tistical model-building process and not circumventing it.

This view of the role of geocomputation (which I freely admit may be narrower than that held by the authors) leads me to my third point, which relates to various concerns over the practical use of some kinds of geocomputational algorithms. The process of model-building is ideally both interactive and iterative. The analyst needs to try out ideas on the data, and this requires exploratory tools that can be guided or steered towards particular chosen ends or hypotheses. At present, many geocomputational algorithms appear too much of a "black box" to make this possible. The very nature of the algorithms makes it difficult to provide simple, readily understood control parameters which enable them to be "steered" towards answering particular questions which one might wish to ask of the data. In a sense they provide an answer in the absence of a question. This detracts from their value as exploratory tools for the model builder. In that sense what is often termed "artificial intelligence" might be better referred to as "artificial un-intelligence". There is also the problem of whether such techniques produce *robust* results as opposed to ones which are pure artifacts of the data. I appreciate that traditional notions of statistical significance and standard error cannot and perhaps should not be looked for in relation to these algorithms and that different algorithmic approaches will naturally reveal different aspects of the data. However, the *sensitivity* of the results from any one of them (e.g. to starting conditions or in repeated application to various subsets of the data) needs to be investigated and is often not. If the data are to be mined then we need to establish whether a vein of gold has been found or a vein of fool's gold, and currently the algorithms are weak on the diagnostics that would enable us to measure that.

In summary I do not wish to appear as a dogged defender of existing spatial statistical models and methods. I am well aware how deficient many of those are. For example, traditional spatial models largely involve space in terms of glib abstractions – "distances", "boundaries", and "edge effects". Of course in reality the areas over which analyses are being conducted are vastly complex, criss-crossed with natural boundaries such as forests, rivers, or ranges of hills, or else human constructions such as roads, industrial estates, recreational parks, and so on. Many commonly used spatial statistical methods and models should be viewed in the cold light of their spatial simplicity compared with what we know to exist in geographical reality and upon which data are now available through GIS and remote sensing. Humility would indeed be wise for anyone defending such models, and it is useful to be reminded of that and presented with some novel algorithmic approaches in this paper which may assist to address it. Therefore I welcome new and improved algorithms for exploratory spatial analysis of health data capable of exploiting the complexity of data and of geography. If geocomputation matures to offer that, then I am very comfortable with using it. However, I think we should be cautious about exaggerating its potential. Data analysis in general involves more than methods; it depends on contextual knowledge of the phenomenon under study, the objectives of the analysis, the quality and origins of the data, and the judgment and experience of the analyst. Because of that there is a long-standing resistance among applied statisticians to the suggestion that what they do is just another branch of mathematics. It would not be surprising to find them equally resistant to the suggestion that it should become a branch of computer science. I also doubt that geographical health and environmental research would necessarily benefit if that were to become the case.

David Waltner-Toews

Department of Population Medicine, University of Guelph, Guelph, Canada.

Epidemiologists, after several decades of favoring non-spatial statistical models, are increasingly realizing the importance of understanding socioeconomic and ecological contexts in the interpretation of disease patterns in populations (McMichael, 1999). As the questions we are asking change in both scope and nature, input from scholars in non-health fields with expertise in studying spatial patterns, such as this paper, are a welcome addition to the health literature.

The authors state that their intent is to *"draw the attention of the public health community to the new analytical possibilities offered by geocomputational techniques"*. While the introduction of these techniques to health researchers is laudable in and of itself, I would like to throw out some cautionary notes, based on some experience working with interdisciplinary teams where these techniques have been proposed.

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