Geostatistical Prediction/Simulation Of Point-Support Values From Areal Data

Phaedon C. Kyriakidis* and Eun-Hye Yoo

Department of Geography, University of California, Santa Barbara,
Santa Barbara, CA 93106-4060, USA

*Corresponding author: Tel: +1 (805) 893-2266,
Fax: +1 (805) 893-3146,
E-mail: phaedon@geog.ucsb.edu

Biography

Phaedon Kyriakidis is an Assistant Professor at the Department of Geography at the University of California Santa Barbara (UCSB) since 2001. He earned his Ph.D. from the Department of Geological and Environmental Sciences of Stanford University (USA) in 1999, with specialization in geostatistics. From 1999 to 2000, he was a Post-Doctoral Fellow at the Earth Sciences Division of Berkeley National Laboratory (USA), where he currently holds a Faculty Staff appointment. His research interests include the development of geostatistical methodologies for GIScience applications, stochastic environmental modeling, data fusion, and spatial accuracy assessment. He has published papers in Mathematical Geology, Atmospheric Environment, IJGIS, Environmental and Ecological Statistics, and other journals.

Eun-Hye Yoo is a graduate student at the Department of Geography at UCSB. She holds a B.A. and a M.A. degree in Geography from Seoul National University in South Korea. Her research focuses on geostatistical areal interpolation/simulation methods and their applications in GIScience.

Introduction

Spatial downscaling issues constitute an important and vibrant research theme in many scientific disciplines, including of course geography (Atkinson and Tate, 2000). Coarse resolution predictions of general circulation models, for example, need to be downscaled to the watershed level for hydrologic impact assessment studies. Similarly, socioeconomic variables reported on census tracts need to be downscaled to smaller regions for detailed modeling. Area-to-point interpolation is a particular case of change of support (the term support referring to the area/volume associated with each datum or unknown value), whereby areal data are used to predict point values at a set of prediction locations; these prediction locations could constitute (or not) a regular grid. For a recent comprehensive review of statistical approaches addressing the change of support problem, see Gotway and Young (2002).

Routine applications of area-to-point interpolation in geography (Lam, 1983), however, tend to ignore several critical issues: (i) the explicit account of the different areas informed by each datum, (ii) the coherence of predictions: the areal-average of point predictions within any area comprising an areal-average datum, should be equal to that datum (if the latter is assumed error free), and (iii) the uncertainty in the resulting point predictions. Perhaps the only existing method of area-to-point interpolation that satisfies requirements (i) and (ii), but not (iii), is the pycnophylactic interpolation method of Tobler (1979).

In this paper, the spatial prediction of point values from areal data of the same attribute is addressed within the general geostatistical framework of change of support. Several little known, but extremely important, characteristics of the proposed geostatistical framework, and in particular its connections with existing methods of area-to-point interpolation are highlighted and discussed.

Proposed Approach

We adopt a geostatistical framework for area-to-point interpolation that can explicitly and consistently account for the support differences between the available areal data and the sought after point predictions. Most importantly, it is demonstrated that appropriate modeling of all area-to-area, and area-to-point covariance values required by the geostatistical approach yields coherent (mass-preserving or pycnophylactic) predictions. In other words, the areal-average (or areal-total) of point predictions within any arbitrary area informed by an areal-average (or areal-total) datum is equal to that particular datum. In addition, the geostatistical approach offers the unique advantage of providing a measure of the reliability (standard error) of each point prediction.

Several existing approaches for area-to-point interpolation can be viewed within the proposed general geostatistical framework as particular solutions corresponding to particular choices of point-support covariance/variogram models (Kyriakidis, 2003). More precisely, it is shown that: (i) the choropleth map case corresponds to the geostatistical solution under the assumption of spatial independence (pure nugget effect or white-noise covariance) at the point-support level, (ii) variants of the kernel smoothing method correspond to alternative (albeit inconsistent) implementations of the general geostatistical approach, whereby the kernel function linked to the covariance model adopted for interpolation, and (iii) Tobler's smooth pycnophylactic interpolation corresponds to the geostatistical solution, when the variogram model adopted at the point-support level is identified to the free-space Green's functions (linear in 1D or logarithmic in 2D) of the Laplacian differential operator.

The geostatistical prediction procedure outlined above is also extended to the case of stochastic simulation of point-support values subject to area-support data constraints. In this latter case, the simulated point-support values reproduce: (i) a point-support histogram, (ii) a point-support covariance/variogram model (both (i) and (ii) are linked to the corresponding histogram and covariance/variogram model of the area-support data),

and (iii) the area-support data, when simulated point values are aggregated with each areal support. These alternative simulated realizations are therefore consistent with all available pieces of information, and can be used to propagate input parameter uncertainty to environmental or socioeconomic model predictions.

Two case studies using simulated and real data are presented to illustrate the application of the proposed geostatistical approach in practice. Future research directions are discussed, which include accounting for known attribute values (or known spatial derivatives) of the same variable at the point-support level (e.g., boundary conditions), as well as incorporating additional auxiliary variables (at the areal- or point-support level) in the prediction/simulation endeavor.

References

Atkinson, P.M., and Tate, N.J. (2000) Spatial scale problems and geostatistical solutions: A review, *Professional Geographer*, 52, 607-623.

Gotway, C.A., and Young, L.J. (2002) Combining incompatible spatial data, *Journal of the American Statistical Association*, 97, 632-648.

Kyriakidis, P.C. (2003) The geostatistical solution of the area-to-point spatial interpolation problem, *Geographical Analysis* (in review).

Lam, N.S-N. (1983) Spatial interpolation methods: A review, *The American Cartographer*, 10, 129-149.

Tobler, W.R. (1979) Smooth pycnophylactic interpolation for geographical regions, *Journal of the American Statistical Association*, 74, 519-536 (with discussion).