abrahart, 1996:665).

Spatial statistics as a basis for zoning: social exclusion/inclusion in São Paulo

In order to assess the validity of local spatial statistics, the authors have conducted a project to study the potential of such indicators as a basis for the design of administrative zoning systems for the city of São Paulo. As is well known, zone design is a major challenge for urban and regional planners, since it involves major decisions on how to distribute public resources.

São Paulo is one of the world's largest cities and presents a major challenge to urban planners and public administrators. Given its present size (over 13 million inhabitants) and enormous socioeconomic inequalities, rational planning of the city requires a careful division of the urban space into administrative regions that are homogenous by some objective criteria. Unfortunately, the current regional division of São Paulo has been driven by historical and political forces and fails to reflect a rational attempt to challenge the city's disparities.

As a basis for a zoning design for São Paulo, we have taken the "Social Exclusion/Inclusion Map of São Paulo", a comprehensive diagnosis of the city coordinated by Prof. Aldaiza Sposati of the Social Research Group at the Catholic University of São Paulo. This map used 49 variables obtained from Census data and local organizations to quantify the social apartheid in 96 districts of São Paulo (Sposati, 1996).

The main results of the "social exclusion/inclusion map" were indicators of social exclusion and disparities in quality of life in São Paulo. Figure 5 shows the map of the social exclusion index (*Iex*), where the values vary from -1 (maximal social exclusion) to +1 (maximal social inclusion) with a value of 0 indicating the attainment of a basic standard of inclusion (Sposati, 1996). Note from the map that two-thirds of the districts in São Paulo have acceptable living standards.

Taking the social exclusion index as a basis, the proposed task was to group the 96 districts into a set of administrative zones, each containing a significant number of districts and homogeneous with respect to social exclusion status. We used two exploratory spatial analysis tools: the Moran Scatterplot Map (Figure 6, left) and the local Moran index significance map (Figure 6, right). The basis for these local spatial statistics indicators is the use of a neighborhood or contiguity matrix W whose elements are $w_{ij} = 0$ if i and j are not neighbors and non-zero otherwise.

The *Moran scatterplot map* is a tool for visualizing the relationship between the observed values Z and the local mean values WZ, where Z indicates the array of attribute values (expressed as deviations from the mean) and WZ: is the array of local mean values, computed using matrix W. The association between Z and WZ can be explored to indicate the different spatial regimes associated with the data and to display graphically as indicated by Figure 6 (left). The Moran Scatterplot Map divides spatial variability into four quadrants:
• Q1 (positive values, positive local means) and Q2 (negative values, negative local means): indicate areas of positive spatial association.
• Q3 (positive values, negative local means) and Q4 (negative values, positive local means): indicate areas of negative spatial association.

Since the *Iex* variable exhibits global positive spatial autocorrelation (Moran I = 0.65, significance = 99%), areas in quadrants Q3 and Q4 are interpreted as regions that do not follow the same global process of spatial dependence, and these points indicate transitional regions between two different spatial regimes.

The local Moran index I_i is computed by multiplying the local normalized value z_i by the local mean (Anselin, 1995):

In order to establish a significance test for the local Moran index, Anselin (1995) proposes a pseudo-distribution simulation by permutation of the attribute values among the areas. Based on this pseudo-distribution, traditional statistical tests are used to indicate local index values with significance of 95% (1.96σ), 99% (2.54σ), and 99.9% (3.20σ). The "significant" indexes are then mapped and posited as "hot spots" of local non-stationarity.

The local Moran index significance map indicated three "hot spots", two of which related to low values of inclusion (located to the South and East of the city) and one related to high values of inclusion (located in the Center of the city). These patterns correspond to the extreme regions of poverty and wealth in the city and were chosen as "seeds" in the zoning procedure.

The remaining regions were defined interactively, taking into account the Moran scatterplot map, which clearly indicates a number of transition regions between the regions of Q1 and Q2 locations (to so-called "high-high" and "low-low" areas), some of which are indicated by the ellipses. These regions were grouped into separate zones. The work proceeded interactively until a final zoning proposal was produced, which can be confronted with the current administrative regions (Figure 7).

In order to assess the resulting map, a regression analysis was performed. This regression analyzes the correlation between the per-

centage of houses with proper sewage facilities (as independent variable) and the percentage of people over 70 years of age (as dependent variable). The rationale behind this choice was that social deprivation is a serious impediment to healthy living, as measured by the percentage of elderly in the population. Three OLS (ordinary least squares) regression analyses were performed: the first, taking all districts of the city overall; the second, using the current administrative division as separate spatial regimes; and the third, using the proposed new zoning as spatial regimes. The results as summarized in Table 1.

These results are a positive indication of the possible use of local spatial statistics as a basis for zoning procedures and show how indicators such as the social exclusion index of Sposati (1996) can be used as a support for urban planning.

## Focus 3 – Neural networks and geographic analysis

Introduction

An Artificial Neural Network (ANN) is a computer paradigm inspired by the way the brain processes information. The key element in this paradigm is a processing system composed of a large number of highly interconnected elements (neurons) working in unison to solve specific problems. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process (Gopal, 1998).

In principle, ANNs can represent any computable function, i.e., they can do everything a normal digital computer can do. In practice, ANNs are especially useful for classification and function approximation and mapping problems which are tolerant of some imprecision and have plenty of training data available. Almost any mapping between vector spaces can be approximated to arbitrary precision by feedforward ANNs (which are the type most often used in practical applications) if there are enough data and enough computing resources.

Given the capabilities of ANNs as exploratory tools in data-rich environments, there has been considerable interest in their use for spatial data analysis, especially in remote sensing image classification (Kannelopoulos, 1997; Leondes, 1997). Other geographical applications include: spatial interaction modeling (Gopal & Fischer, 1996; Oppenshaw, 1993) and classification of census data (Winter & Hewitson, 1994).

Figure 5

Social Exclusion Index (Iex) in São Paulo (96 districts grouped in sextiles).

Social Exclusion

-1 to -0.69

-0.69 to -0.43

-0.43 to -0.27

-0.27 to -0.08

-0.08 to 0.26

0.26 to 1
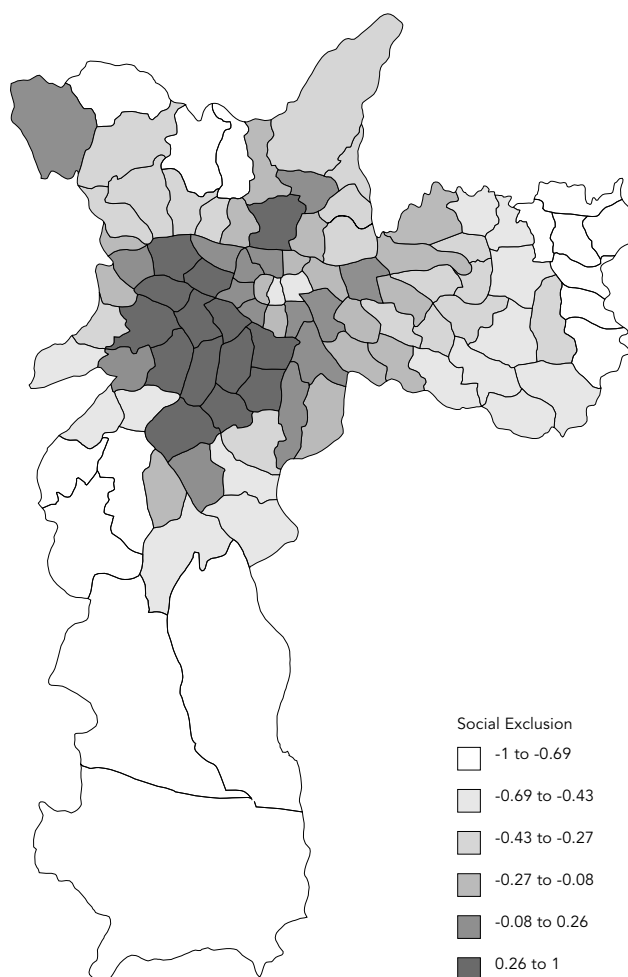
Neural networks for spatial data integration: an economical-ecological zoning application

To illustrate the potential of ANN for spatial data analysis, we have selected one example: the use of neural networks for the integration of multiple spatial information for an environmental zoning application (Medeiros, 1999). Although the chosen application does not involve health data, the integration procedure shown is relevant to heath-assessment applications, which involve multiple data sets as possible sources of epidemiological risk.

One of the more important problems in geographical data analysis is the integration of separate data sets to produce new spatial infor-

Figure 6

Left: Moran Scatterplot Map for Iex in São Paulo. Right: Significant values of Local Moran Index (gray = 95%; dark gray = 99%).
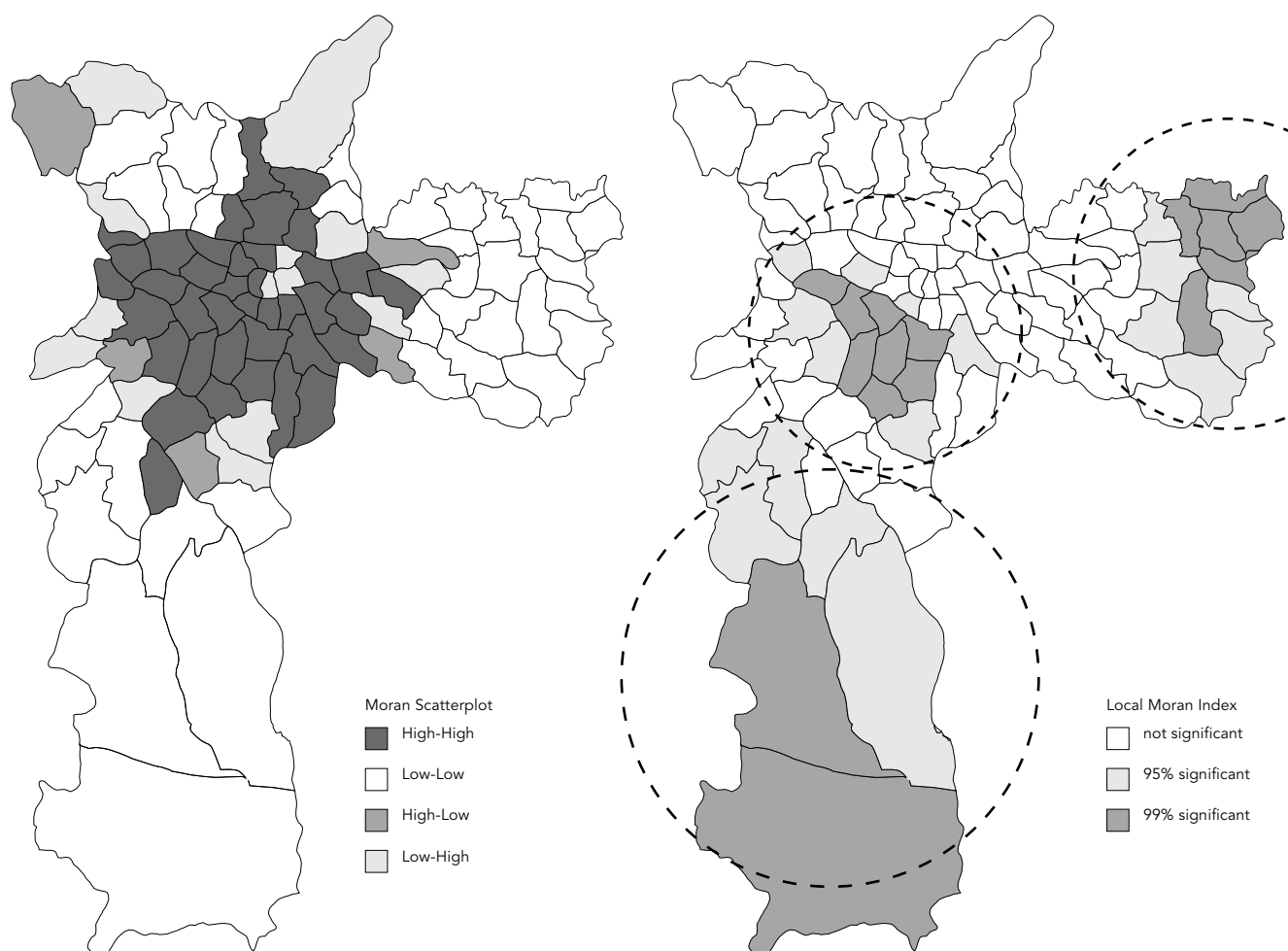
mation. For example, in health analysis, a researcher may be interested in assessing the risks associated with a disease (such as malaria) based on a combination of different conditions (land use and land cover, climatology, hydrological information, and distance to main roads and cities). These conditions can be expressed as maps, integrated into a common geographical database by means of GIS technology.

Once the data has been organized in a common geographical reference, the researcher needs to determine a procedure to combine these data sets. Taking a hypothetical example, a health researcher may want to calculate a *risk map for malaria* based on known disease incidence, climate, distance to cities, and land cov-

er, where the conditions are such that a region is deemed "high risk for malaria" if it rains more that 1000 m/year and the land cover is "inundated forest" and is located less than 50 km from a city.

The main problem with these map inference procedures is their *ad hoc*, arbitrary nature: the researcher formulates hypotheses from previous knowledge and applies them to the data set. The process relies on inductive knowledge of the reality. Additionally, when the input maps have many different conditions, the definition of combinatory rules for deriving the output may be difficult. For example, if an input map has eight different conditions (e.g., land cover classes) and five maps are to

Figure 7

Administrative Zones in São Paulo. Left – Current Division in 11 administrative regions. Right – Proposed division in 13 new regions.
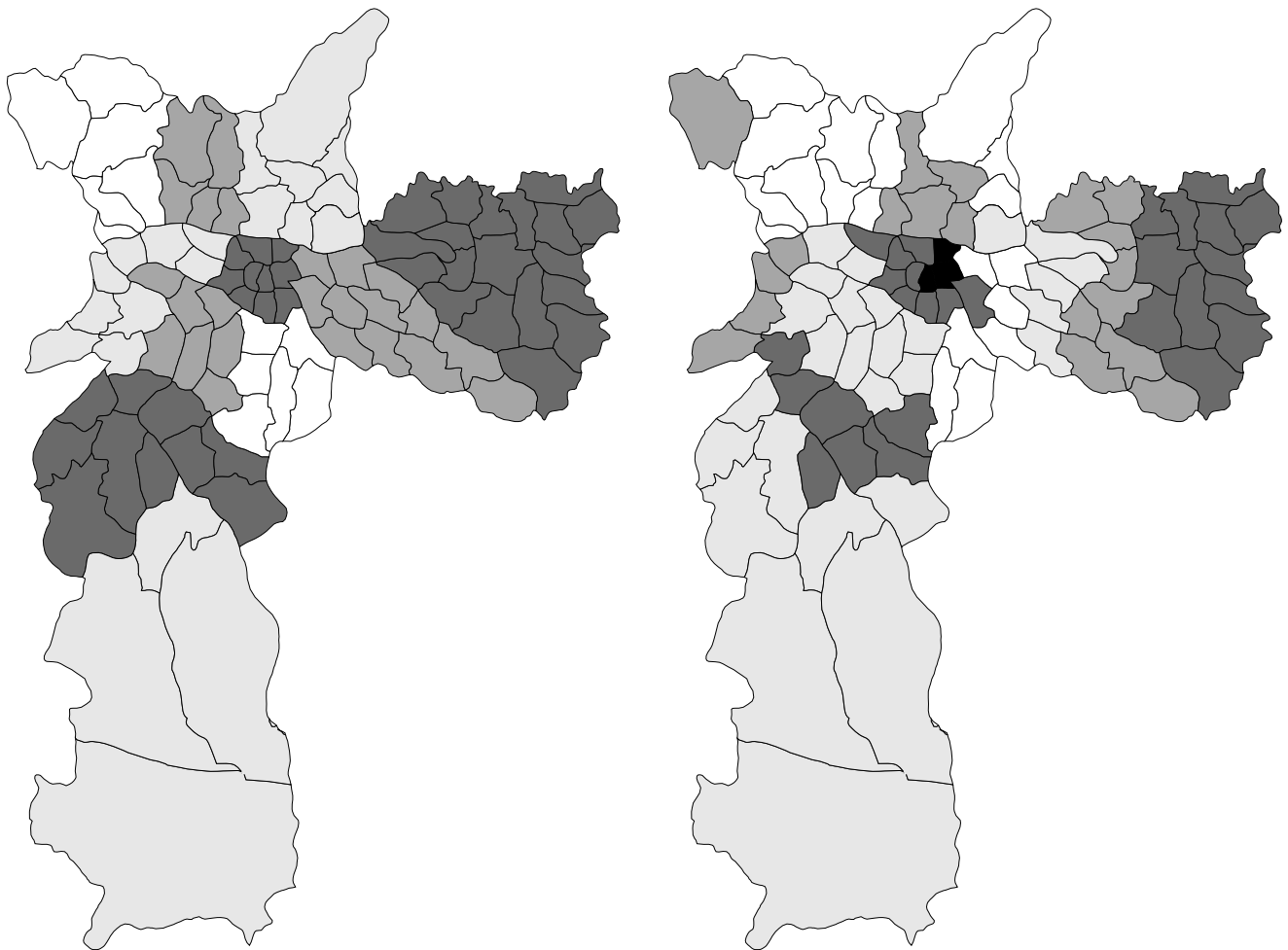
be combined, then 8^5 (= 32,768) different situations have to be taken into account.

There are two main alternative approaches to this problem. One is to use *fuzzy logic* to combine the maps (Câmara et al., 2000). In this case, all input data are transformed into fuzzy sets (in a [0,1] scale) and a fuzzy inference procedure may be used. Alternatively, the use of neural network techniques aims at capturing the researcher's experience, without the need for the explicit definition of the inference procedure. The application of neural networks to map integration can be done using the following steps:

• Create a georeferenced database with the input (conditional maps)

• Select well-known regions as training areas. For these areas, indicate the desired output response (such as health risk).

• Use these training areas as inputs to a neural network learning procedure.

• Using the trained network, apply the inference procedure for the entire study region.

• Evaluate the result and redo the training procedure, if necessary.

This idea was applied by Medeiros (1999) in his study of the integration of natural resources data as a basis for economical-ecological zoning in the Amazon region. Medeiros used five data sets as input: vegetation, geology, geomorphology, soils, and remote sensing images. Medeiros (1999) compared the result obtained

Table 1

Correlation coefficients for (old age, sewage) regression in São Paulo.

Situation	Number of spatial regimes	R² (correlation coefficient)
All city districts	1	0.35
Current Zoning	11	0.72
Proposed Zoning	13	0.83

by the neural network with a subjective operator interpretation and found a very strong spatial coherence between the two maps, with the neural-produced one being more restrictive in terms of results than the subjective one (Figure 8).

Focus 4 – Cellular automata

The computer representation of geographical space in current GIS technology is essentially static. Therefore, one important research focus in geocomputation aims to produce models that combine the *structural* elements of space (geographical objects) to the *processes* that modify such space (human actions as they operate in time). Such models would free us from static views of space (as centuries of map-making have conditioned us) and to emphasize the dynamic components as an essential part of geographical space.

This motivation has led to the use of cellular automata as a technique for simulation of urban and regional growth. Cellular automata (CA) are very simple dynamic spatial systems in which the state of each cell in an array depends on the previous state of the cells within a neighborhood of the cell, according to a set of transition rules. CA are very efficient computationally because they are discrete, iterative systems that involve interactions only within local regions rather than between all pairs of cells. The good spatial resolution that can thus be attained is an important advantage when modeling land use dynamics, especially for planning and policy applications (White & Engelen, 1997).

A conventional cellular automaton consists of: (a) a *Euclidean space* divided into an array of identical cells; (b) a cell neighborhood; (c) a set of discrete *cell states;* (d) a set of *transition rules* which determine the state of a cell as a function of the states of cells in the neighborhood; and (e) *discrete time steps*, with all cell states updated simultaneously.

The application of CA to geographical systems was first proposed by Tobler (1979). More recently, a number of researchers have proposed modifications of the original CA idea to accommodate geographical constraints. The most important characteristic to be discarded is the homogeneous cell space, replaced by a space in which each cell has its own inherent set of attributes (as distinct from its single state) which represent its relevant physical, environmental, social, economic, or institutional characteristics. These advances have been accompanied by an increase in the models' complexity (Couclelis, 1997; White & Engelen, 1997).

This modification has allowed CA models to be linked both conceptually and practically with GIS. Since the CA is running on an inhomogeneous cell space (essentially identical to what would be found in a raster GIS), the CA may be thought of as a sort of dynamic GIS (Batty & Xie, 1994). At present, however, CA models developed in GIS remain simple, because GIS do not yet provide operators with sufficient flexibility to define complex CA transition rules, and in addition they lack the simulation engines needed to run complex models at practical speeds. The more practical approach is to couple GIS to special-purpose CA software modules, and possibly other models as well. White et al. (1997) have developed several CA and CA-based integrated models designed as prototypes of Spatial Decision Support Systems for urban and regional planning and impact analysis (demos of several of these models can be downloaded from <http://www.riks.nl/RiksGeo/freestuff.htm>).

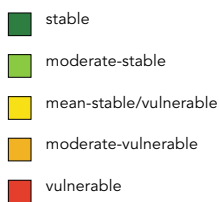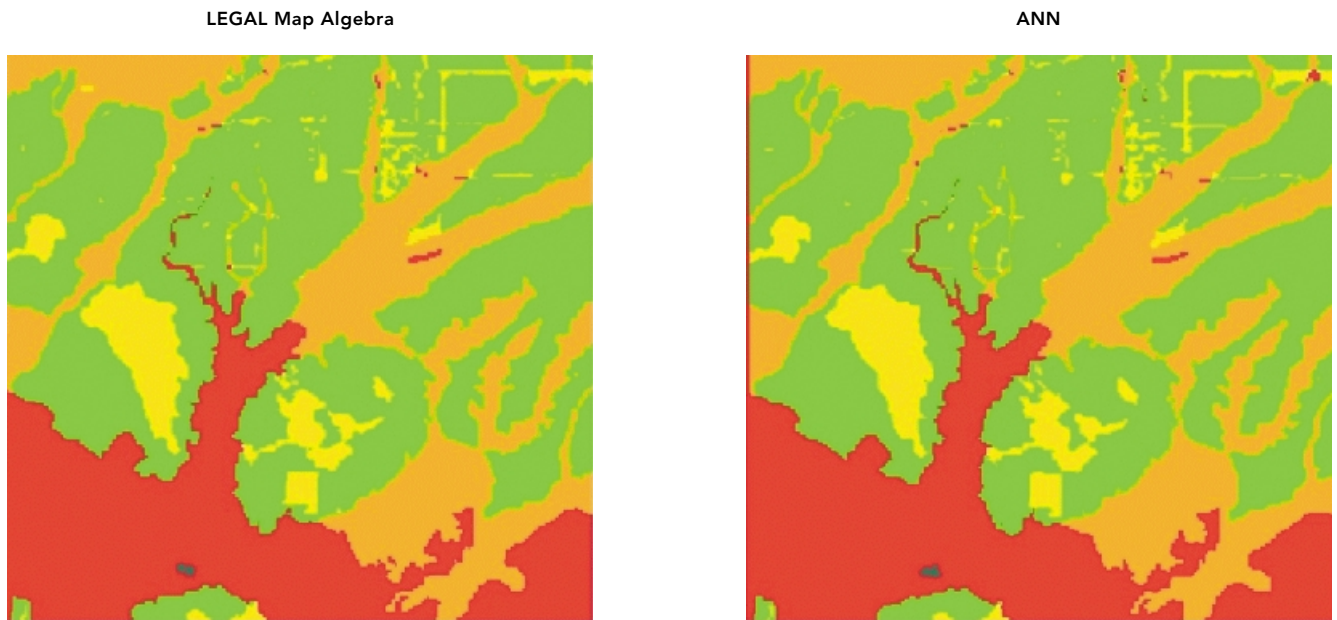## In conclusion: geocomputation as a set of effective procedures

This survey has examined some of the main branches of research in geocomputation, and we conclude the paper with an attempt to provide a unified perspective of this new research field.

We propose that a unifying perspective for geocomputation is the emphasis on *algorithmic* techniques. The rationale for this approach is that the emergence of data-rich spatial databases motivated a new set of techniques for spatial data analysis, most of them originally proposed under the general term "artificial intelligence", such as neural networks, cellular automata, and heuristic search.

Since there are fundamental differences in the perspectives of the set of techniques used by geocomputation, the only unification per-

Figure 8

Neural network based inference.

LEGAL Map Algebra	ANN

- ■ stable
- ■ moderate-stable
- ■ mean-stable/vulnerable
- ■ moderate-vulnerable
- ■ vulnerable

ANN Modeling			Modeling via LEGAL Map Algebra						User exactness (%)	Inclusion error
	Background	Stable	Moderate-stable	Mean-stable/vulnerable	Moderate-vulnerable	Vulnerable	Sum. Xi	Xi/Sum. Xij		
Background	0	0	44	8	16	142	210	0.47	0.00	100.00
Stable	0	15	0	0	0	6	21	0.5	71.43	28.57
Moderate-stable	1	0	19861	434	782	159	21237	47.26	93.52	6.448
Mean-stable/Vulnerable	0	0	473	2598	17	6	3094	6.88	83.97	16.03
Moderate-vulnerable	0	0	671	26	10166	106	10969	24.41	92.68	7.32
Vulnerable	0	2	219	13	122	9053	9409	20.94	96.22	3.78
Sum. Xj	1	17	21268	3079	11103	9472	44940	100.00		
Xj/Sum. Xij	0.00	0.04	47.33	6.85	24.71	21.08	100.00	41693.00		
Producer exactness (%)	0.00	88.24	93.36	84.36	91.56	95.58				
Omission error (%)	100.00	11.76	6.62	15.62	8.44	4.42				
			K % = 91.89					Overall Exactness (%) = 92.8		

ANN = Artificial Neural Network; Sum. = summation; LEGAL = Map Algebra Language.

spective is the *computational* side: such techniques can be thought of as a set of *effective procedures* that, when applied to geographical problems, are bound to produce results. Whatever results are obtained need to be interpreted in light of the basic assumptions of these techniques, and it may be extremely difficult to assign any traditional "statistical significance" criteria to them.

Therefore, the authors propose a tentative definition: "Geocomputation is the use of a set of effective computing procedures to perform spatial data analysis, whose results are dependent on the basic assumptions of each technique and therefore are not strictly comparable".

According to this view, geocomputation emphasizes the fact that the structure and data dependency inherent in spatial data can be used as part of the knowledge-discovery approaches, and the choices involve theory as well as data. This view does not deny the importance of the model-based approaches, such as the Bayesian techniques based on Monte Carlo simulation for the derivation of distribution parameters on spatial data. In fact, in this broader perspective, the use of Bayesian techniques that rely on computationally intense simulations can be considered a legitimate part of the geocomputational field of research.

In conclusion, what can public health researchers expect from geocomputation? When used with discretion, and always bearing in mind the conceptual basis of each approach, techniques such as GAM, local spatial statistics, neural nets, and cellular automata can be powerful aids to a spatial data analysis researcher, attempting to discover patterns in space and relations between its components.

We hope this article serves as inspiration to health researchers and that it will have broadened their notions about what is possible in spatial data analysis.

Acknowledgments

Further reading

For readers interested in more information on geocomputation, we provide a set of references, organized by topics. We suggest that prospective readers begin with Longley (1998) and then proceed to their specific area of interest.

References

ABRAHART, B.; BENENSON, I.; BENWELL, G.; DIBBLE, C.; CALDWELL, D.; GAHEGAN. M.; HARTNETT, J.; LEES, B.; LONGLEY, P.; MACMILLAN, B.; PASCOE, R. & PULLER, D., 2000. *Geocomputation Conference Series Home Page.* 17 April 2000 <http://www.ashville.demon.co.uk/geocomp/index.htm>.

ANSELIN, L., 1995. Local indicators of spatial association – LISA. *Geographical Analysis*, 27:91-115.

ANSELIN, L., 1996. The Moran scatterplot as ESDA tool to assess local instability in spatial association. In: *Spatial Analytical Perspectives on GIS* (M. Fisher, H. J. Scholten & D. Unwin, ed.), pp. 111-126. London: Taylor & Francis.

BAILEY, T. & GATTRELL, A., 1995. *Spatial Data Analysis by Example*. London: Longman.

BATTY, M. & XIE, Y., 1994. From cells to cities. *Environment and Planning B*, 21:531-548.

CÂMARA, G., 2000. *Geoprocessamento: Teoria e Aplicações.* 11 April 2000 <http://www.dpi.inpe.br/gilberto/livro>.

CARVALHO, M. S., 1997. *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. Tese de Doutorado, Rio de Janeiro: Programa de Engenharia Biomédica, Instituto Alberto Luiz Coimbra de Pós-graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro.

COUCLELIS, H., 1997. From cellular automata to urban models: New principles for model development and implementation. *Environment and Planning B: Planning & Design*, 24:165-174.

D'ORSI, E. & CARVALHO, M. S., 1998. Perfil de nascimentos no Município do Rio de Janeiro – Uma análise espacial. *Cadernos de Saúde Pública*, 14: 367-379.

GETIS, A. & ORD, J. K., 1996. Local spatial statistics: An overview. In: *Spatial Analysis: Modelling in a GIS Environment.* (P. Longley & M. Batty, ed.), pp. 261-277, New York: John Wiley.

GOPAL, S., 1998. Artificial neural networks for spatial data analysis. In: *NCGIA Core Curriculum in Geographic Information Science* (M. Goodchild & K. Kemp, ed.), Santa Barbara: National Center for Geographical Information and Analysis. 15 May 2000 <http://www.ncgia.ucsb.edu/giscc/units/u188/u188.html>.

GOPAL, S. & FISCHER, M., 1996. Learning in single hidden layer feedforward neural network models: Backpropagation in a spatial interaction modeling context. *Geographical Analysis*, 28:38-55.

HEWITSON, B. C. & CRANE, R. G., 1994. *Neural Nets: Applications in Geography.* Dordrecht: Kluwer Academic Publishers.

KANELLOPOULOS, I., 1997. *Use of Neural Networks for Improving Satellite Image Processing Techniques for Land Cover/Land Use Classification.* European Commission, Joint Research Centre. 10 April 2000 <http://ams.egeo.sai.jrc.it/eurostat/Lot16-SUPCOM95/final-report.html>.

LEONDES, C., 1997. *Image Processing and Pattern Recognition.* Neural Network Systems Techniques and Applications Series, v. 5. New York: Academic Press.

LONGLEY, P., 1998. *Geocomputation: A Primer.* New York: John Wiley and Sons.

MEDEIROS, J. S., 1999. *Bancos de Dados Geográficos e Redes Neurais Artificiais: Tecnologias de Apoio à Gestão do Território.* Tese de Doutorado, São Paulo: Departamento de Geografia, Universidade de São Paulo.

OPPENSHAW, S., 1993. Modelling spatial interaction using a neural net. In: *GIS Spatial Modeling and Policy* (M. M. Fischer & P. Nijkamp, ed.), pp. 147-164. Berlin: Springer.

OPPENSHAW, S., 1997. Developing GIS-relevant zone-based spatial analysis methods. In: *Spatial Analysis: Modelling in a GIS Environment* (P. Longley & M. Batty, ed.), pp. 55-73, New York: John Wiley.

OPPENSHAW, S., 1998. Building automated Geographical Analysis and Exploration Machines. In: *Geocomputation: A Primer* (P. A. Longley, S. M. Brooks & B. McDonnell, ed.), pp. 95-115, Chichester: MacMillan Wiley.

OPPENSHAW, S. & ABRAHART, R. J., 1996. Geocomputation. In: *Proceedings of the 1st International Conference on GeoComputation* (R. J. Abrahart, ed.), pp. 665-666, Leeds: University of Leeds.

OPPENSHAW, S. & ABRAHART, R. J., 2000. *Geocomputation.* London: Taylor and Francis.

OPPENSHAW, S. & OPPENSHAW, C. 1997. *Artificial Intelligence and Geography.* New York: John Wiley.

ORD, J. K. & GETIS, A., 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27:286-306.

O'SULLIVAN, D., 1999. Exploring the structure of space: Towards geo-computational theory. In: IV International Conference on GeoComputation, *Proceedings.* 12 May 2000 <http://www.geovista.psu.edu/geocomp/geocomp99>.

SPOSATI, A., 1996. *Mapa da Exclusão/Inclusão Social da Cidade de São Paulo.* São Paulo: EDUC.

TAKEYAMA, M. & COUCLELIS, H., 1997. Map dynamics: Integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical Information Science*, 11:73-91.

TERUIYA, R.; CÂMARA, G. & MONTEIRO, A. M. V., 1999. *An Application of GAM for Health Data in Rio de Janeiro.* Internal Report. São José dos Campos: Instituto Nacional Pesquisas Espaciais.

TOBLER, W. R., 1979, Cellular geograph. In*: Philosophy in Geography* (S. Gale & G. Olsson, ed.), pp. 379-386, Dordrecht: D. Reidel Publishing Company.

TURTON, I., 1998. *The Geographical Analysis Machine.* Centre for Computational Geographics, University of Leeds. 29 March 2000 <http://www.ccg.leeds.ac.uk/smart/gam/gam.html>.

WHITE, R. & ENGELEN, G., 1993. Cellular dynamics and GIS: Modelling spatial complexity. *Geographical Systems*, 1:2.

WHITE, R. & ENGELEN, G., 1997. Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B*, 24:235-246.

WHITE, R.; ENGELEN, G. & ULJEE, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B*, 24:323-343.

WINTER, K. E. & HEWITSON, B. C., 1994. Self organizing maps – Application to census. In: *Neural Nets: Applications in Geography* (B. C. Hewitson & R. G. Crane, ed.), Dordrecht: Kluwer Academic Publishers.