Discussion

Object oriented data analysis under spatial correlation

Jorge Mateu*

Department of Mathematics, University Jaume I of Castellon, Spain

Received 2 September 2013; revised 5 December 2013; accepted 5 December 2013

This is a discussion of the following paper: “Overview of object oriented data analysis” by J. Steve Marron and Andrés M. Alonso.

Keywords: Geostatistics; Manifolds; Point patterns; Spatial correlation.

The authors are to be congratulated on a valuable and thought-provoking contribution in the form of a review of the state of the art in the field of Object Oriented Data Analysis (OODA) which opens big avenues for linking mathematics and statistics. Developments beyond Euclidean spaces are becoming more and more useful for finding certain solutions to statistical problems which have to deal with the combination of massive and complex data. This discussion will address statistical complexity in form of spatial or spatio-temporal structure as an important aspect of OODA and which has been given only little attention/coverage in the literature. Among the many points that this manuscript opens for discussion, I would like to focus on functional data analysis in presence of spatial correlation, geostatistics and point patterns, and on prototypes on manifolds.

In Section 2 the authors present curves as data objects, and consider functional data analysis (FDA). The number of practical problems dealing with functions has increased. In particular, FDA has been used to describe, model, and analyse this kind of function-based data. A wide range of statistical tools (ranging from descriptive data analysis to linear models and multivariate techniques) have been extended to handle functions. The standard statistical techniques for FDA are focused on assuming independence among functions (Ferraty and Vieu, 2006), however, there is a growing interest in modeling correlated functional data in many applications. Assume, for example, that spatially correlated functional data are available and one would like to predict curves in particular locations of a region. One possible attempt is to propose an ordinary kriging predictor for functional data whose parameters, as in the univariate case, are scalars. Other approaches may consider kriging predictors with functional parameters (Giraldo et al., 2010). Integration of OODA into this context needs careful consideration, and adapting PCA in this particular case is yet an open question.

In a related case, spatial point processes describe stochastic models where the events have an associated random location in space (Diggle, 2013; Illian et al., 2008). Second-order properties provide information on the interaction between points in a spatial point pattern. In the context of large datasets (e.g., in forestry problems, we may have locations of thousands of trees for each of hundreds of species, i.e., massive and complex data) adapted PCA could play a role. We can consider an approach based on latent process modeling and PCA in order to obtain a computationally feasible exploratory tool for discovering patterns of association between components of a highly multivariate point process. We can model each component process as a log Gaussian Cox process (Møller et al., 1998) where the latent

*Corresponding author: e-mail: mateu@mat.uji.es, Phone: +34-964728391, Fax: +34-964728429

Gaussian fields are obtained as linear combinations of some independent Gaussian processes. Then the coefficients of these linear combinations are estimated using PCA. It is also possible to consider dimension reduction where the spatial variation in the latent fields for each species may be explained by a moderate number of common latent processes. In this context, the integration of the adapted PCA technique into an OODA framework is an open question.

We can also think of functional marked point processes, where we have functions as marks attached to each location. It would be relevant to see how OODA can be applied in this mathematically very complicated context. Jump-diffusion processes applied to some particular medical problems generate such functional marked point patterns. As also commented by the authors, we sometimes need to go beyond the Euclideanity of the space to obtain better properties. If we follow Arafat et al. (2013), we can generate Lévy processes from marked point processes by using jump-diffusion processes. Then we can build new Markov point processes in a particular nilpotent Lie Group. This process lives in such a space that some mathematical properties (such as new classes of distances or various forms of Markovian dependence) arise and nicely apply in these medical problems.

My final point focuses on the case where the objects are not curves but point patterns themselves. Suppose we are interested in considering ways to measure dissimilarity between objects. Under Euclidianity we can follow Mateu et al. (2013) and consider the spike-time distance and its variants, as well as cluster-based distances and dissimilarity measures based on classical statistical summaries of point patterns. Then we can use such measures to summarize and describe collections of repeated realizations of an object (point pattern) via prototypes or multidimensional scaling (MDS). Extending the link between MDS and PCA, as mentioned by Marron and Alonso (2014), into the context where objects are spatial patterns, and even more interesting, when these patterns or objects are defined on spheres or other manifolds is worth exploring in the near future.

References

Arafat, A., Mateu, J. and Gregori, P. (2013). A family of Markov Lévy processes in nilpotent lie groups. Submitted to *Electronic Communications in Probability*.

Diggle, P. J. (2013). *Statistical Methods for Spatial and Spatio-Temporal Point Patterns* (3rd edn.). CRC Press, Boca Raton, FL.

Ferraty, F. and Vieu, P. (2006). *Non parametric functional data analysis. Theory and practice*. Springer, New York, NY.

Giraldo, R., Delicado, P. and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 66–82.

Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, Wiley, Chichester, UK.

Marron, J. S. and Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal* **56**, 732–753.

Mateu, J., Schoenberg, F. P., Diez, D. M., González, J. A. and Lu, W. (2013). On measures of dissimilarity between point patterns and their applications. *Biometrical Journal*, submitted.

Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics* **25**, 451–482.

