CrossMark

EDITORIAL

Advances in spatial functional statistics

Jorge Mateu[1] · Elvira Romano[2]

Abstract This is the editorial letter for the special issue dedicated to *Spatial Functional Statistics*, motivated by the joint VII International Workshop on Spatio-temporal Modelling (METMAVII) and the 2014 meeting of the research group for Statistical Applications to Environmental Problems (GRASPA14), which took place in Turin (Italy) from 10 to 12 September 2014. This special issue summarises and discusses peer-reviewed contributions related to the analysis of functional data showing complex characteristics such as spatial dependence structures. The selection of papers comprises both new methodological proposals and a wide range of applications. In particular, we cover a wide range of statistical aspects, comprising prediction of functional data with spatial dependence, optimal sampling designs using functional covariates, non-parametric clustering methods for dependent functional data, and depth measures for spatially dependent functional data.

Keywords Depth functions · Functional data · Functional regression · Spatial dependence · Spatial-depth functional covariates · Spatial functional data

✉ Jorge Mateu
   mateu@uji.es

   Elvira Romano
   elvira.romano@unina2.it

1 Department of Mathematics, University Jaume I,
   12071 Castellón, Spain

2 Department of Mathematics and Physics, Second University
   of Napoli, Caserta, Italy

1 Introduction

In the last decade, the increasing progress of modern technologies allows to handle large complex and high-dimensional spatio-temporal data. Such data is often based on dense sampling schemes of observations over space, time, and over other continuum measures. Accompanying this growth of (high quality, informative and big) data, there has been a rising scientific interest in new statistical methods able to handle and analyse such open problems.

Although classical multivariate statistical techniques can still be applied to this kind of data, they do not capture the additional information coming from the underlying generating process that underpins the data. Spatial functional statistics is a recent research area combining together the well-developed branches of functional statistics and spatial statistics, showing ability to analyse such complex, multivariate data. It has been developed in the framework of the Functional Data Analysis (FDA) paradigm [Ramsay and Silverman (2005); Ferraty and Vieu (2006)] by taking into account the spatial structure of the data. FDA includes methods and theory for data coming in form of functions, and spatial functional statistics extend this approach to deal with samples of functions recorded at different locations of a region (the so-called spatially correlated functional data), or functions observed over a discrete set of time points (temporally correlated functional data).

In contrast to other methods used to analyse spatio-temporal data, spatio-functional techniques enlarge spatial techniques making no parametric assumptions about time effects. As happens with FDA, spatial functional statistics can also be seen as a different way of thinking, where the basic unit of information is the entire observed function rather than a string of numbers with the additional condition related to the spatial component. In particular, the

emerging characteristics about the developed methods are mainly related to the several spatial data structures (geostatistical data, point patterns and areal data) that can be combined with functional data. Thus, as it happens in univariate or multivariate spatial data analysis, methods developed in spatial functional statistics can be largely classified into two main categories: geostatistical functional methods, and purely spatial functional methods.

Geostatistical functional methods aim to describe the spatial correlation of functional phenomena and provide adaptations of regression techniques to account for this spatial dependence. One of the main scopes is the spatial prediction of curves when a sample of spatially correlated curves is collected at a set of locations of a region. In the other category of purely spatial functional methods, we find techniques dealing with point pattern data associated with functional marks, and with functional areal data (groups of functional data with a spatial reference). The objectives (among others) are to study the spatial distribution of curves, testing hypothesis about the observed pattern, detection and modelling of such spatial dependence, and identification of spatial clustering among the curves.

Good and focused motivated examples of spatial indexing problems come from many areas of environmental fields including climate studies. Environmental indicators through widely accessible electronic devices are noise-free, and can be seen to have spatial functional characteristics. In these cases the information consists of a set of curves associated to different geographical locations on a spatial domain.

A substantial and interesting illustration about the underlying philosophy of spatial functional statistics can be found in Delicado et al. (2010a) and Ruiz-Medina (2012). These papers provide an overview of spatial functional statistical methods, starting with the simple notion of a spatial functional observation and its diversification into the different types of spatial functional data structures according to the domain where the observation belongs to.

However, searching in the literature, it is clear that amongst the several categories of spatial functional methods, functional geostatistics has been much more developed considering both new methodological approaches and analysis of a wide range of case studies covering a wealth of varied fields of applications. Indeed, a partial aim in writing this special issue is to encourage the development of further methods in all the branches according to the spatial functional perspective. The focus is not only on case studies, which is per se an interesting and very pragmatic angle of new developments, but also on methodological contributions. In fact, the area of spatially dependent functional structures has several specific and difficult aspects that make it a research topic in its own right. The articles selected here move beyond the current state-of-the-art in various ways, from spatial prediction methods to clustering, going along the way through depth measures for spatially dependent functional data.

In the following section we provide a sketch of the current state-of-the-art of spatial functional statistics by describing the general areas of research, including the new selected proposals which are considered into their bibliographical context through some short discussion on the current literature.

2 An overview of the selected approaches and applications

This special issue comprises six papers considering computational and analytical techniques to deal with functional data with spatial dependence. The issue collects works from experts in this area coming from Sweden, Spain, Colombia, United States, and Italy.

In particular, there are four papers dealing with spatial prediction. Briefly, Aguilera-Morillo et al. (2016) deal with prediction for spatial functional variables whose observations are a set of spatially correlated sample curves obtained as realisations of a spatio-temporal stochastic process. Bernardi et al. (2016) propose a method of regression with partial differential regularisations for spatially dependent curves. Then, Bohorquez et al. (2016) show how to perform optimal spatial prediction of a functional variable at unsampled locations. Finally, Espejo et al. (2016) propose a dynamic spatial-depth functional regression model, where the functional response and the covariates are indexed in time, and take their values in the space of square integrable functions over the spatial-depth domain.

Two more papers complete the special issue. Abramowicz et al. (2016) propose a functional non-parametric clustering method which simultaneously clusters and aligns spatially dependent curves. Balzanella et al. (2016) address the problem of getting order statistics for georeferenced functional data by means of depth functions for spatially dependent functional data.

Note that prediction methods in spatial functional statistics have been developed from different points of view as adapted extensions of more classical prediction methods. Kriging techniques and spatial regression methods have been adapted to the case of spatially correlated curves. Kriging Chiles and Delfiner (1999) is a well-known prediction method in classical geostatistics. It allows to predict values of a (scalar) random field based on sampled surrounding data points, weighted according to a spatial covariance function. The assumption under which it is developed (if the process is considered stationary or non-stationary) drives different types of kriging techniques,

such as ordinary kriging (OK), universal kriging (UK), indicator kriging, cokriging and others. There has been a number of adaptations of kriging approaches for spatially correlated functional data. After the pioneering work of Goulard and Voltz (1993), the papers of Delicado et al. (2010b) and Nerini et al. (2010) have been crucial. Both propose ordinary kriging approaches allowing to predict a curve at an unsampled site under the assumption of stationarity. They are based on the functional linear pointwise model adapted to the case of spatially correlated curves. The idea behind this procedure consists of a direct adaptation of the more classical prediction problem in geostatistics to curves, after a particular smoothing process has been applied. The prediction problem is then solved by estimating a linear model of coregionalisation to capture the spatial dependence among the fitted coefficients. The main methodological difference among these two approaches is that Nerini et al. (2010) use the condition of orthonormal basis functions, whereas orthogonality is not a required condition in Delicado et al. (2010b). A somehow different definition of an ordinary kriging predictor for functional data can be found in Delicado et al. (2011), and in its implementation in R in Giraldo et al. (2012b).

We note that quite often geostatistical methods assume that the spatial functional process considered is stationary, that is, the mean function is constant (no trend), the variance function is constant, and the covariance function depends on the distance between the locations. However, in many applied cases, the assumption of a constant mean function is clearly not realistic. To address this problem, there have been a number of contributions dealing with this situation (see Caballero et al. 2013; Menafoglio et al. 2013; Ignaccolo et al. 2014; Reyes et al. 2015). In all these cases, the stationarity assumption is relaxed. Caballero et al. (2013) propose a new predictor by extending the classical universal kriging predictor for univariate data to the context of functional data. Menafoglio et al. (2013) establish a kriging theory for random fields in any separable Hilbert space, allowing for the analysis of a broad range of object data, such as curves, surfaces or images. Reyes et al. (2015) generalise the classical residual kriging method used in univariate geostatistics proposing a three step procedure. Finally, by considering more complex forms of non-stationarity when the mean function depends on exogenous variables (either scalar or functional), the work of Ignaccolo et al. (2014) develops the so-called kriging with external drift or regression kriging in a functional data setting.

With a focus on estimation procedures, spatio-temporal modeling and non-parametric spatial regression models for functional data have also been proposed. In the former cases, covariance functions for data evolving in space and time are directly modelled by some kind of spatial-functional model. Some of these alternative methods working with spatio-temporal models for spatial-functional data have been proposed by Yamanishi and Tanaka (2003) and Bel et al. (2010). Bel et al. (2010) use a functional linear model to model the relationship between the genetic diversity in European beech forests and curves of temperature and precipitation reconstructed from the past. In addition, in order to take into account the spatial dependence they estimate the covariance matrix of the residuals in a spatial framework. In contrast, a regression model where both response and predictors are functional data, and the relation among the variables may change over the space is proposed by Yamanishi and Tanaka (2003). In a similar way, a non-parametric kernel regression with scalar response and functional predictors, as observations of a continuous spatial process, is proposed by Dabo-Niang and Yao (2007). With the main aim to answer an important space physics question regarding global changes in the ionosphere, Gromenko (2013) proposes a new functional regression approach that handles unevenly spaced, partially observed curves.

In connection with this substantial amount of research, four new methods proposed in this special issue deal with the problem of prediction for spatially correlated functional data.

Aguilera-Morillo et al. (2016) provide a spatial regression method based on a penalised estimation criterion with the aim to predict a random variable continuously in time and space. In particular a three-dimensional P-spline penalty at the least squares fitting criterion is proposed in order to take handle the spatial component of the data. Their method is compared to the classical ordinary kriging, and its performance is illustrated with simulated data. An application on the Canadian temperature data set, introduced by Ramsay and Silverman (2005), is also analysed. This data set has been many times object of study using functional techniques for spatial data (see good examples in Giraldo et al. (2012a, b), and Delicado et al. (2010a, b).

Bernardi et al. (2016) propose a regression method with partial regularisation with the aim of accounting specifically for the geometry of the domain of interest. The focus is in this case to deal with complex time and space structures by considering two roughness penalties that account separately for the regularity of the field in space and in time. The proposed method has advantages with respect to classical spatial data analysis techniques, since as the authors show, it is able to efficiently deal with data distributed irregularly over shaped domains, with complex boundaries, strong concavities and interior holes. What is evident is that combining the spatial and time texturing allows to take advantage with respect to kriging methods for complex spatial functional data structures. The proposed method is compared via simulation studies to other

spatio-temporal techniques, and it is applied to the analysis of the annual production of waste in the towns of Venice province.

These two previous papers propose regression methods that focus on different ways to account for the spatial dependence in the functional data, and to predict a functional observation at a non-observed spatial location. The most relevant difference consists in the way they include the spatial component into the model, and as a consequence the entire estimation process. In the former case, both components are included in a spatial regression model and a combination of an estimation method of spatial smoothing with a P-spline penalty provides a method to predict spatially correlated functional data. In the latter case, based on the idea of regression with differential regularisations, the model merges functional and numerical techniques to deal with the spatial functional covariates.

Espejo et al. (2016) propose a dynamic spatial-depth functional regression model, where the functional response and the covariates are indexed in time, and take their values in the space of square integrable functions over the considered spatial-depth domain. The authors offer a new perspective to exploit the spatial depth ocean temperature field. It is one important challenge in the environmental field that has been analysed by Ruiz-Medina and Espejo (2012, 2013) focusing on the problem of spatial functional extrapolation of ocean surface temperature profiles and surface temperature anomalies.

The problem of prediction of a functional variable has to be considered also in the case where there is cross-variability with other functional variables. To analyse the spatial cross-dependencies with other functional variables is complicated due to the infinite dimensionality of the data. The key difficulty is in specifying and estimating the function responsible for the relationship between distinct variables, that is the cross-covariance function.

To overcome this drawback, Bohorquez et al. (2016) propose a functional cokriging method based on the representation of each function in terms of its empirical functional principal components. The basic idea the authors develop is a generalisation of their previous paper in the univariate case. They thus show how the functional cokriging only depends on the auto-covariance and cross-covariance of the associated scores vectors, which are scalar random fields. In addition, they propose an approach to find optimal sampling designs that ensure the quality of the spatial functional predictions in presence of covariates.

The idea used in the procedure proposed by Bohorquez et al. (2016) has the advantage that it uses the functional principal component representation of each random field. Thus, the functional cokriging method does not require multivariate functional principal component analysis. Among the contents of the paper, there is a part devoted to clarify the characteristics of a coregionalisation model for functional data, where they carefully show the solution they give to the problem they have to ensure that the covariance matrix is positive definite. Several important ideas are emphasised in the paper, among these the possibility to extend functional geostatistics under Gaussianity or goodness-of-fit tests for a joint Gaussian distribution in Hilbert spaces.

All these proposed prediction methods have been proven to provide statistical tools for many applied problems related to environmental, weather, and climate studies.

Another interesting topic of research in these fields is the use of clustering methods to describe local spatial functional characteristics of the observed phenomena. The state-of-the-art about clustering has not been so straightforward developed as much as the prediction problem. To the best of our knowledge there are just few approaches which consider the inclusion of the spatial correlation within clustering methods. Examples of spatial functional clustering are provided in Romano et al. (2010, 2016) and Secchi et al. (2012). These use iterative algorithms to partition geographically referenced data. In particular, a first approach Romano et al. (2010) proposes to classify curves by minimising the spatial variability in each cluster and proposes, as prototype of a cluster, a kriging prediction at an unsampled location. Thus, the optimised objective function performs not only the prediction of the representative curve of each cluster, but it also estimates its location. The approach of Romano et al. (2016) consists of solving an important issue of environmental fields: the description of clusters in terms of spatial dispersion. The method is an extension of the dynamic clustering algorithm to a set of spatially dispersed functions. A further alternative to these methods has been proposed in Secchi et al. (2012), with a bagging strategy in which the nal partitioning of the data is obtained by bagging together weak analysis performed on reduced datasets.

Other contributions provided two types of hierarchical clustering methods. A first one is a generalisation of a hierarchical approach for spatial data to functional data via weighting the dissimilarity matrix by a measure of a spatial functional covariance [see Giraldo et al. (2012a)]. Performances of this method have been shown in Romano et al. (2015). A second one allows to solve the practical problem of identifying groups of stations along a river network which are spatially coherent, calculates the spatial covariance function between functions from sites along a river network, and applies the measure as a weight within the functional hierarchical clustering step [Haggarty et al. (2015)].

Finally, a non-parametric model-based method with a spatially correlated error structure to classify service accessibility patterns for the financial services industry has been proposed by Jiang and Serban (2012).

For completeness of this special issue, two more key topics are developed. Abramowicz et al. (2016) propose a

novel method, the Bagging Voronoi K-medoid Alignment algorithm (BVKMA), that jointly handles clustering, misalignment, and spatial dependence of functional data. As claimed by the authors, this method is the first proposal in the literature that jointly deals with these three sources of variability. Moreover, it allows many different families of warping functions to address the problem of misalignment.

And, additionally, introducing the concept of depth function for establishing the centrality? of an observation among a set of functions, and to provide a natural center-outward ordering of the sampled curves is also a hot topic. Depth definitions are mainly obtained as the generalisation of the classical depth concept, and these definitions can come from integrals of univariate depth, from the graphical representation of the functions, or from depth-based projections for functions. A good reference is Lopez-Pintado and Romo (2011), where it is shown that most of the proposed depth functions have degenerate behaviour in infinite dimensional spaces. Only one of these, named the modified version of the Half Region Depth (HRD), does not suffer from such behaviour, and it is simple and computationally fast.

In spatial functional statistics this topic has been addressed by Balzanella and Romano (2015), where the spatial dependence among the curves is introduced in the definition of the band depth. The spatial covariance function plays the role of a weighting scheme among the geostatistical functional data. Indeed, the considered spatial covariance function measures the spatial dependence of all the curves in the space, but does not consider each single contribution that a curve provides to the whole spatial variability. Moreover, it suffers from a degenerate behaviour for some standard probability models in functional spaces.

The contribution of Balzanella et al. (2016) overcomes these problems. It is a generalisation of the graphical approach based on the modified version of the HRD proposed by Lopez-Pintado and Romo (2011) consisting of defining spatial dispersion functions computed for each site of the observed functional data. By introducing the concept of spatial dispersion function as a transformation of the functional data, this proposal has the following advantages: (a) it furnishes a criterion for ranking simultaneously the spatial and the functional component of the data; (b) it allows to define a distribution of the spatial dispersion functions characterised by robust location estimates, such as the median spatial dispersion and the quartile functions.

3 Final conclusions

In this introductory paper to the present special issue we have discussed some new techniques for spatially dependent functional data, opening new areas for future research.

We hope that these contributions will further enhance the current interest in statistical methods in the spatial functional framework.

We would like to express our gratitude to the reviewers that collaborated in the edition of this special issue. We are specially grateful to the Editor-in-Chief and the Editorial Board of Stochastic Environmental Research and Risk Assessment for their support.

References

Abramowicz K, Arnqvist P, Secchi P, Sjöstedt de Luna S, Vantini S, Vitelli V (2016) Clustering misaligned dependent curves applied to varved lake sediment for climate reconstruction. Stoch Environm Res Risk Assess. **(In press)**

Aguilera-Morillo MC, Durbán M, Aguilera AM (2016) Prediction of functional data with spatial dependence: a penalized approach. Stoch Environ Res Risk Assess. **(In press)**

Balzanella A, Romano E (2015) A depth function for geostatistical functional data. Advances in statistical models for data analysis, studies in classification, data analysis, and knowledge organization. Springer, Berlin, pp 9–16

Balzanella, A., Romano, E., Verde, R. (2016). Modified half-region depth for spatially dependent functional data. Stoch Environ Res Risk Assess. **(In press)**

Bel L, Bar-Hen A, Cheddadi R, Petit R (2010) Spatio-temporal functional regression on paleoecological data. J Appl Stat 38:695–704

Bernardi MS, Sangalli LM, Mazza G, Ramsay JO (2016) A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. Stoch Environ Res Risk Assess. **(In press)**

Bohorquez M, Giraldo R, Mateu J (2016) Multivariate functional random fields: prediction and optimal sampling. Stoch Environ Res Risk Assess. **(In press)**

Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. Stoch Environ Res Risk Assess 27:1553–1563

Chiles JP, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Dabo-Niang S, Yao AF (2007) Kernel regression estimation for continuous spatial processes. Math Methods Stat 16:298–317

Delicado P, Giraldo R, Comas C, Mateu J (2010a) Statistics for spatial functional data: some recent contributions. Environmetrics 21:224–239

Delicado P, Giraldo R, Mateu J (2010b) Continuous time-varying kriging for spatial prediction of functional data: an environmental application. J Agr Biol Environ Stat 15:66–82

Delicado P, Giraldo R, Mateu J (2011) Ordinary kriging for function-valued spatial data. Environ Ecol Stat 18:411–426

Espejo RM, Fernández-Pascual RM, Ruiz-Medina MD (2016) Spatial-depth functional estimation of ocean temperature from non-separable covariance models. Stoch Environ Res Risk Assess. **(In press)**

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

Giraldo R, Delicado P, Mateu J (2012a) Hierarchical clustering of spatially correlated functional data. Stat Neerl 66:403–421

Giraldo R, Mateu J, Delicado P (2012b) Geofd: An R package for function-valued geostatistical prediction. Revista Colombiana de Estadistica 35:383–405

Goulard M, Voltz M. (1993) Geostatistical interpolation of curves: a case study in soil science. Soares A (ed) GeostatisticsTroia92. Kluwer Academic Press, Boston

Gromenko O (2013) Spatially indexed functional data. Ph.D. Thesis

Haggarty R, Miller C, Scott EM (2015) Spatially weighted functional clustering of river network data. J R Stat Soc Ser C Appl Stat 64:491–506

Ignaccolo R, Giraldo P, Mateu J (2014) Kriging with external drift for functional data for air quality monitoring. Stoch Environ Res Risk Assess 28:1171–1186

Jiang H, Serban N (2012) Clustering random curves under spatial interdependence: classification of service accessibility. Technometrics 54:108–119

Lopez-Pintado S, Romo J (2011) A half-region depth for functional data. Comput Stat Data Anal 55:1679–1695

Menafoglio A, Secchi P, Dalla Rosa M (2013) A universal kriging predictor for spatially dependent functional data of a Hilbert space. Electron J Stat 7:2209–2240

Nerini D, Monestiez P, Mantè C (2010) Cokriging for spatial functional data. J Multivar Anal 101:409–418

Ramsay JE, Silverman BW (2005) Functional data analysis, 2nd edn. Springer, Berlin

Reyes A, Giraldo R, Mateu J (2015) Residual kriging for functional spatial prediction of salinity curves. Commun Stat Theory Methods 44:798–809

Romano E, Balzanella A, Verde R (2010) Clustering spatio-functional data: a model-based approach. Studies in classification, data analysis, and knowledge organization. Springer, New York

Romano E, Balzanella A, Verde R (2016) Spatial variability clustering for spatially dependnet functional data. Stat Comput. **(In press)**

Romano E, Mateu J, Giraldo R (2015) On the performance of two clustering methods for spatial functional data. Adv Stat Anal 99:467–492

Ruiz-Medina MD (2012) New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. Spatial Stat 1:82–91

Ruiz-Medina MD, Espejo RM (2012) Spatial autoregressive functional plug-in prediction of ocean surface temperature. Stoch Environ Res Risk Assess 26:335–344

Ruiz-Medina MD, Espejo RM (2013) Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming. Spatial Stat Mapping Environ 22:27–39

Secchi P, Vantini S, Vitelli V (2012) Bagging Voronoi classifiers for clustering spatial functional data. Int J Appl Earth Obs Geoinform 22:53–64

Yamanishi Y, Tanaka Y (2003) Geographically weighted functional multiple regression analysis: A numerical investigation. J Jpn Soc Comput Stat 15:307–317