Computational Simplifications Needed for Efficient Implementation of Spatial Statistical Techniques in a GIS

Daniel A. Griffith* and Zhiqiang Zhang†

*Department of Geography, Syracuse University
Syracuse, NY 13244-1020
†Phone.com, 800 Chesapeake Drive
Redwood City, CA 94063

Abstract

This paper contributes to the ongoing debate about which spatial analysis functions should be coupled with a GIS by identifying research problems that need to be solved before a richer toolbox of spatial statistical techniques can be implemented in a GIS. Three general problem areas are addressed. The first replaces a sequential ordinary least squares linear regression implementation with a single regression analysis. The second establishes the effective sample size for a single variable in a georeferenced data set, a result useful when calculating confidence intervals for means. The third establishes the effective sample size for pairs of variables in a georeferenced data set, a result useful when calculating the significance of correlation coefficients. These three general problems allow four more specific research problems to be identified that are in need of definitive solutions before a richer toolbox of spatial statistical techniques can be relatively easily implemented in a GIS. Their complete solutions will involve both empirical assessments and simulation experiments. These four problems are represented by four principal equations posited in this paper, equations that offer considerable computational simplification for the implementation of spatial statistical techniques within a GIS. Sufficient evidence in support of them is presented here to allow their implementation at this time on an experimental basis. These equations remove the need for eigenfunction and nonlinear optimization routines, and maintain the standard linear regression technique as the workhorse of a GIS statistical analysis. They also strengthen the inferential basis for a spatial scientist.