Geostatistics: classical multivariate statistics from a spatial perspective?

Roxana Tesileanu

roxana.te@web.de
INCDS, Romania

November 2017

Contents

1 Abstract

In its essence geostatistics adds georeferenced spatial information to the vector of recorded variables of each individual observation, which represents a spatial location. It further adds into models the space dependent random error. In this paper I state that the classical multivariate techniques like multiple regression analysis and multidimensional scaling, can include this spatial perspective into modeling by treating the X and Y coordinates as "classical" variables, thus reducing the space dependent random error to the "classical" random error of multivariate statistics.

2 Introduction

After reading through introductory chapters of several geostatistics books, I can now state that geostatistics is a departure from classical statistics just because of the sentence that "it takes spatial autocorrelation of observations into account when predicting values for new points". It is not more than that to it. Maybe the most important fact is that geospatial analysis doesn't treat variables as we are used to in classical statistics but uses individual observations (i.e. individual points) and investigates the relationships between them from a spatial perspective [1], [2], [3], [4], [5]. It is adding space as a variable in the vector of recorded variables of each individual observation/point. The highlight of individual points is actually like in object-based classical multivariate statistics. It is treating space as an autocorrelated variable across a series of individual points, and letting all the other variables be "classical".

For me the most important moment in this introductory phase was when I've realized that we don't talk about classical samples of observations, where we concentrate on variables, but instead in geostatistics we concentrate on pairs of observations (i.e. of points). We compute covariances for such pairs, not for the whole sample as in classical statistics. We don't have weights for entire variables, we have weights for individual points caring those values of the variables studied. It is I believe very important the moment when you understand this. Spatial models treat individual points in ways similar to treating individual variables in classical statistics. But we must be aware these are individual points we are talking about, and we very much use distance measures for objects (like the Euclidean distance) like in the multivariate object-based classical statistics.

That being said I think I can use the same matrix calculations as in classical object-based multivariate analyses (where we use objects to predict values for variables), i.e. a n by n MATRIX OF DISSIMILARITIES BETWEEN OBJECTS by means of which we derive variables as LINEAR COMBINATIONS OF THE OBJECTS (Q-mode analyses) - see [6]. And of course classical variable-based multivariate analyses are equally possible - see [**?**] and [7]. The example from Quinn and Keough (2002) at the multiple regression chapter, where the study of Paruelo and Lauenroth is presented in which they've modeled the relative abundance of C3 plants against longitude and latitude is an implementation of this perspective. If the spatial analysis includes a random error with spatial dependence, then why not include in the model the X and Y coordinates as two separate variables and make the random error spatial independent? Adding appropriate variables to a model is the approach used in multivariate statistics to reduce the unexplained variation [6] , [7].

Maybe spatial analysis is just classical multivariate analysis (variable- or object-based, or combined); the important thing is to include X and Y variables in the model.

This doesn't mean I give up the "spatial perspective". I will still use the X-Y coordinate plane to inspect how the residuals from fitted linear models are located. Eventually, delineate more than one target population. And reevaluate the sampling design based on these preliminary conclusions.

I will still keep the classification of Cressie (1993) which delineates three types of geospatial analyses:

- on continuous surfaces (raster)
- on discrete spatial features based on multiple points (lines, polygons)

- on discrete spatial features based on individual points (points).

Note

This document is "under construction". The current version is available on my GitHub profile under the multivariate_analyses project repository: https://github.com/RoxanaTesileanu/multivariate_analyses/blob/master/literature_analysis/geospatial_scala/geostats_multivariate.pdf.

References

[1] N. A. C. Cressie, *Statistics for Spatial Data.* Wiley, 1993.

[2] R. Webster and M. A. Oliver, *Geostatistics for Environmental Scientists*, second edition ed. Wiley, 2007.

[3] E. H. Isaaks and R. M. Srivastava, *Applied Geostatistics.* New York: Oxford University Press, 1989.

[4] Hengl, *A Practical Guide to Geostatistical Mapping*, 2009.

[5] K. Johnston, Ver Hoef, Jay M., Krivoruchko, Konstantin, and Lucas, Neil, "Using ArcGis Geostatistical Analyst," 2003.

[6] G. Quinn and M. Keough, *Experimental design and data analysis for biologists.* Cambridge: Cambridge University Press, 2002.

[7] J. D. Carroll and P. E. Green, *Mathematical Tools for Applied Multivariate Analysis*, revised ed. San Diego: Academic Press, 1997.