Maria de Fátima Militão de Albuquerque

Núcleo de Estudos em Saúde Coletiva, Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Brasil.

What do public health researchers expect of geocomputation?

The article by Gilberto Câmara and Antônio Miguel Vieira Monteiro is highly interesting and objective. While it introduces the concept of geocomputation in a clear and didactic way, demonstrating its potential as a tool for analyzing spatial data, it also invites the reader to answer the question at the end with the same clarity as the authors: what do we expect from geocomputation?

Epidemiology seeks to improve the methods and techniques that allow it to describe, explain, and predict health and disease phenomena in populations, with a view towards prevention. Therefore it plays a fundamental role in public health. From this perspective, the analysis of spatial distribution of diseases has contributed to the production of knowledge in the field and should not be seen as a "second-class" replacement for studies focusing on the individual as the unit of analysis (Susser, 1994a).

Depending on the problem one wishes to solve, the ecological approach has its indications and specificities. Thus, studies can focus on mapping the geographical distribution of diseases with the identification of spatial clusters of cases and the analysis of associations between the incidence of diseases/events and environmental or contextual exposures related to the collective sphere.

How can geocomputation help improve such studies? We must first ask if we really understand what is being offered to us.

Reading the article was certainly enlightening, providing us with the scope of development of techniques and the analytical possibilities offered by the various methods. The authors facilitated an understanding of the concepts by giving a detailed development of the theme through examples of health-related and environmental situations. For example, we are left with the idea that the four types of approaches presented by the authors have different premises and objectives but can be viewed as complementary.

Thus the use of GAM (the Geographic Analysis Machine) is capable of revealing clusters of events/diseases and constructing maps when the excess rates found are statistically significant. For example, this would be a useful technique for detecting priority areas for public health interventions, and would not aim at helping explain the occurrence of phenomena.

Meanwhile, techniques for the detection of "spatial autocorrelation", measured by the Moran coefficient or through semi-variograms, would detect dependence between geographically proximate events, *"explicitly considering the possible importance of their spatial arrangement in the analysis or interpretation of the results"* (Bailey & Gatrell, 1995). There are thus specific indications for this type of research, for example: when one's point of departure is the hypothesis that the event at issue is generated by environmental factors that are difficult to detect at the individual level.

The other two approaches described by the authors involve more sophisticated techniques, incorporating functions aimed at contemplating the complexity of the phenomena. The authors explain that an Artificial Neural Network (ANN) can be used as an exploratory tool in data-rich environments and that it is capable of integrating different types of nature in a single geographic data base using Geographic Information Systems (GIS) technology. The information to be introduced into the model should be chosen by the researcher, which obviously presupposes the existence of an underlying theoretical basis.

Meanwhile, cellular automata move even further in the sense of incorporating dynamic elements into the models. These models "would free us from static views of space" and would be capable of representing the change in space over time as the product of human actions.

We may be closer to achieving the ambitious objective identified by Susser (1994b) (speaking of the logic of the ecological approach): to understand how the context affects the health of individuals and groups. In other words, it appears increasingly possible to develop studies that reveal the effects not only of the structural elements of space but also those of the *processes*, not perceptible within the sphere of studies whose unit of analysis is the individual. Hence epidemiology turned to critical geography for the concept of "socially organized space".

Finally, the authors point out that computational technology for solving health problems should always be applied keeping in mind the conceptual underpinnings of each approach. This concern has its counterpart in the health field. The conceptual basis to be considered in studies should be related to the theoretical and methodological issues of public health. This underscores the need for an interdisciplinary dialogue, where the respective challenge for the public health researcher is to guarantee the epidemiological content of the studies, allowing for better knowledge of the target phenomenon, prediction of new occurrences, and the organization of interventions aimed at prevention.

BAILEY, T. C. & GATRELL, A., 1995. *Interative Spatial Data Analysis*. Longman: Scientific & Technical.
SUSSER, M., 1994a. The logic in ecological I: The logic of analysis. *American Journal of Public Health*, 84:825-829.
SUSSER, M., 1994b. The logic in ecological II: The logic of design. *American Journal of Public Health*, 84:830-835.

Claudio J. Struchiner

Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

First I would like to express my appreciation to the authors for this impressively wide-ranging paper. It is a review paper that provides an introduction to geocomputation techniques, i.e., computer-intensive techniques for knowledge discovery in physical and human geography. The authors seem to favor the view that this new interdisciplinary area is to be distinguished from the simple extension of statistical techniques to spatial data. My comments are motivated by questions I have posed to myself after reading their review: How do such methods compare to established techniques? What are their advantages and disadvantages? What are their ranges of applications? Do the new techniques challenge or extend any of the existing paradigms in data analysis?

The computational dimension appears to be the common denominator of the techniques described in this review and goes into the definition of the key concept at stake, geocomputation. Faster and more powerful computers and advances in software engineering have had a profound impact on all areas of statistics. Bootstrap and Monte Carlo Markov Chain (MCMC) methods, for example, allow the estimation of parameters in richer and more realistic model-based representations of natural phenomena, thereby freeing the imagination of the scientific community. In this context, the boundaries of statistical models and statistical theory have been extended, while preserving the current paradigms, i.e., good statistical thinking is based on solid philosophical principles.

Algorithmic thinking also plays an important role in other areas of science. Complex systems can be generated through the use of very simple building rules, which resemble the functioning of DNA chains. In this context, computer-intensive algorithmic techniques are intimately related to the mechanisms of pattern formation that supposedly occur in nature. In opposition, the procedures under the heading of geocomputation also seek to uncover pattern formation, but their search mechanisms are general in nature and do not bear any relationship to the various possible mechanisms that generate those spatial patterns.

In my view, the geocomputational methods reviewed in this paper do not share the same principles as these extensions. These algorithmic techniques appear to be a computerized version similar, in spirit, to a once very fashionable set of techniques developed by J. Tukey and known as Exploratory Data Analysis. Other statistical techniques put together under different headings such as Data-Driven Procedures and Data Mining attempt to answer similar questions raised here, i.e., "Are there any patterns, what are they, and what do they look like?"

The literature on quantitative methods has acknowledged, at least since the beginning of this century, the existence of two dimensions in research practice, i.e., exploratory versus analytical. For example, R. Ross opposed the concepts of *a priori* versus *a posteriori* pathometry in his Theory of Happenings. Most textbooks make a distinction between descriptive and analytical epidemiology. The debate seems endless and can be naively put by such questions as: "Are there purely descriptive studies? Without knowing what one is looking for, how can one tell when one has found it? If there is some previous knowledge or intuition of a subject, why not make it explicit in a model and see how the available empirical evidence modifies this knowledge or intuition? Do pattern-discovery algorithms carry some sort of built-in intelligence?"

Therefore, by analogy with other computer-intensive techniques mentioned above, one could wonder whether geocomputation, and other modern exploratory data analysis techniques, could benefit from incorporating a causal structure or more specific pattern-formation mechanisms.