SPATIAL CORRELATES AND AREAL INTERPOLATION: A PROPOSED NEURAL NETWORK SOLUTION TO THE MISSING DATA PROBLEM IN AREA-BASED SPATIAL ANALYSIS

David A. Merwin

Department of Environmental, Earth and Ocean Sciences, University of Massachusetts-Boston, Boston, MA 02125-3393, Tel-(617)287-5270, Fax-(617)287-3929, email-david.merwin@umb.edu.

Biography

Assistant Professor of Geography, University of Massachusetts-Boston (2003 -). Research interests: Spatial data analysis, spatial interpolation, GIS.

INTRODUCTION

This study proposes an alternative solution to the "missing data" problem; a scenario that requires the use of areal interpolation procedures to estimate spatially extensive data (count data) for areal units. In spatial analysis, a common statistical dilemma encountered with polygonal boundaries is the "alternative geography" problem (Mrozinski and Cromley, 1999). This problem arises when spatially extensive data is only available in set of areal units (source zones) that is not the desired set of areal units (target zones). If the spatial extents for both sets of units align, common areal interpolation methods, such as areal weighting, can be used to estimate the data values for the intersectional polygons that result (Goodchild and Lam, 1980). When source zones do not spatially align with target zones, the "missing data" problem arises and target zone polygons will likely contain unknown data values. A number of areal interpolation methods have been proposed to address the issue of missing data in area-based statistics, including dasymetric mapping techniques (Eicher and Brewer, 2001) and "intelligent" interpolation methods (Flowerdew and Green, 1992; Goodchild et al., 1992; Flowerdew and Green, 1994). Both methods mentioned above incorporate the use of ancillary spatial data to assist in the interpolation process when missing data exists in the transfer from source zones to target zones.

GOALS AND METHODOLOGY

The purpose of this study is to propose an alternative areal interpolation procedure to solve the missing data problem. The method that is put forward is a neural network approach that estimates missing spatially extensive data for target zones based on spatial correlates that align with the geographic extent of the areal units. Spatial correlates represent a spatial distribution of a phenomenon that is typically correlated with the statistic that is being estimated by the interpolation procedure. Unlike the approach taken for intelligent interpolation procedures, the method proposed in this study does not consider an overlay between source zones and target zones. The proposed method uses a back-propagation artificial neural network to estimate aggregated population values for census block groups in Hartford County, Connecticut. Identical spatial correlates and known population values for another county of similar size are used as a training set for the neural network and resulting model is tested on the census block groups in Hartford County. To determine the accuracy the neural network interpolation procedure, the results from the neural network model are compared to predicted population values derived from multivariate regression.

REFERENCES

Eicher C.L. and Brewer, C.A. (2001) Dasymetric Mapping and Areal Interpolation: Implements and Evaluation, Cartography and Geographic Information Science 28(2), 125-138.

Flowerdew, R. and Green, M. (1992) Developments in Areal Interpolation Methods and GIS, Annals of Regional Science 26, 67-78.

Flowerdew, R. and Green, M. (1994) Areal Interpolation and Types of Data, in S. Fotheringham and P. Rogerson (Eds), Spatial Analysis and GIS, 121-45, London: Taylor and Francis.

Goodchild, M. and Lam, N. (1980) Areal Interpolation: A Variant of the Traditional Spatial Problem, Geo-Processing 1, 297-312.

Goodchild, M., Anselin, L. and Deichmann, U. (1993) A Framework for the Areal Interpolation of Socioeconomic Data, Environment and Planning A 25, 383-97.

Mrozinski, R. and Cromley, R. (1999) Singly and Doubly-Constrained Methods of Areal Interpolation for Vector-Based GIS, Transaction in GIS 3, 285-301.