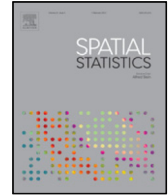Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

SPATIAL
STATISTICS

Check for
updates

Spatio-temporal statistical methods in environmental and biometrical problems

Jorge Mateu [a,*], Rosa M. Crujeiras [b], Raquel Menezes [c], Francisco Montes [d]

[a] *Department of Mathematics, University Jaume I of Castellón, Spain*
[b] *Department of Statistics, Mathematical Analysis and Optimization, University of Santiago de Compostela, Spain*
[c] *Department of Mathematics and Applications, University of Minho, Portugal*
[d] *Department of Statistics and O.R., University of Valencia, Spain*

A R T I C L E   I N F O

A B S T R A C T

This is the editorial letter for the Special Issue dedicated to the VIII International Workshop on Spatio-temporal Modelling (MET-MAVIII) which took place in Valencia (Spain) from 1 to 3 June 2016, and to the second Galician-Portuguese meeting of Biometry, with applications to Health Sciences, Ecology and Environmental Sciences (BIOAPP2016) held in Santiago de Compostela (Spain), 30–2 July 2016.

This special issue summarises and discusses selected peer-reviewed contributions related to spatial and spatio-temporal statistical methodologies comprising both new methodological approaches and a wide range of applications related to environmental and biometrical problems. Point processes, lattice data and geostatistical methods are covered. These methods are illustrated with statistical analyses of animal or plant species in ecological studies, seismic data, temperatures and monthly precipitation, daily ozone concentration values, air pollution data, breast cancer incidence rates, mussels, wildfires, pore structures in pharmaceutical coatings, hake recruitment and cancer mortality data.

© 2017 Published by Elsevier B.V.

* Corresponding author.
E-mail addresses: mateu@uji.es (J. Mateu), rosa.crujeiras@usc.es (R.M. Crujeiras), rmenezes@math.uminho.pt (R. Menezes), montes@uv.es (F. Montes).

1. Introduction

In recent years, spatio-temporal modelling has become one of the most interesting and, at the same time, challenging research areas of natural sciences. This has been largely fueled by the increased availability of inexpensive, high-speed computing. Such availability has enabled the collection and management of large spatial and spatio-temporal datasets across many fields, has facilitated the widespread usage of sophisticated geographic information systems (GIS) software to create attractive displays, and has endowed the ability to investigate challenging, evermore appropriate and realistic models (Gelfand et al., 2010). The relevant literature is growing fast and along directions that range from theoretical works to methodological developments to real world applications. Spatio-temporal systems modelling involves the synthesis of a rich interdisciplinary body of knowledge for which it is necessary to establish a solid theoretical foundation and a science-based methodology with both researchers and practitioners in mind.

A spatial point process is a stochastic process each of whose realisations consists of a finite or countably infinite set of points in the plane. Spatio-temporal point processes are considered as being a hybrid of the spatial and temporal components, by extending the definition of spatial point processes to include time. Because the spatial location can always be considered as one component of a multidimensional mark, the evolution of spatial features with time is often of special interest. Despite such considerations, studies of spatio-temporal models have lagged well behind those of simple temporal models, and even those of purely spatial models. No doubt the reasons have been largely practical, notably the difficulty of compiling good spatio-temporal datasets and the heavy computations needed to analyse them. One way to observe these processes is to consider the spatial location itself viewed as a mark for a simple point process in time, thereby providing one route to likelihood analyses of spatio-temporal models. Further characteristics, such as magnitude, spatial extent, or even duration, can be added as additional marks. Thus, the study of spatio-temporal point processes leads almost inevitably to the more general study of evolving spatial fields, although practical modelling in this direction is still limited and very subject-specific. In this context, principled statistical modelling and residual analysis are at the core of many outstanding research works. Indeed, for gridded forecasts, deviance residuals seem ideally suited for model comparison. And for replicated spatial or spatio-temporal point patterns, looking for methods of assessing whether all spatial point patterns share a common spatial distribution or there are specific features is of primary interest.

This special issue is dedicated to the VIII International Workshop on Spatio-temporal Modelling (**METMAVIII**) which took place in Valencia (Spain) from 1 to 3 June 2016. The purpose of this workshop is to promote the development and application of spatial, temporal, and mainly spatio-temporal statistical methods to different fields related to the environment. This meeting is an opportunity to bring together several communities with common research interests, such as the development and use of statistical methods in the environmental sciences. In addition, the special issue is also related to the second Galician-Portuguese meeting of Biometry, with applications to Health Sciences, Ecology and Environmental Sciences (**BIOAPP2016**) that was held in Santiago de Compostela (Spain), 30–2 July 2016. The aim of this meeting is to disseminate the latest advances in the development and application of statistical methods in life sciences.

Papers coming from these two conferences span the field of spatio-temporal statistical contributions with a focus on environmental and biometrical problems. Papers presented at METMA workshop are mainly focused on spatial and spatio-temporal statistical techniques solving problems in environment, health and biometry. And a large proportion of papers presented at BIOAPP also focuses on spatial or spatio-temporal problems. The conference topics included statistical analyses of animal or plant species in ecological studies, seismic data, temperatures and monthly precipitation, daily ozone concentration values, air pollution data, breast cancer incidence rates, mussels, wildfires, pore structures in pharmaceutical coatings, hake recruitment and cancer mortality data.

In light of the above considerations, the articles of this special issue have been carefully selected to present a variety of conceptual frameworks, powerful methods and comprehensive techniques that address a number of interesting problems in environmental and biometrical sciences. In particular, this special issue comprises papers mainly dealing with point processes (Gabriel et al., 2017; Fuentes-Santos et al., 2017; Baddeley, 2017; Eckardt and Mateu, 2017; Häbel et al., 2017), geostatistical

techniques (Hristopulos and Tsantili, 2017; García-Soidán and Cotos-Yáñez, 2017; Fernández-Casal et al., 2017; Monteiro et al., 2017; Alegría et al., 2017), geoaddtive modelling (Duarte et al., 2017), Bayesian inference approaches (Thaden et al., 2017; Paradinas et al., 2017; Ugarte et al., 2017), and finally one paper dealing with complexity measures in the multifractal domain (Esquivel et al., 2017).

We now comment on the different lines within each of the above mentioned fields of research developed in the papers comprising this special issue.

In the field of spatial and spatio-temporal point processes, the first-order intensity function and its corresponding density function play a fundamental role in the statistical analysis. In particular, mapping the intensity of objects means a difficult task as soon as these objects are not accessible by automated methods. The knowledge at large scale of the underlying process variability can then only be obtained through sampling and spatial prediction. Thus prediction of the intensity of a point process at locations where it has not been observed, conditional to the observation using the best linear unbiased combination of the point process realisation in the observation window is a key aspect as performed in Gabriel et al. (2017). Here, the weight function associated to the predictor is the solution of a Fredholm equation of second kind, and the kernel and the source term of the Fredholm equation are related to the second-order characteristics of the point process through the pair correlation function. In this context, comparing the spatial distribution of two spatial point patterns is an important issue in many scientific areas such as ecology, epidemiology or environmental risk assessment. However, up to date, the analysis of multitype point processes has been mainly focused on searching for interactions between events of different patterns, i.e. on the second-order structure, while the first-order structure has received less attention. Fuentes-Santos et al. (2017) propose testing the similarity between two spatial point patterns through the comparison of their densities of event locations. They consider a squared discrepancy measure to propose a nonparametric statistical test. The asymptotic normal distribution of the associated statistic provides a calibration procedure.

The intensity function constitutes also the base for the analysis of the spatial inhomogeneity, and forms a building block of performing inference through likelihood analysis. In this context, Baddeley (2017) develops a general approach to spatial inhomogeneity in the analysis of spatial point pattern data. The ideas of local likelihood are applied to the composite likelihoods that are commonly used for spatial point processes. For Poisson point processes, local likelihood is already known; for Gibbs point processes Baddeley (2017) develops a local version of Besag's pseudolikelihood, and for Cox point processes and Neyman–Scott cluster processes a local version of the Palm likelihood of Ogata and Katsura is developed. Using recent results for composite likelihood and for spatial point processes, Baddeley (2017) develops tools for statistical inference, including intensity approximations, variance estimators, localised tests for the significance of a covariate effect, and global tests of homogeneity. Bandwidth selection methods are also considered.

When we focus on events that occur randomly in space or space–time on networks, a different story comes into play. Eckardt and Mateu (2017) develop spatial dependence graph models (SDGMs), and network intensity functions. While SDGMs are undirected graphical models which capture the conditional independence structure of multivariate spatial point processes, network intensity functions describe the first-order properties of point patterns that occur on arbitrary network structures.

A final point considered here is when combining both inhomogeneity and anisotropy. Spatial characterisation and modelling of the structure of a material may provide valuable knowledge on its properties and function. In particular, for a drug formulation coated with a polymer film, understanding the relationship between pore structure and drug release properties is important to optimise the coating film design. Häbel et al. (2017) use methods from image analysis and spatial statistics to characterise and model the pore structure in pharmaceutical coatings. More precisely, they use and develop point process theory to characterise the branching structure of a polymer blended film with data from confocal laser scanning microscopy. Point patterns, extracted by identifying branching points of pore channels, are both inhomogeneous and anisotropic. Therefore, Häbel et al. (2017) introduce a directional version of the inhomogeneous K-function to study the anisotropy and then suggest two alternative ways to model the anisotropic three-dimensional structure. The methods presented will be useful for anisotropic inhomogeneous point patterns in general and for characterizing porous material in particular.

On a different vein of spatial statistics, covariance functions constitute the core of a more formal geostatistical modelling approach. Motivated by the generation of non-separable, physically motivated covariance functions, Hristopulos and Tsantili (2017) construct a non-separable space–time covariance function based on a diffusive Langevin equation. They employ ideas from statistical mechanics to express the response of an equilibrium random field to a driving noise process by means of a linear, diffusive relaxation mechanism. The equilibrium field is assumed to follow an exponential joint probability density which is determined by a spatial local interaction model. They use linear response theory to express the temporal evolution of the random field around the equilibrium state in terms of a Langevin equation. The latter yields an equation of motion for the space–time covariance function, which can be solved explicitly at certain limits. By means of the turning bands transform, the authors derive a non-separable space–time covariance function in three space dimensions and time.

Theoretical models are assessed by variograms, and these are often estimated by non-parametric Nadaraya–Watson type estimators. It is then needed to analyse properties of the L_2-deviations of these estimators, as proposed by García-Soidán and Cotos-Yáñez (2017) for both the anisotropic and the isotropic settings. Their convergence in distribution is established, which provides the basis for addressing practical problems, such as the construction of goodness-of-fit tests for the variogram and, therefore, for modelling the spatial dependence. For estimation purposes, the authors propose proceeding through the least squares criteria, whose consistency is proved, together with a reformulation of the global measures for the kernel-type estimators. Then, the resulting critical points can be approximated by appealing to the bootstrap approaches. In addition, Fernández-Casal et al. (2017) provides nonparametric estimators of the conditional variance and the dependence structure of a heteroscedastic spatial process. When assuming zero mean along the domain, the approximation of the variance can be addressed by linear smoothing of the squared observations. Then, the variogram can be estimated from the standardised data. In the presence of a non-zero deterministic trend, the author suggests a modification of the latter method that involves the residuals obtained from a local linear estimation of the trend, together with corrections of the biases derived from the use of these residuals.

A more practical approach is motivated when analysing the spatial and temporal dynamics of spatiotemporal data sets, characterised by high resolution in the temporal dimension, which is a common fact in many application areas, namely environmental modelling. In particular, air pollution data, such as NO_2 concentration levels, often incorporate also multiple recurring patterns in time imposed by social habits, anthropogenic activities and meteorological conditions. Monteiro et al. (2017) propose a two-stage modelling approach which combined with a block bootstrap procedure correctly assesses uncertainty in parameters estimates and produces reliable confidence regions for the space–time phenomenon under study.

Finally, geostatistical modelling of data on spheres motivated by problems of data on the planet Earth is a welcome modern area of research. Alegría et al. (2017) consider a multivariate spatial random field with each component having univariate marginal distributions of the skew-Gaussian type. They assume that the field is defined spatially on the unit sphere embedded in $\mathbb{R}^3$, allowing for modelling data available over large portions of planet Earth. Their model admits explicit expressions for the marginal and cross-covariances. Since inference based on the full likelihood is computationally unfeasible, Alegría et al. (2017) propose a composite likelihood approach based on pairs of spatial observations, and illustrate the effectiveness of the method through simulation experiments and the analysis of minimum and maximum temperatures.

When analysing data from cancer screening programs, flexible regression specifications are required to account for the highly complex structure in such data. Duarte et al. (2017) analyse data from a breast cancer screening program conducted in central Portugal and consider an extension of structured additive regression models where, in addition to the possibility to include nonlinear and spatial effects, they can include a trivariate interaction between attendance rate, detection rate and mortality rate in the screening program. While spatial effects capture unobserved heterogeneity at the level of municipalities, the trivariate interaction proves to be important to understand the complex interaction effects resulting from the diversity in municipalities' coverage and attendance rates.

A wide area of interest for statisticians is that of modelling and understanding ecological interrelationships between species and their environment, and the corresponding spatio-temporal

dynamism. Thaden et al. (2017) investigate how adult mussels and mussel seeds (of Mytilus gallo-provincialis) interact while simultaneously accounting for environmental factors. The authors develop a recursive bivariate simultaneous equation model which considers the environmental endogeneity of the mussels. In order to address the potentially highly complex ecological patterns, they include environmental factors via integrating multivariate autoregressive discrete spatial effect priors into the framework of simultaneous equation models. This allows for general correlation structures of the spatial effects and thus avoids the restrictive (implicit) assumption that adult mussels and mussel seeds of the same species react independently to their environment. Having a neater idea of the spatio-temporal dynamism and environmental relationships of species is essential for the conservation of natural resources. Many spatio-temporally sampled processes result in continuous positive $(0, \infty)$ abundance datasets that have many zero values observed in areas that lie outside their optimum niche. In such cases the most common option is to use two-part or hurdle models, which fit independent models and consequently independent environmental effects to occurrence and conditional-to-presence abundance. This may be correct in some cases, but not as much in others where the detection probability is related to the abundance. Paradinas et al. (2017) infer the spatio-temporal dynamism of ecological processes and fit more robust environmental effects in two-part models. On the one hand they propose different spatio-temporal structures to infer the fundamental spatio-temporal behaviour of the process under study. On the other hand, they propose the use of shared component modelling to estimate more robust model effects in related semi-continuous datasets by combining information from occurrence and conditional-to-presence abundance. The authors implement all the proposed model structures in a case study on hake recruitment.

We finally make a note on two interesting research lines. One is disease mapping, and here models incorporating splines have been considered for smoothing risks. Although these models are very flexible, they can be computationally demanding in certain cases. One, two, and three-dimensional B-splines (penalized or unpenalized) are considered in Ugarte et al. (2017) to model space–time interactions. Model identifiability issues are discussed and appropriate constraints are clearly established. As computing time could be a limitation in real practice, integrated nested Laplace approximations are used for model fitting and inference. The complete set of proposed models are illustrated using cancer mortality data in small areas. The other one is the definition of new parametric families of complexity measures in the multifractal domain, as proposed in Esquivel et al. (2017). Their families of complexity measures come from the scaling limiting behaviour of generalised Renyi-entropy-based product complexity measures. These families are related to incremental functionals of the curve of generalised Renyi dimensions. In particular, the significance in this context of the first derivative of the generalised dimensions curve, and specifically the indicators given by its minimum value and the corresponding value of the deformation parameter, is justified. The practical usefulness of the multifractal complexity measures proposed, for characterisation and assessment of structural dynamical changes in a spatiotemporal system with multifractal behaviour, is illustrated with the study of real data from a seismic series involving a central period of high activity within a regular regime.

References

Alegría, A., Caro, S., Bevilacqua, M., Porcu, E., Clarke, J., 2017. Estimating covariance functions of multivariate skew-gaussian random fields on the sphere. Spatial Statist. 22, 388–402.

Baddeley, A., 2017. Local composite likelihood for spatial point processes. Spatial Statist. 22, 261–295.

Duarte, E., de Sousa, B., Cadarso-Suárez, C., Kneib, T., Rodrigues, V., 2017. Exploring risk factors in breast cancer screening program data using structured geoadditive models with high order interaction. Spatial Statist. 22, 403–418.

Eckardt, M., Mateu, J., 2017. Analysing highly structured point patterns in space. Spatial Statist. 22, 296–305.

Esquivel, F.J., Alonso, F.J., Angulo, J.M., 2017. Multifractal complexity analysis in space–time based on the generalized dimensions derivatives. Spatial Statist. 22, 469–480.

Fernández-Casal, R., Castillo-Páez, S.A., García-Soidán, P., 2017. Nonparametric estimation of the small-scale variability of heteroscedastic spatial processes. Spatial Statist. 22, 358–370.

Fuentes-Santos, I., González-Manteiga, W., Mateu, J., 2017. A nonparametric test for the comparison of first-order structures of spatial point processes. Spatial Statist. 22, 240–260.

Gabriel, E., Coville, J., Chadœuf, J., 2017. Estimating the intensity function of spatial point processes outside the observation window. Spatial Statist. 22, 225–239.

García-Soidán, P., Cotos-Yáñez, T.R., 2017. Convergence in distribution of the L_2-deviations of the kernel-type variogram estimators with applications. Spatial Statistics 22, 338–357.

Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P., 2010. Handbook of Spatial Statistics. Chapman and Hall/CRC Press, Boca Raton.

Häbel, H., Rajala, T., Marucci, M., Boissier, C., Schladitzd, K., Redenbach, C., Särkkä, A., 2017. A three-dimensional anisotropic point process characterization for pharmaceutical coatings. Spatial Statist. 22, 306–320.

Hristopulos, D.T., Tsantili, I.C., 2017. Space–time covariance functions based on linear response theory and the turning bands method. Spatial Statist. 22, 321–337.

Monteiro, A., Menezes, R., Silva, M.E., 2017. Modelling spatio-temporal data with multiple seasonalities: the NO2 Portuguese case. Spatial Statist. 22, 371–387.

Paradinas, I., Conesa, D., López-Quílez, A., Bellido, J.M., 2017. Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. Spatial Statist. 22, 434–450.

Thaden, H., Pata, M.P., Klein, N., Cadarso-Suárez, C., Kneib, T., 2017. Integrating multivariate conditionally autoregressive spatial priors into recursive bivariate models for analyzing environmental sensitivity of mussels. Spatial Statist. 22, 419–433.

Ugarte, M.D., Adin, A., Goicoa, T., 2017. One-dimensional, two-dimensional, and three-dimensional B-splines to model space–time interactions in Bayesian disease mapping: model fitting and model indentifiability. Spatial Statist. 22, 451–468.