GEOSTATISTICS: A COMMON LINK BETWEEN MEDICAL GEOGRAPHY, MATHEMATICAL GEOLOGY, AND MEDICAL GEOLOGY

PIERRE GOOVAERTS*

BioMedware Inc, Ann Arbor, 48103, MI, USA
goovaerts@biomedware.com

Geostatistics provides a set of statistical tools for the analysis of data distributed in space and time. It allows the description of spatial patterns in the data, the incorporation of multiple sources of information in the mapping of attributes, the modeling of the spatial uncertainty and its propagation through decision-making. Since its development in the mining industry, geostatistics has emerged as the primary tool for spatial data analysis in various fields, ranging from earth and atmospheric sciences, to agriculture, soil science, environmental studies, and more recently exposure assessment In the last few years, these tools have been tailored to the field of medical geography or spatial epidemiology, which is concerned with the study of spatial patterns of disease incidence and mortality and the identification of potential "causes" of disease, such as environmental exposure or sociodemographic factor.

Data available for human health studies fall within two main categories: individual-level data (e.g. location of patients and controls) or aggregated data (e.g. cancer rates recorded at county or ZIP code level). Although none of these datasets falls within the category of "geostatistical data" as classically defined in the spatial statistics literature, geostatistics offers a promising alternative to common methods for analysing spatial point processes and lattice data. One of the most challenging tasks in environmental epidemiology is the analysis and synthesis of data collected at different scales and over different spatial supports. For example, one might want to explore relationships between health outcomes aggregated to the ZIP code level, census-tract demographic covariates, and exposure data measured at a few point locations. Geostatistics provides a theoretical framework for performing the various types of changes of support, while providing a measure of the reliability of the predictions.

This paper provides an overview of geostatistical methods available for the analysis of geological and health data, with a focus on the issue of error propagation, that is how the uncertainty in input data (e.g. arsenic concentrations) translates into uncertainty about model outputs (e.g. risk of cancer). Methods for uncertainty propagation, such as Monte-Carlo analysis, are critical for estimating uncertainties associated with spatially-based policies in the area of environmental health, and in dealing effectively with risks.

Keywords: uncertainty, interpolation, cancer