None of the motivations listed by the authors, apart from what they refer to as the "abstract" searching for patterns, are grounded in asking scientific or scholarly questions. Certainly the identification of disease patterns is an important first step, but without carefully thinking through the nature of the disease and how it is spread, combining maps of various outcomes and characteristics can be both misleading and also dangerous, to the extent that it leads to misdirection of funds to attack diseases in particular ways.

As they themselves acknowledge, the major motivating force behind the use of many of these techniques is simply "the emergence of computerized data-rich environments" and the availability of "affordable computational power." My experience has been similar, and as a scientist I am very skeptical of such motivation. It leads to researchers confusing their units of analysis, slipping between individuals, communities, and regions, or combining them in the same maps, and making false inferences across scales. The determinants of cases and the determinants of incidence rates are often quite different (Rose, 1985). I shudder to think that we are training young scholars who are driven by a mere fascination with technology and who have forgotten how to frame clear, important questions and design studies to answer them.

For instance, the example they give of regressing percentage of people over 70 years on percentage of houses with proper sewage facilities is based on the problematic assumption that the populations and sewage disposal of urban neighborhoods have been stable over time. Older people may have grown up in the countryside and only moved to those urban areas as adults (poverty often being associated with old age): thus migration patterns may be the major determinants of percentage of people over 70. Or increasing population densities may have interacted with sewage disposal methods to create problems over time; in this case it is most important to understand demographic and sewage production and disposal dynamics of those urban neighborhoods over the past seven decades. It seems to me that before jumping into the computational techniques, the researchers need to propose a clear theoretical framework and a biological and socially substantive logic which leads to specific questions to be answered in the research.

A further concern I have with the focus on these newer techniques of analyses is that researchers sometimes ignore the sources and quality of data, how they were collected, and their real spatio-temporal distribution. Data collected from referral hospital and health center records, based on diagnostic tests and questionnaires with a wide range of sensitivities, specificities, and precision cannot simply be lumped together with satellite data to produce meaningful information. Sometimes simple hand-drawn maps combined with intensive community survey or focus group work may be what is needed most.

Health researchers are facing important and often unprecedented questions in the 21st century. How can we create sustainably healthy societies? What are the relationships between economic policy, environmental change, and human health? How might global warming affect changes in regional disease patterns? I have no doubt that geocomputational techniques can make important contributions to answering these questions. The authors recognize that *"the results are dependent on the basic assumptions of the technique"*, and that researchers should use these techniques *"with discretion, and always bearing in mind the conceptual basis of each approach"*. I only wish they had spent a little more time and space exploring those assumptions and concepts, to enable those of us who are novices to more carefully select those techniques most suitable to the questions we seek to answer.

McMICHAEL, A. J., 1999. Prisoners of the proximate: Loosening the constraints on epidemiology in an age of change. *American Journal of Epidemiology,* 149:887-897.

ROSE, G., 1985. Sick individuals and sick populations. *International Journal of Epidemiology,* 14:32-38.

Flávio Fonseca Nobre

Programa de Engenharia Biomédica, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

I thoroughly enjoyed the paper by Drs. Gilberto Câmara and Antonio Miguel Viera Monteiro and hope that more researchers will be enticed by the main ideas presented above. I hope to see a stronger cross-fertilization of this emerging interdisciplinary field, connecting the use of so-called intelligent systems to spatial health data analysis.

The difficult task is to sum up and provide a brief discussion of this paper. Reading the first part I learned a term with which I had little or no familiarity – geocomputation – posited by the authors as a new interdisciplinary field using computer-intensive methods, including neural networks, fuzzy logic, genetic algorithms, and cellular automata for spatial data analysis.

The study of spatial and spatial-temporal epidemiological data is a timely issue which is driven by both decreasing technology costs and increasing availability of information. For example, it is becoming increasingly possible to access georeferenced public health data in a speedy manner through the Internet for analyzing and merging with other information. Several models and methods to work with spatial health-related data have cropped up in the literature in the last twenty years. Most of these were developed in other areas, like geostatistics, which originated in the mining industry and was later borrowed to help understand and explain the spatial distribution of health events. As is common in many applied sciences, the method is first introduced in an intuitive way, and once the heuristic results prove encouraging, there is major involvement by mathematical and statistical theorists to get the technique soundly established. The wave of progress following this pattern continues with Câmara and Monteiro's paper, presenting a basic review of existing possibilities for the use of different computing procedures to perform spatial health data analysis.

On the application side, I would partially support the motivating statement of the paper citing Oppenshaw (1996), that *"many end users merely want answers to fairly abstract questions ..."*. However, some care should be exercised here. Some twenty years ago I heard in a Brazilian workshop on statistical methods for epidemiologists, particularly on multiple regression, that the basic concepts are cumbersome and difficult to be understood by public health workers, and that they should be more involved in collecting good data to be analyzed by the "foreigners", i.e., specialists in statistics. Obviously, the authors of the paper would not wish us to merely engage in using these "black box" tools (which are well understood in the artificial intelligence community) but rather, that we begin close collaboration to both further the knowledge of these new methods and convince ourselves that they could be included in the analytical tool box of epidemiologists and public health professionals.

The authors provide examples of real analyses in the hope of giving a genuine applied flavor to the methods reviewed. I wish to make some comments on these applications. The first concerns the use of the GAM (Geographical Analysis Machine) to find clusters in data that are originally areal data. Although the authors emphasized that it is only an example, there is no mention of the large differences in area sizes and population distribution in Rio de Janeiro's districts, which I believe could substantially influence the results. If one uses some sort of altered or transformed data set, one must interpret it with caution and be certain that the alteration is stated clearly to avoid misuse by newcomers to this field of spatial analysis research.

My other point concerns Section 3, on neural networks and geographical analysis, where the authors present a classification problem to produce a map of environmental vulnerability. One of the most fundamental aspects of neural network modeling is the requirement of "plenty of training data", which is properly identified in the paper. Neural networks are "adaptive computing" in that they learn from data to build a model. Therefore, the training data set should contain all examples of possible sets of explanatory and outcome variables if one uses the workhorse of neural network modeling: a feed forward network with a back propagation algorithm. Users interested in applying this new technology should be aware of this important aspect. In addition, analysts must be willing to both tolerate the large amount of time for training and have a "black box" model which unfortunately does not provide the ability to explain the reasoning used to arrive at a result. This still limits the usefulness of this technique in some areas, particularly when one is interested in measuring the effects of input variables rather than prediction.

Recent developments in computing performance have provided a wealth of opportunities for advancement of new analytical approaches to spatial data analysis. These include the increasing use of Bayesian thinking, particularly with the introduction of the Monte Carlo Markov Chain (MCMC) approach to tackle intractable integrals. For the unfamiliar reader, the paper provides a brief introduction to various techniques. Some of these techniques were derived from the so-called intelligent systems, and it is the hope of the authors, and also mine, that they may assist our capability to convert data into information.