

## Discussion on the paper “Analysis of spatio-temporal mobile phone data: a case study in the metropolitan area of Milan” by P. Secchi, S. Vantini and V. Vitelli

Wenceslao González-Manteiga<sup>1</sup> ·  
Rosa M. Crujeiras<sup>1</sup>

Accepted: 23 April 2015 / Published online: 29 April 2015  
© Springer-Verlag Berlin Heidelberg 2015

### 1 Introduction

First of all, we would like to congratulate the authors for their interesting paper, a nice combination of methodological proposal and data analysis. This work is focused on geo-referenced high-dimensional data, with the purpose of identifying spatial and/or temporal patterns. After a first reading, one may think about possible applications of the proposed methodology in environmental sciences (e.g. pollutants concretion patterns), which certainly opens a fruitful application field. In this area, prediction surfaces can be produced by kriging methods adapted to functional data (see [1], for an application to temperature curves). Although prediction is usually the main goal, the identification of similar spatial and/or temporal patterns provides useful information for characterizing and understanding the behaviour of the (complex) underlying process, as it can be clearly seen in the example analyzed in this work.

We would like also to highlight the (purley nonparametric) innovative route followed by the authors: temporal and spatial variation is addressed by combining a treelet analysis with a previous bagging Voronoi strategy, where data representatives are computed as weighted averaged of the data belonging to each cell. With the algorithm presented in Figure 2, the *aggregation step* offers as outputs a collection of time varying functions and their coupled surfaces.

---

✉ Rosa M. Crujeiras  
rosa.crujeiras@usc.es

Wenceslao González-Manteiga  
wenceslao.gonzalez@usc.es

<sup>1</sup> Department of Statistics and Operations Research, University of Santiago de Compostela, Santiago de Compostela, Spain

In our discussion, we would like to address some issues that may possibly lead to an extension or adaption of the method, as well as raising some concerns about the current proposal.

## 2 Space–time interaction

The final output of the data analysis is provided in Section 4, where Figures 7 and 8 show the selected elements of the reference treelet analysis (TA) basis and the coupled surfaces. The temporal patterns are justified and described in detailed and, as a consequence, the estimated surfaces are also interpreted. Hence, the spatial and the temporal components do not play a symmetric role, which can be clearly noticed in Sect. 2, where  $K$  [the number of latent fields in model (3)] does not depend on the lattice cell.

At this stage, a modification of the proposed model can be easily proposed: let's reverse the role of time and space, and consider the Erlang value at location  $\mathbf{x}$  for a certain time  $t$  as the response of an additive model with spatially varying *reference signals* and time varying coefficients, that is:

$$E_t(\mathbf{x}) = \sum_{k=1}^K D_k(t)\psi_k(x) + \varepsilon. \quad (1)$$

Maybe such a model is not of interest for the project proposal, focused on population dynamics, but it may be useful in other applications. We would like to ask the authors to explain the methodological problems that may arise in analyzing such a model. Can the bagging Voronoi treelet analysis (BVTA) be adapted to this scenario? Is there any room for allowing the parameter  $K$  to vary in space in the original proposal [or in time for model (1)]? Is it possible to estimate space–time signals directly?

Another issue regarding the space–time interaction is the construction of local representatives at different sites, namely  $g_i$  ( $i = 1, \dots, n$ ), which are obtained as a weighted mean with Gaussian isotropic weights, and the same applies for the simulation examples in the supplementary material. So implicitly, the method is assuming an isotropic spatial dependence pattern. We are not sure about the impact of this assumption on the current application, but it would be interesting to have some guidelines on how can the procedure be adapted in order to account for anisotropic patterns, that is, how flexible can be the formulation in step 2.

## 3 Some issues about tuning parameters

The algorithm for the BVTA (Figure 2) depends on some tuning parameters that must be fixed in advanced. Specifically, one must chose  $B$  (number of bootstrap replicates),  $n$  (number of Voronoi elements in the partition),  $J$  (number of functions in the TA orthogonal basis) and  $K$  (which has been commented above). Regarding the bagging strategy (introduced by [2] and analysed by [3]), a large value of  $B$  translates into a higher accuracy of the estimates, as the authors point out in Section 2.2. In the example

as well as in the Supplementary Material, the authors consider  $B = 50$  (and the results seem satisfactory), but how could a practitioner select the *right* number of bootstrap replicates? It would be nice to have a (possibly heuristic) idea on how to deal with this parameter. The same happens for  $n$  (selected by minimizing the total averaged variance) and for  $J$ , although in the first case the simulations support the selection method and in the second case, it seems that this parameter should be *large enough*.

Again, we would like to thank the authors for this nice and interesting work and the Editor for giving us the opportunity to read this paper.

## References

1. Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. *Stoch Env Res Risk Assess* 27:1553–1563
2. Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
3. Buhlmann P, Yu B (2002) Analyzing bagging. *Ann Stat* 30:927–961