MEASUREMENT AGGREGATION EFFECTS ON THE SPATIAL STATISTICS OF SEA SURFACE TEMPERATURE AND CHLOROPHYLL-a OVER THE NORTH AEGEAN

VASIOS G.K.[1], KYRIAKIDIS P.C.[1,2] and KITSIOU D.[3]

[1] Department of Geography, University of the Aegean, University Hill, Mytilene, Lesvos, 81100, Greece; [2] Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA; [3] Department of Marine Sciences, University of the Aegean, University Hill, Mytilene, Lesvos, 81100, Greece
E-mail: vasios@aegean.gr

ABSTRACT

The effects of measurement scale (spatial resolution) on the statistics derived from environmental data constitute a critical factor in spatial pattern investigation. In this work, we introduce a geostatistical approach to analytically predict the effects of spatial aggregation (upscaling) on spatial statistics (correlation or semivariogram functions) of geospatial data pertaining to environmental variables. More precisely, fine and coarse resolution data are linked through an aggregation function (linear in this work) quantifying the contribution of the former to the latter. The sample semivariogram of the coarse data is then analytically approximated based on a known or estimated semivariogram model for the fine resolution data and that aggregation function.

We illustrate the application of the proposed analytical approach using monthly and seasonal measurements of sea surface temperature (SST) and chlorophyll-a (CHL) concentrations for a decay (period 2005-2014), obtained from the MODIS Aqua sensor over the North Aegean Sea, Greece. More precisely, we study the effect of upscaling on the empirical semivariogram functions of SST and CHL for different pixel sizes. The results show that the proposed geostatistical approach can accurately and consistently predict the effects of spatial aggregation (upscaling) on the semivariogram functions derived at coarser spatial resolutions.

Keywords: upscaling, Modifiable Area Unit Problem, geostatistics, semivariogram, MODIS

1. Introduction

The spatial resolution, or measurement scale, of geospatial data is a critical factor for the elucidation of spatial patterns in environmental variables. In addition, spatial resolution plays a critical role in environmental data integration, where it is often necessary to transform attribute values from one existing set of geographical units (supports) to another (Atkinson and Tate, 2000; Haining, 2003). Numerous studies in the geographical and statistical literature have been carried out seeking to acknowledge, account for, and possibly mitigate the effects of spatial aggregation on the statistical analysis of spatial data (Gotway and Young, 2002).

A major caveat in making inferences with spatial data is the Modifiable Areal Unit Problem (MAUP), which pertains to the extrapolation of spatial statistics from data of one spatial resolution to another. The MAUP, i.e., the notion that generalizations made at one level (spatial resolution) depend on the size and configuration (grouping) of the underlying spatial units, has been long considered “a fundamentally geographical phenomenon” (Openshaw, 1984). Although various attempts towards its solution have been proposed (Fotheringham and Wong, 1991; Cressie, 1996; Holt *et al*., 1996), a general solution to the MAUP does not yet exist (Cao and Lam, 1997).

In this work, we introduce an analytical geostatistical approach for studying the MAUP, within the general framework of change of support (Gotway and Young, 2002; Kyriakidis, 2004). More precisely, fine and coarse spatial resolution attribute values are linked via a measurement function that encapsulates the linear aggregation of point attribute values into their coarse-level counterparts. The MAUP is then formulated analytically as the forward problem of estimating coarse-level statistics, possibly involving multiple variables, based on knowledge of the corresponding, fine resolution statistics.

2. Materials and methods

In the proposed geostatistical approach, any coarse resolution attribute value $\overline{z}(v)$ is specified as an aggregate (e.g., average) of M_v fine resolution attribute values $z(\mathbf{c})$ within the coarse support (pixel) v of the former:

$$\overline{z}(v) = \sum_{m=1}^{M_v} g_v(\mathbf{c}) z(\mathbf{c}) \quad \text{e.g.} \quad \overline{z}(v) = \frac{1}{M_v} \sum_{m=1}^{M_v} z(\mathbf{c})$$

where $\mathbf{c}$ denotes the coordinate vector of the centroid of a fine resolution pixel, and $g_v(\mathbf{c})$ the value of the aggregation function. In this work, $\overline{z}(v)$ is defined as an equally weighted average, and $z(\mathbf{c})$ is deemed representative of the entire fine resolution pixel of area a; hence, aM_v approximates the area of the coarse pixel.

The spatial correlation, or better dissimilarity, of fine resolution attribute values is quantified by a semivariogram model $\gamma(\mathbf{h};\boldsymbol{\theta})$, which is typically an increasing function (e.g., exponential or Gaussian) of the separation vector $\mathbf{h}$ between pixel centroids; here $\boldsymbol{\theta}$ is a model parameter (sill, range, nugget) vector.

The semivariogram $\gamma_v(\mathbf{h})$ of coarse resolution values can then be analytically derived as the convolution of the fine resolution semivariogram $\gamma(\mathbf{h};\boldsymbol{\theta})$ and the aggregation function:

$$\gamma_v(\mathbf{h}) = \frac{1}{M_v M_{v'}} \sum_{m=1}^{M_v} \sum_{m'=1}^{M_{v'}} \gamma(\mathbf{c}_m^v - \mathbf{c}_{m'}^{v'};\boldsymbol{\theta}) - \frac{1}{M_v^2} \sum_{m=1}^{M_v} \sum_{m'=1}^{M_v} \gamma(\mathbf{c}_m^v - \mathbf{c}_{m'}^v;\boldsymbol{\theta}) = \overline{\gamma}(v, v_{\mathbf{h}}) - \overline{\gamma}(v,v)$$

where v and $v_{\mathbf{h}}$ denote two pixels separated by vector $\mathbf{h}$; $\overline{\gamma}(v,v)$ and $\overline{\gamma}(v,v_{\mathbf{h}})$ denote, respectively, averages of semivariogram values between fine resolution pixel centroids spanning the same v or two different v and $v_{\mathbf{h}}$ coarse pixels.

We illustrate the application of the proposed geostatistical approach to the study of aggregation effects using monthly and seasonal data of sea surface temperature (SST) and chlorophyll-a (CHL) for the period 2003-2014, obtained from the MODIS Aqua sensor over the North Aegean Sea, Greece (Fargion and McClain, 2003). As an example, the original ($4\times4km^2$) monthly SST and CHL data for December 2003 are shown in the left column maps of Figure 1. The objective is to analytically predict the semivariogram of upscaled (five- and nine-fold) SST and CHL values, shown in the central (five-fold upscaling) and rightmost (nine-fold upscaling) columns of Figure 1, from the semivariograms of the corresponding fine resolution values and the aggregation scheme.

3. Results

To mitigate the effects of spatial trends, the analysis (upscaling and variography) was conducted using de-trended SST and log(CHL) residuals (not shown).

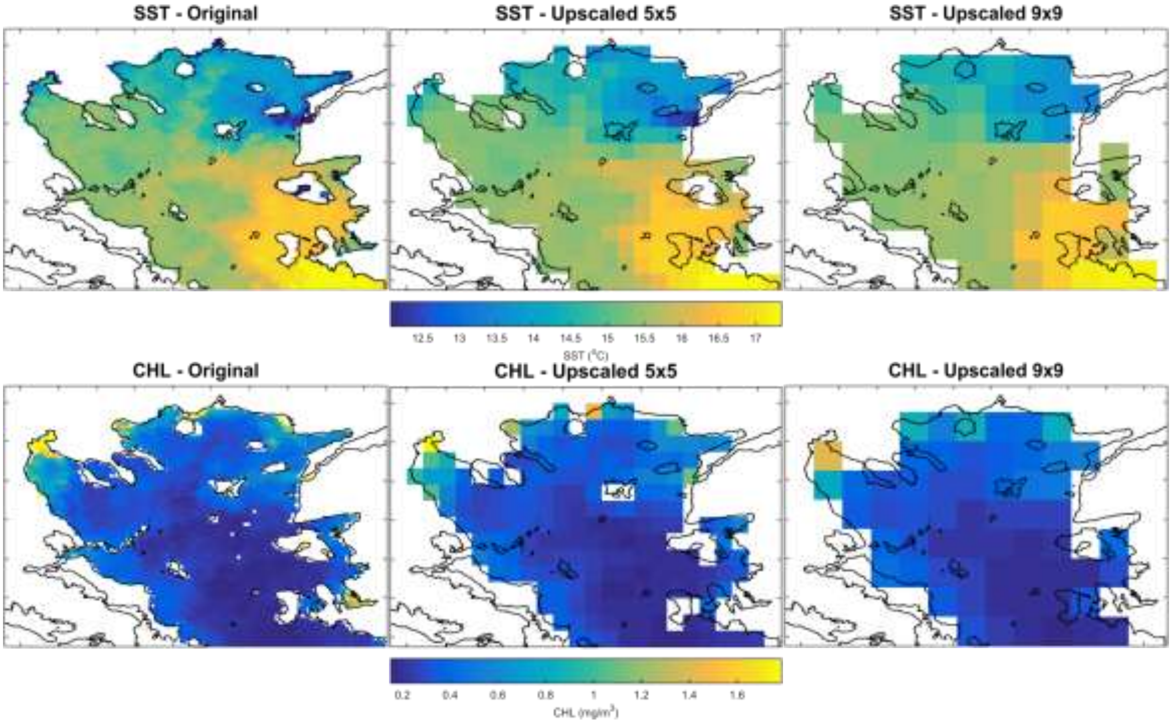

Figure 1: Original (left column) and upscaled (five- and nine-fold) monthly concentrations of SST (top row) and CHL (bottom row) in the North Aegean region for December 2003.

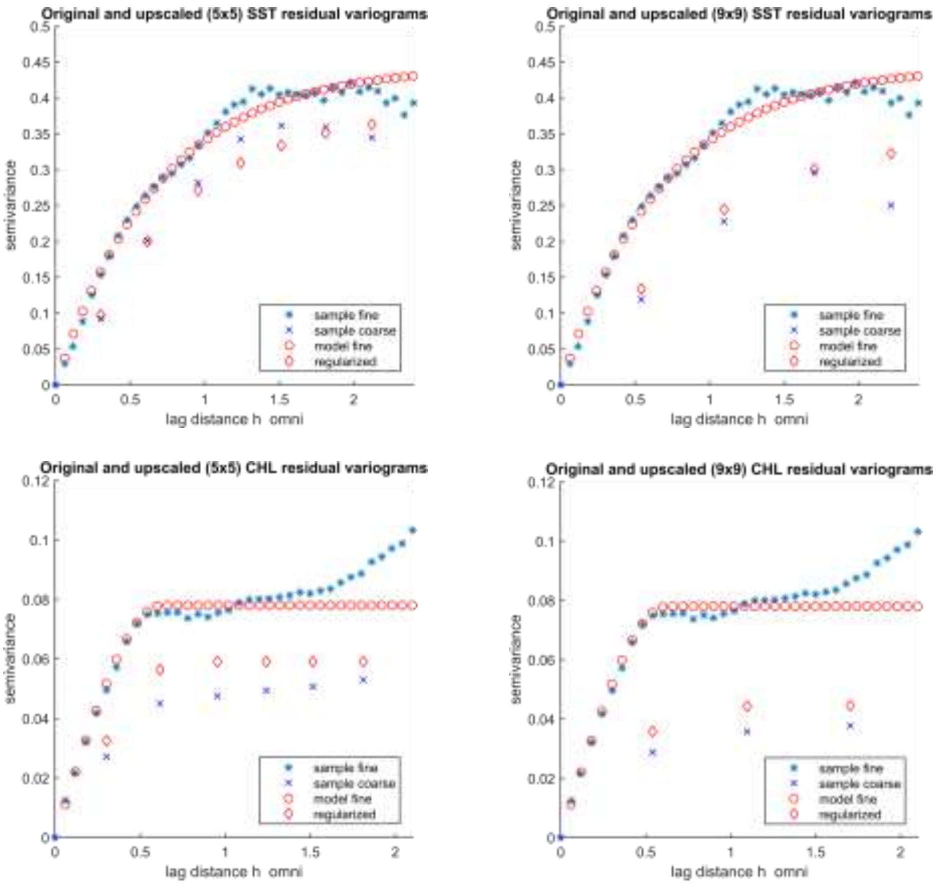

Figure 2: Original and upscaled semivariograms for SST (top row) and log(CHL) (bottom row) residuals for December 2003. Sample semivariograms of original (fine) and upscaled (coarse) residuals are shown in blue and modeled semivariograms in red.

The coarse level semivariograms were computed from upscaled data for two levels of aggregation and compared to their analytically predicted counterparts. The original (fine level) and upscaled (coarse level) semivariograms for December 2003 are shown in Figure 2. It can be seen that the sample coarse level semivariograms are satisfactory predicted from the proposed analytical methodology; any discrepancy stems from the sampling variability of the coarse level semivariograms due to the limited sample size at those resolutions.

4. Conclusions and discussion

A geostatistical approach is proposed in this work to analytically predict the effects of spatial aggregation (upscaling) on the spatial statistics of fine resolution data. The fine and coarse resolution data are linked through an aggregation function, and the sample semivariogram at the coarse data is analytically approximated based on a known or estimated semivariogram model for the fine resolution data and that aggregation function.

We illustrated the application of the proposed geostatistical approach to the study of aggregation effects using data of SST and CHL, for December 2013 as an example, obtained from the MODIS Aqua sensor over the North Aegean Sea, Greece. The results indicated that sample coarse level semivariograms are satisfactory predicted from the proposed analytical methodology. Research efforts are currently underway to extend this approach to address the effects of spatial aggregation on measures of multivariate spatial association.

ACKNOWLEDGEMENTS

This work is part of the research project (2269) “Advances in Geostatistics for Environmental Characterization and Natural Resources Management” with acronym “GEOSTATENV” is implemented within the framework of the Action “Aristeia I” of the Operational Program “Education and Lifelong Learning” (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

REFERENCES

1. Atkinson P.M. and Tate N.J. (2000), Spatial scale problems and geostatistical solutions: A review, Professional Geographer, **52**, 607-623.
2. Cao C. and Lam N.S. (1997), Understanding the scale and resolution effects in remote sensing and GIS, in: Scale in Remote Sensing and GIS, Quattrochi, D.A. and Goodchild, M.F. (Eds.), Lewis Publishers, Boca Raton, 57-72.
3. Cressie N.A.C. (1996), Change of support and the modifiable areal unit problem, Geographical Systems, **3**, 159-180.
4. Fargion G.S. and McClain C.R. (2003), MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003, NASA, Maryland, USA. URL: http://oceancolor.gsfc.nasa.gov
5. Fotheringham A.S. and Wong D.W.S. (1991), The modifiable areal unit problem in multivariate statistical analysis, Environment and Planning A, **23**, 1025-1044.
6. Gotway C. and Young L. (2002), Combining incompatible spatial data, Journal of the American Statistical Association, **97**, 632-648.
7. Haining R. (2003), Spatial Data Analysis: Theory and Practice, Cambridge University Press, Cambridge.
8. Holt D., Steel D.G., Tranmer M. and Wrigley N. (1996), Aggregation and ecological effects in geographically based data, Geographical Analysis, **28**, 244-261.
9. Kyriakidis P.C. (2004), A geostatistical framework for area-to-point spatial interpolation, Geographical Analysis, **36**(3), 259-289.
10. Openshaw S. (1984), The Modifiable Areal Unit Problem, Concepts and Techniques in Modern Geography (CATMOG), **38**, Geobooks, Norwich, UK.