Speakers:

Roberto Benedetti (1), Federica Piersimoni (2)

(1) Department of Economic Studies, University “G. d’Annunzio” of Chieti-Pescara, Viale Pindaro 42, Pescara, IT-65127, Italy

(2) Istat, Directorate for Methodology and Statistical Process Design, Via Cesare Balbo 16, Rome, IT-00184, Italy

Title:

IDENTIFYING SPATIAL REGIMES IN ENVIRONMENTAL AND ECOLOGICAL MODELS

Abstract:

This paper is aimed at examining the presence of spatial pattern of parameters when modeling geocoded environmental and ecological data. The empirical evidence often suggests that we are very far from being able to accept the assumption that the parameters vary in space smoothly and continuously as is expected from the typical solution suggested by geographically weighted regression (GWR). The existence of natural boundaries and of areas of similar ecology, delineated by homogenous topography, temperature, and land cover, gives rise to the adoption of a more realistic hypothesis concerning spatial regime regressions that allow the model coefficients to vary between discrete spatial subsets of the data. We observed that the fit of several models is not satisfactory if, in the empirical analysis, we consider the entire data set as one sample. The phenomenon should be analyzed considering zones as belonging to different sub-samples with similar ecological status. The existence of this heterogeneity affects spatially distributed data, and often is neglected in the analysis. We review and compare the performance on real data sets of four different approaches that have been introduced in literature. The findings reveal that, computationally intensive, purely spatial algorithms (Postiglione et al 2010; 2013) accomplish better results than multipurpose solutions as the EM algorithm (Dempster et al, 1977) used to identify finite mixture regressions (Turner, 2000) and binary recursive partitioning based on parametric models (Zeleis et al, 2008).

Keywords: Spatial clusters, latent variables, spatial heterogeneity, finite mixtures, regression trees.

REFERENCES

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society B, 39: 1–22.

Postiglione, P, Andreano, M. S. and Benedetti, R. (2013) Using Constrained Optimization for the Identification of Convergence Clubs. Computational Economics, 42(2):151–174.

Postiglione, P, Benedetti, R. and Lafratta, G (2010) A regression tree algorithm for the identification of convergence clubs. Computational Statistics and Data Analysis, 54(11):2776–2785.

Turner, T. R. (2000) Estimating the rate of spread of a viral infection of potato plants via mixtures of regressions. Applied Statistics, 49(3): 371 – 384.

Zeileis, A, Hothorn, T and Hornik, K (2008). Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics, 17(2): 492–514.