University of Granada
Doctorate Program in Mathematics and Statistics

Doctoral Thesis

Spatial Functional Statistics analysis of panel data

Rosa María Espejo Montes

Thesis supervised by:

Prof. María Dolores Ruiz Medina

Granada 2014

Editor: Editorial de la Universidad de Granada
Autor: Rosa María Espejo Montes
D.L.: GR 2138-2014
ISBN: 978-84-9083-157-1

Universidad Granada
Programa de Doctorado en matemáticas y Estadística

Tesis Doctoral

Análisis Estadístico Funcional Espacial de datos de panel

Rosa María Espejo Montes

Tesis supervisada por:

Prof. María Dolores Ruiz Medina

Granada 2014

Contents

Summary

This thesis opens new research lines in the field of Spatial and/or Temporal Functional Statistics. The original motivation of the thesis was to contribute to the field of panel data analysis adopting a functional perspective. Specifically, the primary objective of this thesis was to derive flexible statistical models and methodologies for the analysis of correlated curve data in space, or alternatively, for correlated surfaces in time, covering, in particular, the infinite-dimensional multivariate framework. Summarizing, the main contributions of this thesis are related to the following topics:

- The implementation of Spatial Autoregressive Hilbertian extrapolators by projection.
- The formulation of new classes of models in the context of spatial functional multiple regression when response and regressors are Hilbert-valued variables, considering the case where the regression operators are spatially heterogeneous.
- The introduction of spatial functional classification techniques based on the estimation of spatial mixed effect models with fixed and random effect curves having a heterogeneous behavior in space.
- The extension of the proposed weak-dependent models in the context of spatial time series theory to the long-range dependence case in terms of Gegenbauer polynomials.
- The formulation of new RKHS based estimation methodologies for autoregressive Hilbertian processes.

- The derivation of sufficient conditions for the asymptotic normal distribution of maximum likelihood estimators for Gaussian autoregressive Hilbertian processes.

- Finally, the consideration of the theory of multifractional pseudodifferential operators for the representation of heterogeneous curve data behaviors in space (i.e., variable order of differentiation in the weak and mean-square sense), solving the associated functional least-squares estimation problem.

Resumen

Esta tesis desarrolla nuevas líneas de investigación en el campo de la Estadística Funcional Espacial o/y Temporal. La motivación original de la tesis es contribuir en el campo del análisis de datos de panel adoptando una perspectiva funcional. Específicamente, el principal objetivo de esta tesis fue la derivación de modelos estadísticos y metodologías más flexibles para el análisis de curvas correladas en el espacio, o alternativamente, para superficies con correlación temporal, cubriendo, en particular, el marco multivariante infinito-dimensional. Resumiendo, las principales aportaciones de esta tesis están relacionadas con los siguientes temas:

- La implementación mediante diferentes metodologías de proyección del extrapolador espacial autorregresivo Hilbertiano en modelos SARH(1).
- La formulación de nuevas clases de modelos en el contexto de la regresión múltiple funcional espacial, cuando la respuesta y los regresores son variables Hilbert-valuadas, considerando el caso donde los operadores de regresión son espacialmente heterogéneos.
- La introducción de nuevas técnicas de clasificación espacial funcional, basadas en la estimación de modelos de efectos mixtos espaciales Hilbert-valuados, cuyas curvas de efectos fijos y aleatorios presentan un comportamiento heterogéneo en el espacio.
- La extensión de los modelos con dependencia débil propuestos al contexto de series espaciales con dependencia de largo rango, en términos de los polinomios Gegenbauer.
- La formulación de una metodología de estimación penalizada, basada en la teoría de espacios del núcleo reproductor, para procesos autorregresivos Hilbertianos de orden uno.

- La derivación de condiciones suficientes para la normalidad asintótica de estimadores máxima verosímiles de los operadores que definen los parámetros de la ecuación de estados satisfecha por procesos autorregresivos Hilbertianos Gaussianos.

- Finalmente, la consideración de la teoría de los operadores pseudodiferenciales multifracionales para la representación de curvas aleatorias heterogéneas espacialmente (es decir, con orden variable de diferenciación en sentido débil y en media cuadrática), resolviendo el problema asociado de estimación funcional por mínimos cuadrados.

Part I

Chapter 1

Introduction

The classical time series theory is applied to the statistical analysis of the evolution of a phenomena over time. However, the statistical analysis of panel data requires the introduction of new classes of models processing sample information from multiple phenomena (dependent variable) crossed with the observation of several independent variables whose effect on the response should be studied through different time periods incorporating temporal correlations. Dimension reduction techniques then appear as a fundamental topic in panel data analysis, since inference techniques must be implemented incorporating the sample information provided by long record of observations in time in a large dimensional multivariate framework. These techniques should avoid the loss of information in the temporal dimension, which enriches the analysis of panel data, and constitutes a key aspect in most of the areas of application (e.g., Econometrics), and, in particular, in the treatment of great variety of problems such as migration flow, carbon dioxide emissions, energy consumption, environmental studies, shares price control of companies in stock market, the determinant factors of international tourism (see Mahía, 2000; Mayda, 2008; Nayaran and Narayan, 2010; Schweinberger, 2012; Wang, *et al.*, 2011, among others).

Dynamical linear modeling constitutes one of the most popular tools in the statistical analysis of panel data analysis. Specifically, temporal correlated fixed, random and mixed effect models have been widely applied (see, Arellano and Honoré, 2001; Aquaro and Cižek, 2014; Wu and Su,

2010, among others). In the framework of linear dynamical fixed effect models, robust estimation, based on the median of two consecutive pairs of correlated values, is addressed in Aquaro and Cižek (2014) (see also Verardi and Wagner, 2010). In the mixed effect contexto, Meintanis (2011) introduces a new method to check the normality of the error components of statistical panel data models, based on the empirical distribution function and the empirical characteristic function. Generalized quasi-likelihood estimation of mixed effect panel data models has been applied in Sun and Sutradhar (2013). Arellano and BonHomme (2012) consider the method of moments for the parametric estimation of linear panel data models with random coefficients.

Dynamical linear regression modeling also constitutes a key subject in the statistical analysis of panel data (see, for example, Cardot and Sarda, 2011, where a functional framework is adopted; Hu, Sun and Wei, 2003; You and Zhou, 2006; Zhao and Tong, 2011, in relation to the classical approach for dynamical regression). A semi-parametric estimator of weighted least-squares is introduced by You *et al.* (2010) for the partially linear regression models with heteroscedastic errors. See also Galvao (2011) and Lin, Li and Sun (2014) in the fixed effect framework, and He *et al.* (2008) or McKitrick, McIntyre and Herman (2010) in the multivariate regression context.

The pioneer work by Kalbfleisch and Lawless (1985) introduces the panel data modeling approach based on continuous Markov models. Qian and Wang (2012) proposed a new semi-parametric estimation method to approximate the non-linear component in panel data models. See also the papers on non-linear modeling by Chen, Gao and Li (2012); Honoré (2002); Huang, Wang and Zhang (2006); Lee (2014); Li, Chen, and Gao (2011); Su and Lu (2013); Zhang *et al.* (2011).

Interaction between individuals with a specific spatial location, or, in general, interaction between multivariate variables measured at different locations in space can be reflected applying classical tools for Spatial Statistics to panel data modeling. This fact allows the consideration of

a new dimension in the panel data analysis in the presence spatial correlations. Some of the works where spatial dependence is incorporated in the statistical analysis of panel data are Burnett, Bergstromb, and Dorfman (2013); Debarsy and Ertur (2010); Driscoll and Kraay (1998); Elhorst (2010); Lee and Yu (2010); Yu, Jong and Lee (2012)). Non-parametric techniques are applied in Driscoll and Kraay (1998) for the estimation of covariance matrices characterizing the presence of spatial correlations. Lee and Yu (2010) derive the asymptotic properties of quasi-likelihood estimators for spatial mixed effects models. In this spatial mixed-effect framework, different inference approaches can be implemented since the maximum-likelihood based approach until the Empirical Bayesian and purely Bayesian frameworks (see, Meintanis, 2011; Olokoyo, 2013; Salamh, 2011; Wu *et al.*, 2011; Wu and Li, 2014; Wu and Zhu, 2012, Ugarte *et al.*, 2009, among others). A spatial autoregressive framework is adopted in Badinger and Egger (2013) to modeling spatial dynamical error components. The spatial prediction problem in the context of spatial dynamical regression models has been addressed in the papers by Baltagi and Li (2004); Baltagi and Pirotte (2010); Baltagi, Bresson and Pirotte (2012); Baltagi, Fingleton and Pirotte (2014), to mention just a few.

In the context of Spatiotemporal Statistics (see Ugarte *et al.* (2009, 2010, 2012) in relation to statistical inference from the spatiotemporal mixed-effect approach the analysis of disease mapping), new results on panel data analysis incorporating the temporal and spatial dimensions in the multivariate response and transversal random variables, as well possible temporal and spatial correlations have been derived, for example, in the papers by Debarsy, Ertur and LeSage (2012); Parent and LeSage (2010, 2011); Salmerón and Ruiz-Medina (2011), and the references therein.

The aim of the present thesis is to contribute to the panel data analysis in a spatiotemporal framework adopting a functional statistical approach in time or space. Specifically, the main contributions of this thesis are related to the statistical analysis of spatial correlated curves, as

well as temporal correlated surfaces. Most of the spatial functional statistical models and techniques proposed, for the analysis of curves displaying spatial interactions, are motivated by real data problems in the area of Finances and Environment. In the context of functional dynamic modeling the contributions of this thesis are mainly given within the Hilbertian autoregressive time series framework where our main reference is the book by Bosq (2000). A brief sketch of the most significant references in the development of this thesis, from the temporal and spatial functional statistical frameworks, is now provided.

The availability of long-record of observations at numerous databases in many fields of application, environment, economy, finance, education, climate change, road safety, sport, health, psychology, medicine, biology, among other areas, motivates an extensive literature in the last few decades in relation to statistical analysis of functional data. In a probabilistic context, infinite-dimensional probability distributions are considered as suitable models to characterize the behavior of functional observations, like curves, surfaces, etc. Several approaches to functional prediction in time from the Hilbert-valued time series approach have been investigated in the papers by Antoniadis, Paparoditis and Sapatinas (2006); Antoniadis, and Sapatinas (2003); Mourid (2002), among others. Indeed, the framework of functional (Hilbert-valued) series has been widely discussed (see Besse, Cardot and Stephenson, 2000; Damon and Guillas, 2002, 2005; Hörmann and Kokoszka, 2011; Horváth, Kokoszka and Reeder, 2013), although there are still many open research lines in this field. As commented, the texts of Bosq (2000) and Bosq and Blanke (2007) constitutes one of the key references in the development of this thesis that have inspired most of the models and estimation methodologies proposed. These books provide the introduction, nonparametric estimation, and prediction based on autoregressive Hilbertian processes. Several results describing the asymptotic behavior of the estimators derived are obtained as well (see also Cardot *et al.* 2003; Damon and Guillas, 2005; Mas, 2007; Ruiz-Medina and Salmerón, 2010; Ruiz-Medina, Salmerón and Angulo, 2007; Salmerón and Ruiz-Medina, 2009).

On the other hand, principal component analysis and canonical correlation analysis for infinite-dimensional random variables have been extensively applied in the context of linear models in function spaces (see, for example, the classical text by Ramsay and Silverman, 2005).

A classical text on functional non-paranetric statistics is the book by Ferraty and Vieu (2006), (see also the references by Ferraty, Kudraszow and Vieu, 2012; Ferraty and Vieu, 2004). Antoniadis, Paparoditis and Sapatinas (2006) consider nonparametric estimation techniques for functional regression. Giraitis, Kapetanios and Yates (2014) introduce a new class of nonparametric autoregressive processes which decompose in a temporal series with varying coefficients in time. Shang (2013) considers the estimation of the error component in a nonparametric functional regression model with functional predictor and scalar response by a Bayesian approach (see also, in the context nonparametric statistics, Li and Hsing, 2010; Laksaci, Rachdi and Rahmani, 2013; Lian, 2012, among others). Aneiros-Perez, Cao and Vilar-Fernandez (2011) consider the problem of the prediction of temporal series using nonparametric functional techniques, performing an extension of the linear regression method with functional explanatory variables. Earls and Hooker (2014) have proposed a Bayesian inference method for functional data characterized by a Gaussian process. Angelini, De Canditiis, and Pensky (2012) analyze the temporal correlation between Gaussian curves in a nonparametric regression framework, using the theory of reproducing kernel Hilbert spaces generated by wavelet bases.

Spatial Functional Statistics is a relatively new branch of Statistics allowing the analysis of spatial correlated functional observations. In particular, we will refer to the papers by Guillas and Lai (2010) where spatial functional regression models based on bivariate splines are introduced, and least squares estimation is applied with/without penalty term from the approximation of the surface of explanatory random variables by projection. Bayesian inference from spatial correlated curves is implemented in Baladandayuthapani *et al.* (2008). Spatial correlation analysis for high-dimensional data from a functional perspective is considered in Fan and

Zhang (2000). Cokriging techniques for spatial functional data are implemented in Nerini, Monestiez and Manté (2010). Giraldo, Delicado and Mateu (2010) implement a spatial functional predictor, based on temporal kriging techniques from projection on suitable functional bases. In the context of nonparametric statistics, Basse, Diop and Dabo-Niag (2008) propose kernel-based density estimators for spatial correlated functional random variables. Motivated by ocean variable studies, Nerini and Ghattas (2010) apply functional statistical classification tools for the analysis of spatial correlated curve data. We also mention the papers by Giraldo, Delicado, and Mateu (2012); Romano, Giraldo and Mateu (2011), and Secchi, Vantini and Vitelli (2011) on functional statistical classification of spatially correlated curves. Finally, we refer to the Bayesian framework adopted in Ma *et al.* (2008) for spatial functional statistical classification based on information criteria.

Spatial time series models with values in a separable Hilbert-space are introduced in Ruiz-Medina (2011), where conditions for their stationarity and invertibility are investigated. The non-parametric estimation of the operators defining the functional parameters of the state space equation associated with the Spatial Autoregressive Hilbertian process of order one (SARH(1) process) is obtained in Ruiz-Medina (2012a) by projection into the dual eigenvector systems of such autocovariance operators. A review on the state of the art of the current literature on Spatial Functional Statistics, as well as on the most relevant open problems that could be addressed is reflected in Ruiz-Medina (2012b).

As commented before, the aim of the present thesis is to contribute with new results of theoretical and computational nature to the development of new tools in the field of Spatial Functional Statistics, as well as to solve some problems that still remain open in the field of Temporal Functional Statistics motivated by their possible application to the field of panel data analysis. Summarizing, in this thesis, new models and methodological approaches are proposed for the statistical analysis of spatial and temporal correlated curves and surfaces.

These contributions can be distributed in three parts according to the field where they are located. Specifically, the first part of the thesis reflects the main results derived in the area of Spatial Functional Statistics for the analysis of correlated curves in space, and in particular, for the analysis of panel data under this perspective. The second part reflects the main results obtained in the field of Hilbert-valued time series models. In particular, we concentrate in the study of the class of Autoregressive Hilbertian processes of order one (ARH(1) processes). Finally, the third part is related to the spatial functional modeling under the multifractional pseudodifferential equation approach, providing the regularization of the associated least-squares functional estimation problem, in terms of suitable functional bases related to the corresponding Reproducing Kernel Hilbert Space (RKHS), whose elements display a spatial heterogenous local regular behavior (variable order of differentiation).

Appendix A1 implements the SARH(1) extrapolator defined in Ruiz-Medina (2012a) by projection into the eigenvectors of the auto-covariance operator of SARH(1) process. The optimal dimension reduction, in the sense of variance, obtained with the application of this projection methodology is illustrated with a real-data example related to spatial extrapolation of ocean surface temperature curves at different areas of the earth globe, where global warming and climate change effects must be analyzed. Alternatively, Appendix A2 provides the implementation of SARH(1) extrapolator in terms of the discrete compactly supported wavelet transform to weaken strong spatial correlations between curve data that hinder the SARH(1) analysis previously developed in terms of the eigenvectors of the auto-covariance operator. The application of this transform is motivated by the high spatial concentration of weather stations in the ocean islands, that induce strong spatial correlations in the observed ocean surface temperature curves at such weather stations. In Appendix A3, a new class of spatially heterogeneous multiple regression models for random variables with values in a separable Hilbert space is introduced. The error term is assumed to be weak correlated in space according to SARH(1) state space equation.

The implementation of the proposed non-parametric estimators of the regression operators and the functional parameters defining the SARH(1) state space equation, satisfied by the functional error term, is also derived in this appendix. The formulated model is motivated by financial applications. Specifically, in previous empirical studies developed for the analysis of firm panel data, it has been tested the non-stability in space of regression coefficients defining the linear relationship between firm factors and indebtedness curves. In addition, the effect of institutional factors and the industrial sector (both aspects usually introduced in the statistical analysis as dummy variables), is reflected in terms of possible heterogeneities in this linear relationship, and it had also constituted the subject of several empirical studies in the financial context. Both aspects motivate the introduction of the new class of models formulated in Appendix A3, in the Hilbert-valued context, to reflect possible spatial heterogeneities in the linear relationship between the firm factors and indebtedness curves, as well as the presence of possible spatial correlations between firm indebtedness curves within an industry sector. The spatial dependence between firm indebtedness curves could also be introduced through a random functional parameter in the model. This fact motivates the subject of Appendix A4, where a new class of spatial functional statistical models is derived in the mixed effect framework. A new classification methodology for spatial correlated curves is formulated, based on the introduced spatial functional mixed effect model, leading to the detection of local spatial homogeneous patterns in the regression operators involved in the definition of the fixed effect curves, as well as the detection of local spatial homogeneous patterns in the mean-quadratic functional local variation in space of random effect curves. Finally, in this part of the thesis a new model family is derived to represent strong correlation in the space. The extension of the class of temporal autoregressive Gegenbauer process to the spatial and Hilbert-valued context is performed in Appendix A5. Their structural properties are analyzed as well.

In the context of Gaussian ARH(1) processes, a penalized functional least-squares estimation

technique, based on the RKHS geometry, is proposed to stabilize standardized mortality ratio (SMR) in disease mapping. The motivation of this appendix relies on the real-data problem analyzed in relation to breast cancer mortality in Spain. Specifically, classical approaches like the ones based on Conditional Autoregressive Models (CAR modeling) and spline-based smoothing, in a mixed effect framework, present the problem of an hyper-smoothing of the breast cancer mortality log-relative-risk curves at each one of the Spanish provinces analyzed (small areas analyzed). This is the reason why in Appendix A6, a new approach is presented for stabilization of SMR, based on the combination of ARH(1) and RKHS approaches, to interpolate local variability in a finer scale than the integer-order scale usually considered, avoiding false negatives in the detection of risk areas. On the other hand, the smoothing parameter in the proposed penalized estimation methodology is fitted by cross-validation, while in previous spline-based smoothing approaches (see, for example, Ugarte *et al.*, 2010) this parameter is approximated as a variance component by applying maximum likelihood (ML) estimation methodology. This is the reason why in Appendix A7 maximum-likelihood estimation methodology for ARH(1) processes is studied. Specifically, in Appendix A7, the results derived in Ruiz-Medina and Salmerón (2010) for Gaussian ARH(1) processes are extended to a more general family of ARH(1) processes whose infinite-dimensional distribution belongs to the exponential family introduced in this appendix. Additionally, the asymptotic normal distribution of ML estimators within such H-valued random variable family is derived.

The thesis ends with the introduction of a new family of Hilbert-valued Gaussian models for curves, based on the theory of multifractional pseudodifferential equations. The associated functional least-squares estimation problem for approximation of the functional values of the solution is ill-posed. Its regularization can not be obtained in a direct way applying classical numerical regularization methods, since the functions in the corresponding RKHS present a variable order of differentiation depending on space. This is the reason why a new projection

methodology is derived in Appendix A8 for regularization of this estimation problem removing its ill-posed nature. Such a projection methodology is based on dual Riesz bases providing a spectral diagonalization of the operator that generates the closed bilinear form defining the inner product in the corresponding RKHS.

Before the eight appendixes that constitute the eight papers published in relation to the contents of this thesis, we provide a brief sketch of the objectives, methodology and main results and conclusions derived in these appendixes, in order to facilitate their reading by summarize of the key aspects and points addressed in theses appendices.

Chapter 2

Introducción

La teoría clásica de series temporales surge como herramienta clásica para el análisis estadístico de la evolución temporal de magnitudes relacionadas con un determinado fenómeno en el tiempo. Sin embargo, el análisis estadístico de datos de panel requiere la introducción de nuevas clases de modelos para procesar la información de múltiples magnitudes o bien, de diferentes fenómenos que se hayan relacionados y cuyo efecto sobre la respuesta debe ser estudiado a través de diferentes periodos de tiempo incorporando la correlación temporal. Las técnicas de reducción de la dimensión juegan entonces un papel fundamental en el análisis de datos de panel donde se analiza la evolución temporal de la variable de interés y variables independientes (cortes transversales) en un contexto multidimensional. Estas técnicas deben evitar la pérdida de información en la dimensión temporal, aspecto crucial que enriquece el análisis de datos de panel, constituyendo un aspecto clave en la mayoría de los ámbitos de aplicación (por ejemplo, econometría), y, en particular, en el tratamiento de una gran variedad de problemas tales como el flujo de la migración, las emisiones de dióxido de carbono, consumo de energía, estudios ambientales, control de acciones del precio de las empresas en el mercado de valores, los factores determinantes del turismo internacional (ver Mahía, 2000; Mayda, 2008; Nayaran y Narayan, 2010; Schweinberger, 2012; Wang, *et al.*, 2011, entre otros).

Los modelos dinámicos lineales constituyen una de las herramientas más populares en el

análisis estadístico de datos de panel. Específicamente, los modelos fijos de correlación temporal, los modelos de efectos mixtos y aleatorios han sido extensamente aplicados (véase, Arellano y Honoré, 2001; Aquaro y Cižek, 2014; Wu y Su, 2010, entre otros). En el contexto de los modelos de efectos fijos con dinámica lineal, Aquaro y Cižek (2014) abordan la estimación robusta, basada en la mediana, de valores correlacionados (ver también Verardi y Wagner, 2010). En el marco de los efectos mixtos, Meintanis (2011) introduce una nueva metodología para comprobar la normalidad de las componentes de error de los modelos estadísticos de datos de panel, basándose en la función de distribución empírica y la función característica empírica. La estimación de la cuasi-verosimilitud generalizada de los modelos de efectos mixtos de datos de panel ha sido aplicada en Sun y Sutradhar (2013). Arellano y BonHomme (2012) consideran el método de momentos para la estimación de modelos lineales de datos de panel con coeficientes aleatorios.

Los modelos de regresión con dinámica lineal también constituyen un aspecto clave en el análisis estadístico de datos de panel (véase, por ejemplo, Cardot y Sarda, 2011, donde se adopta un marco funcional; Hu, Sun y Wei, 2003; You y Zhou, 2006; Zhao y Tong, 2011, en relación con el enfoque clásico para la regresión dinámica). Un estimador semi-paramétrico de mínimos cuadrados ponderados es introducido por You *et al.* (2010) para los modelos de regresión parcialmente lineales con errores heterocedásticos. Ver también Galvao (2011) y Lin, Li y Sun (2014) en el marco de efectos fijos, y He *et al.* (2008) o McKitrick, McIntyre y Herman (2010) en el contexto de la regresión multivariante.

El trabajo inicial de Kalbfleisch y Lawless (1985) introduce la modelización, basada en procesos markovianos, en el contexto de datos de panel. Qian y Wang (2012) proponen un nuevo método de estimación semi-paramétrico para aproximar la componente no lineal en modelos para datos de panel. Ver también los trabajos desarrollados sobre modelización no-lineal de Chen, Gao y Li (2012); Honoré (2002); Huang, Wang y Zhang (2006); Lee (2014); Li, Chen, y Gao (2011); Su y Lu (2013); Zhang *et al.* (2011).

Las herramientas clásicas de la Estadística Espacial permiten incorporar una nueva dimensión en el análisis de datos de panel teniendo en cuenta la interacción entre individuos con una ubicación espacial específica, o en general, la interacción espacial entre variables multivariantes medidas en diferentes localizaciones. Algunos de los trabajos donde se incorpora la correlación espacial en el análisis estadístico de los datos de panel son Burnett, Bergstromb, y Dorfman (2013); Debarsy y Ertur (2010); Driscoll y Kraay (1998); Elhorst (2010); Lee y Yu (2010); Yu, Jong y Lee (2012)). Driscoll y Kraay (1998) aplican técnicas no paramétricas en la estimación de las matrices de covarianza que reflejan la presencia de correlación espacial. Lee y Yu (2010) derivan las propiedades asintóticas de estimadores cuasi-verosímiles para modelos espaciales de efectos mixtos. En este contexto de modelos espaciales de efectos mixtos, se han adoptado diferentes enfoques para la inferencia que cubren desde el enfoque basado en la estimación por máxima verosimilitud hasta el enfoque desarrollado en un marco empírico Bayesiano y puramente bayesiano (véase, Meintanis, 2011; Olokoyo, 2013; Salamh, 2011; Ugarte *et al.*, 2009; Wu *et al.*, 2011; Wu y Li, 2014; Wu y Zhu, 2012, entre otros). Badinger y Egger (2013) consideran un marco autorregresivo espacial para la modelización dinámica de las componentes de error. Diferentes contribuciones en el ámbito de la predicción a partir de modelos de regresión espaciales dinámicos se pueden encontrar en los trabajos de Baltagi y Li (2004); Baltagi y Pirotte (2010); Baltagi, Bresson y Pirotte (2012); Baltagi, Fingleton y Pirotte (2014) por mencionar algunos de los más representativos.

Mediante la aplicación de técnicas de la Estadística Espacio-Temporal se pueden incorporar simultáneamente las dos dimensiones en el análisis de datos de panel, así como las correlaciones en espacio y tiempo en términos de modelos más sofisticados de covarianzas espacio-temporales (ver Ugarte *et al.*, 2009, 2010, 2012, en relación con la inferencia estadística a partir de modelos de efectos mixtos espacio-temporales para el análisis de mapas de mortalidad de enfermedades). En este ámbito del análisis estadístico espacio-temporal de datos de panel nos referiremos a

los trabajos de Debarsy, Ertur y LeSage (2012); Parent y LeSage (2010, 2011); Salmerón y Ruiz-Medina (2011), entre otros, así como a las referencias allí reflejadas.

El objetivo de esta tesis es contribuir al análisis de datos de panel en un marco espacio-temporal adoptando un enfoque estadístico funcional en el tiempo o el espacio. Específicamente, las principales aportaciones de esta tesis están relacionadas con el análisis estadístico de curvas con correlación espacial, así como de superficies con correlación temporal. La mayoría de los modelos y técnicas estadísticos funcionales espaciales propuestos en esta tesis han sido motivados por problemas con datos reales en diferentes áreas aplicadas, especialmente, en las áreas de finanzas y de medio-ambiente. Las series temporales autorregresivas Hilbertianas constituyen el marco por excelencia en el desarrollo de modelos funcionales dinámicos temporales en esta tesis. En este contexto nuestra referencia fundamental es el libro de Bosq (2000). Seguidamente nos referiremos brevemente a las referencias más destacadas que han sido consideradas en el desarrollo de esta tesis en el referido marco de la Estadística Funcional Temporal y Espacial.

La disponibilidad un amplio registro de observaciones en numerosas bases de datos en muchos campos de aplicación, tales como medio ambiente, economía, finanzas, educación, cambio climático, seguridad vial, deporte, salud, psicología, medicina, biología, entre otras áreas, motivan una extensa literatura en las últimas décadas en relación con el análisis estadístico de datos funcionales. En un contexto probabilístico, las distribuciones de probabilidad infinito-dimensionales son consideradas como modelos adecuados para caracterizar el comportamiento aleatorio de curvas, superficies, etc. En el contexto de series temporales Hilbert-valuadas, se han considerado diversos enfoques funcionales predictivos según se refleja en los trabajos de Antoniadis, Paparoditis y Sapatinas (2006); Antoniadis, y Sapatinas (2003); Mourid (2002), entre otros. De hecho, las propiedades asintóticas de los estimadores de los operadores de covarianza y operadores de autocorrelación derivados en este contexto, incluyendo consistencia y normalidad asintótica, han sido ampliamente investigadas, por ejemplo, en los trabajos de Besse, Cardot

y Stephenson (2000); Damon y Guillas (2002, 2005); Hörmann y Kokoszka (2011); Horváth, Kokoszka y Reeder (2013), aunque todavía hay muchas líneas de investigación abiertas en este campo. Como se ha comentado, los textos de Bosq (2000) y Bosq y Blanke (2007) constituyen una de las referencias clave en el desarrollo de esta tesis que ha inspirado la mayor parte de los modelos y las metodologías de estimación propuestas en el ámbito temporal y espacial. Estos libros proporcionan un marco de referencia para el desarrollo de la estimación y predicción no paramétrica en series de curvas, así como para la inferencia asintótica, en particular, a partir de procesos autorregresivos Hilbertianos (ver, por ejemplo, Cardot *et al.* 2003; Damon y Guillas, 2005; Mas, 2007; Ruiz-Medina y Salmerón, 2010; Ruiz-Medina, Salmerón y Angulo, 2007; Salmerón y Ruiz-Medina, 2009). Por otro lado, el análisis de componentes principales y el análisis de correlación canónica para variables aleatorias infinito-dimensionales ha sido ampliamente aplicado en el contexto de modelos lineales en espacios de funciones (ver, por ejemplo, el texto clásico de Ramsay y Silverman, 2005).

Un texto clásico en la estadística funcional no-paramétrica es el libro de Ferray y Vieu (2006), (véase también las referencias de Ferraty, Kudraszow y Vieu, 2012; Ferraty y Vieu, 2004). Antoniadis, Paparoditis y Sapatinas (2006) consideran las técnicas de estimación no paramétrica para la regresión funcional. Giraitis, Kapetanios y Yates (2014) introducen una nueva clase de procesos autorregresivos no paramétricos, que se descomponen en una serie temporal con coeficientes variables en el tiempo. Shang (2013) considera la estimación de la componente de error en un modelo de regresión funcional no paramétrico con predictor funcional y respuesta escalar bajo un enfoque bayesiano (véase también, en el contexto de la estadística no paramétrica, Li y Hsing, 2010; Laksaci, Rachdi y Rahmani, 2013; Lian, 2012, entre otros). Aneiros-Pérez, Cao y Vilar-Fernández (2011) consideran el problema de la predicción de series temporales utilizando las técnicas funcionales no paramétricas, realizando una extensión del método de regresión lineal con variables explicativas funcionales. Earls y Hooker (2014) han propuesto un método de

inferencia bayesiana para los datos funcionales caracterizados en el marco de procesos Gaussianos. Angelini, De Canditiis y Pensky (2012) analizan la correlación temporal entre las curvas gaussianas en el ámbito de la regresión no paramétrica, utilizando la teoría de espacios del núcleo reproductor generados a partir de bases de wavelets.

La Estadística Funcional Espacial es una rama relativamente nueva de la estadística permitiendo el análisis de observaciones funcionales correlacionadas espacialmente. En particular, se hará referencia a los trabajos de Guillas y Lai (2010), donde se introducen los modelos de regresión funcional espacial basados en splines bivariantes, y la estimación de mínimos cuadrados es aplicada con/sin término de penalización a partir de la aproximación de las superficies de variables aleatorias explicativas mediante proyección. La inferencia bayesiana a partir de curvas correlacionadas espacialmente se implementa en Baladandayuthapani *et al.* (2008). El análisis de correlación espacial de los datos de alta dimensión desde una perspectiva funcional es considerado en Fan y Zhang (2000). Las técnicas cokriging para datos funcionales espaciales se consideran en Nerini, Monestiez y Manté (2010). Giraldo, Delicado y Mateu (2010) derivan técnicas de predicción espacial funcional, basadas en el *kriging* a partir de curvas mediante proyección en bases de funciones adecuadas. En el contexto de la estadística no paramétrica, Basse, Diop y Dabo-Niag (2008) proponen estimadores no paramétricos tipo núcleo para variables aleatorias funcionales correlacionadas espacialmente. Motivado por los estudios de variables en el océano, Nerini y Ghattas (2010) aplican herramientas funcionales de clasificación estadística para el análisis de curvas con correlación espacial. También mencionamos los trabajos de Giraldo, Delicado y Mateu (2012); Romano, Giraldo y Mateu (2011) y Secchi, Vantini y Vitelli (2011) sobre clasificación estadística funcional de curvas correlacionadas espacialmente. Por último, nos referimos al enfoque bayesiano adoptado en Ma *et al.* (2008) para la derivación de técnicas de clasificación estadística espacial funcional basadas en criterios de información.

Los modelos de series temporales espaciales con valores en un espacio de Hilbert separable son

introducidos en Ruiz-Medina (2011), donde se investigan las condiciones para su estacionariedad e invertibilidad. La estimación no paramétrica de los operadores que definen los parámetros funcionales de la ecuación de estados, que refleja la dinámica espacial de un proceso espacial autorregresivo Hilbertiano de orden uno (proceso SARH(1)), se obtiene en Ruiz-Medina (2012a) a partir de su proyección en los sistemas de autovectores duales de dichos operadores. Una revisión sobre los últimos avances en el campo de la Estadística Espacial Funcional, así como sobre algunos de los posibles problemas abiertos a tratar se refleja en Ruiz-Medina (2012b).

Como se ha comentado anteriormente, el objetivo de la presente tesis es contribuir con nuevos resultados teóricos y computacionales para el desarrollo de nuevas herramientas en el campo de la Estadística Espacial Funcional, así como resolver algunos problemas que aún permanecen abiertos en el campo de la Estadística Funcional Temporal, motivados por su posible aplicación en el análisis de datos de panel. Resumiendo, en esta tesis, se proponen nuevos modelos y metodologías para el análisis estadístico de curvas y superficies correlacionadas espacial y temporalmente. Estas contribuciones pueden ser distribuidas en tres partes según el campo en el que están ubicadas. Más concretamente, la primera parte de la tesis recoge los principales resultados derivados en el área de la Estadística Espacial Funcional para el análisis de curvas correlacionas en el espacio, y en particular, para el análisis de datos de panel bajo esta perspectiva. La segunda parte se recoge los principales resultados obtenidos en el campo de los modelos de series temporales Hilbert-valuadas. En particular, nos centramos en el estudio de la clase de los procesos autorregresivos Hilbertianos de orden uno (procesos ARH(1)). Finalmente, la tercera parte está relacionada con la modelización espacial funcional bajo el enfoque de operadores pseudodiferenciales multifracionarios, proporcionándose una regularización del problema de estimación funcional de mínimos cuadrados, en términos de bases ortogonales de funciones adecuadas, relacionados con el espacio del núcleo reproductor (RKHS), cuyos elementos muestran un comportamiento local con heterogeneidad espacial regular (orden variable de

diferenciación en sentido débil) .

Apéndice A1 implementa el extrapolador SARH(1) definido en Ruiz-Medina (2012a) mediante proyección en los autovectores de los operadores de auto-covarianza del proceso SARH(1). La reducción óptima de la dimensión, en el sentido del máximo nivel de varianza explicada con el mínimo número de términos, se obtiene con la aplicación de esta metodología de proyección. Se ilustra dicha metodoogía de estimación espacial funcional mediante un ejemplo con datos reales relacionados con la extrapolación espacial de las curvas de temperatura de la superficie del océano en diferentes zonas del planeta tierra, donde los efectos del calentamiento global y del cambio climático deben analizarse. Alternativamente, el Apéndice A2 proporciona la implementación del extrapolador SARH(1) en términos de la transformada wavelet discreta con soporte compacto para debilitar la fuerte correlación espacial entre curvas de datos que dificultan el análisis SARH(1), desarrollado anteriormente en términos de los auto-vectores del operador de auto-covarianza. La aplicación de esta transformación está motivada por la elevada concentración espacial de las estaciones meteorológicas en las islas del océano, que inducen una fuerte correlación espacial en las curvas de la temperatura de la superficie del océano observadas. En el apéndice A3, se introduce una nueva clase de modelos de regresión múltiple espacialmente heterogéneos para variables aleatorias con valores en un espacio de Hilbert separable. Se supone que el término de error es un proceso SARH(1). La implementación de los estimadores propuestos para los operadores de regresión y de los parámetros funcionales que definen la ecuación espacial de estados del proceso SARH(1), verificada por el término de error funcional, se derivan en este apéndice. El modelo formulado está motivado por una aplicación financiera. Específicamente, en estudios empíricos previos desarrollados para el análisis de datos de panel en finanzas, se había probado la no estabilidad en el espacio de los coeficientes de regresión que definen la relación lineal entre los factores empresariales y las curvas de endeudamiento. Además, el efecto de los factores institucionales y el sector industrial (ambos aspectos suelen ser introducidos en

el análisis estadístico como variables *dummy*), se refleja usualmente en términos de posibles heterogeneidades en la realción lineal asumida entre las curvas de endeudamiento y factores empresariales. Este hecho ha impulsado el desarrollo de diversos estudios empíricos en el contexto financiero. Ambos aspectos motivan la introducción de una nueva clase de modelos formulados en el Anexo A3, en el contexto Hilbert-valuado, para reflejar las posibles heterogeneidades espaciales en la relación lineal entre los factores empresariales y las curvas de endeudamiento, así como la presencia de una posible correlación espacial entre las curvas de endeudamiento de las empresas dentro de un sector industrial. La dependencia espacial entre las curvas de endeudamiento de las diferentes empresas también puede ser introducida a través de un parámetro funcional aleatorio en el modelo. Este hecho motiva el tema del Apéndice A4, donde una nueva clase de modelos estadísticos espaciales funcionales es derivada en el marco de los efectos mixtos. Una nueva metodología de clasificación para las curvas con correlación espacial es formulada entonces, basada en la introducción y estimación de modelos espaciales funcionales de efectos mixtos, dando lugar a la detección de patrones locales espaciales homogéneos en los operadores de regresión que se hallan invoucrados en la definición de las curvas de efectos fijos, así como se obtiene la detección de patrones locales espaciales homogéneos en la variación local funcional en media cuadrática en el espacio de las curvas de efectos aleatorios. Finalmente, en esta parte de la tesis, se extiende la familias de modelos presentadas para curvas débilmente correlacionadas en el el espacio al ámbito de las series espaciales que permiten representar dependencias espaciales de largo rango. La extensión de la clase de procesos temporales autorregresivos de Gegenbauer al contexto espacial y al contexto Hilbert-valuado se realiza en el Apéndice A5. Sus propiedades estructurales son también analizadas.

En el contexto Gaussiano de los procesos ARH(1), se formula una técnica de estimación funcional penalizada mínimo cuadrática, basada en la geometría del espacio del núcleo reproductor, para estabilizar la ratio de mortalidad estandarizada (SMR) en los mapas de enfermedades. La

motivación de este apéndice surge con un problema con datos reales relacionado con el cáncer de mama en España. Específicamente, los enfoques clásicos basados en los modelos condicionales autorregresivos (modelos CAR) y en las bases de splines, dentro del marco general de los modelos de efectos mixtos, presentan el problema de que producen un suavizamiento local excesivo (también a gran escala) de las curvas de log-riesgo relativo de la mortalidad por cáncer de mama en cada una de las provincias españolas analizadas (pequeñas áreas analizadas). Esta es la razón por la cual, en el Apéndice A6, se presenta un nuevo enfoque para la estabilización de los SMRs, basado en la combinación de los enfoques ARH(1) y RKHS, para interpolar la variabilidad local que se produce en la escala más fina (microescala) que suele ser menos suave que la variación local usualmente reflejada con la diferenciación de orden entero. Se evita así el problema de falsos negativos en la detección de áreas de alto riesgo. Por otra parte, el parámetro de suavizado en la metodología de estimación penalizada propuesta se ajusta por validación cruzada, mientras que en los enfoques mencionados (ver, por ejemplo, Ugarte *et al.*, 2010) este parámetro se estima como una componente de la varianza por el método de máxima verosimilitud (ML). Esta es la razón por la cual en el Apéndice A7 se estudia la metodología de estimación por máxima verosimilitud para procesos ARH(1). Más concretamente, en el apéndice A7, los resultados derivados en Ruiz-Medina y Salmerón (2010) para procesos ARH(1) Gaussianos son extendidos a una familia más general de los procesos ARH(1), cuya distribución infinito-dimensional pertenece a la familia exponencial introducida en este apéndice. Además, la distribución normal asintótica de los estimadores ML dentro de esta familia aleatoria $H-$valuada es derivada.

La tesis finaliza con la introducción de una nueva familia de modelos Gaussianos Hilbert-valuados para curvas, a partir de la teoría de ecuaciones pseudodiferenciales multifraccionarias. Se aborda asimismo el problema de estimación funcional mínimo cuadrática asociado para aproximar los valores funcionales de la solución. Nótese que la regularización directa de este problema mediante aplicación de las técnicas númericas usuales de proyección no es posible, ya que las

funciones del RKHS asociado presentan un orden de variable de diferenciación que depende de la localización espacial. Esta es la razón por la cual se deriva una nueva metodología de proyección en el Apéndice A8 para la regularización de este problema de estimación, lo que permite una inversión continua para el cálculo de la solución. Dicha metodología de proyección está basada en las bases Riesz-duales que proporcionan una diagonalización espectral del operador que genera la forma bilineal cerrada asociada al producto escalar en la correspondiente RKHS.

Antes de los ocho anexos que constituyen los ocho artículos publicados en relación con los contenido de esta tesis, ofrecemos un breve resumen de los objetivos, la metodología y los principales resultados y conclusiones que se derivan de estos apéndices, con el fin de facilitar su lectura proporcionando una visión global y concisa de los aspectos y puntos clave abordados en su desarrollo.

Chapter 3

Objectives

- **Appendix A1.** This appendix has as main objective to provide an alternative projection methodology to the one considered in Ruiz-Medina (2012a) for implementation of the SARH(1) extrapolator. Specifically, the spatial functional extrapolator derived in Ruiz-Medina (2012a) for a SARH(1) process Y, satisfying the state space equation

$$Y_{i,j} = R + L_1(Y_{i-1,j}) + L_2(Y_{i,j-1}) + L_3(Y_{i-1,j-1}) + \epsilon_{i,j}, \tag{3.1}$$

with $R, \epsilon \in H$, and ϵ being strong Hilbertian white noise, is implemented by projection into the common biorthogonal eigenvector systems of the bounded linear operators $L_i \in \mathcal{L}(H)$, $i = 1, 2, 3$, acting on the separable Hilbert space H. These eigenvectors provide a continuous inversion of the following equation system, satisfied by operators L_i, $i = 1, 2, 3$,

$$\begin{aligned} R_{1,0} &= L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1}, \\ R_{0,1} &= L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0}, \\ R_{1,1} &= L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0}, \end{aligned} \tag{3.2}$$

and a spectral diagonalization of equation (3.1). However, they do not provide an optimal dimension reduction in the sense of variance. This is the reason why Appendix 1 has as primary objective to test the eigenvector system of the auto-covariance operator of process Y.

- **Appendix A2.** The presence of strong spatial correlations hinders the SARH(1) statistical analysis based on projection into the auto-covariance eigenvector system. Hence, A2 has as aim to implement SARH(1) extrapolator in terms of compactly supported wavelet bases that weaken correlations and provide a suitable processing of border effects without lengthening the support of the analyzed functional data.

- **Appendix A3.** In Ruiz-Medina (2011) the SARH(p) and MAH(q) models are introduced. They are able to represent homogeneous spatial dynamics and dependence. However, several nature phenomena display heterogeneity in space. This fact motivates the objective of Appendix A3 where a new class of spatial heterogeneous functional multiple regression models is introduced having Hilbert-valued spatially correlated error term satisfying the SARH(1) state space equation (3.1).

- **Appendix A4.** Spatial dependence between curves could also be introduced through a functional random parameter. This is the approach adopted in Appendix A4 that has the objective to formulate a functional statistical classification methodology in the spatial functional mixed effect framework, based on the following model:

$$\mathbf{Y}(\cdot,\mathbf{x}) = \mathbf{X}_{FE}\vec{\beta}(\cdot,\mathbf{x}) + \mathbf{X}_{RE}\vec{\nu}(\cdot,\mathbf{x}) + \sigma\vec{\epsilon}(\cdot,\mathbf{x}), \quad \mathbf{x} \in D \subset \mathbb{R}^n, \tag{3.3}$$

  where $\mathbf{X}_{FE}$ and $\mathbf{X}_{RE}$ represent the respective functional fixed and random effect design matrices. For each spatial location $\mathbf{x} \in D$, each component of the functional vectors $\mathbf{Y}(\cdot,\mathbf{x})$, $\boldsymbol{\beta}(\cdot,\mathbf{x})$, $\vec{\nu}(\cdot,\mathbf{x})$ and $\vec{\epsilon}(\cdot,\mathbf{x})$ lies in the real separable Hilbert space H. In a general setting, D could be an open bounded domain in $\mathbb{R}^n$. $\mathbf{Y}(\cdot,\mathbf{x})$, $\boldsymbol{\beta}(\cdot,\mathbf{x})$, $\vec{\nu}(\cdot,\mathbf{x})$ and $\vec{\epsilon}(\cdot,\mathbf{x})$ respectively denote the vectors of response, fixed effect, random effect and error curves observed at spatial location $\mathbf{x} \in D$.

- **Appendix A5.** In the previous appendixes weak-dependence spatial functional models are analyzed within the time series theory. The aim of Appendix A5 is to extend this

class of models to the strong dependence case. Specifically, autoregressive Gegenbauer processes are introduced in the spatial domain. The case of spatial autoregressive Hilbert-valued Gegenbauer processes is also considered. Note that this class of models can display multiple spectral singularities.

- **Appendix A6.** Inference on ARH(1) processes has been developed in Bosq (2000) under a nonparametric perspective. The objective of Appendix A6 is to provide a penalized RKHS approach for estimation and prediction of ARH(1) processes based on the implementation of Kalman filtering algorithm, from the moment-based empirical estimators of the auto-covariance and cross covariance operators of an ARH(1) process, as well as from the corresponding estimator of the autocorrelation operator, obtained in Bosq (2000).

- **Appendix A7.** Maximum likelihood estimation of Gaussian ARH(1) processes has been implemented in Ruiz-Medina and Salmerón (2010) by combination of Kalman filtering with Expectation Maximization algorithm (EM algorithm). Appendix A7 has an primary objective to extend maximum likelihood estimation for ARH(1) processes from the Gassian context to the Exponential Hilbert-valued family framework introduced. The asymptotic normal distribution of the derived maximum likelihood estimator is derived as well.

- **Appendix A8.** A new class of Gaussian fractional generalized random fields is introduced in terms of multifractional pseudodifferential equations in Ruiz-Medina, Anh and Angulo (2004). The Hilbert-valued formulation of these models constitutes the aim of Appendix 8, where the regularization of the associated least-squares estimation problem is addressed, in terms of projection into dual Riesz bases that provide a spectral diagonalization of the auto-covariance operator of the solution, and they generate the associated RKHS constituted by functions with variable order of differentiation.

Chapter 4

Methodology

- **Appendix A1.** The estimation methodology proposed in this appendix consists of the following steps:

  Step 1. To project equation system (3.2) into the auto-covariance eigenvector system ϕ_k, $k \geq 1$, satisfying

  $$R_X \phi_k = \lambda_k, \quad k \geq 1,$$

  where $R_X = E[X_t \otimes X_t] = E[(Y_t - R) \otimes (Y_t - R)]$, with $E[Y_t] = R \in H$, for all $t \geq 0$, and Y_t being a SARH(1) process satisfying state space equation (3.1).

  Step 2. Given a truncation order M, a finite-dimensional approximation of equation system (3.2) is obtained, whose matrix solution is defined as follows:

  $$\begin{bmatrix} \widehat{\Lambda}(L_1) \\ \widehat{\Lambda}(L_2) \\ \widehat{\Lambda}(L_3) \end{bmatrix} = \begin{bmatrix} \Lambda(\hat{R}_{0,0}) & \Lambda(\hat{R}_{1,1}) & \Lambda(\hat{R}_{0,1}) \\ \Lambda(\hat{R}_{1,1}) & \Lambda(\hat{R}_{0,0}) & \Lambda(\hat{R}_{1,0}) \\ \Lambda(\hat{R}_{0,1}) & \Lambda(\hat{R}_{1,0}) & \Lambda(\hat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \Lambda(\hat{R}_{1,0}) \\ \Lambda(\hat{R}_{0,1}) \\ \Lambda(\hat{R}_{1,1}) \end{bmatrix}, \tag{4.1}$$

  where, for a Hilbert-Schmidt operator K on $\mathcal{H}$, $\Lambda(K)$ represents the matrix whose entries are the coordinates of operator K with respect to the tensorial product basis

  $$\{\phi_l \otimes \phi_m,\ l, m = 1, \dots L\},$$

  i.e., the matrix with entry (i, j), given by $\lambda_{i,j}(K)$, for $i, j = 1, \dots L$. Also, $\widehat{\Lambda}(L_k)$ denotes the estimate of matrix $\Lambda(L_k)$, for $k = 1, 2, 3$.

Note that, the presented projection methodology is selected because it provides a sparse finite-dimensional version of (3.2) after truncation, since $\Lambda(\hat{R}_{0,0})$ are diagonal matrices.

- **Appendix A2.** The projection methodology adopted in this appendix for implementation of the SARH(1) extrapolator

$$\begin{aligned} & E\left[Z_{i,j} \mid Z_{k,l},\ k<i,\ \text{or } l<j,\ \text{or } k<i \text{ and } l<j\right] \\ & = L_1 Z_{i-1,j} + L_2 Z_{i,j-1} + L_3 Z_{i-1,j-1} \\ & \simeq \widehat{L_1} Z_{i-1,j} + \widehat{L_2} Z_{i,j-1} + \widehat{L_3} Z_{i-1,j-1} \end{aligned} \tag{4.2}$$

is based on the compactly supported wavelet transform, which as commented before, it weakens correlations and provides a suitable processing of border effects. The resulting finite dimensional approximation of the solution to the SARH(1) estimation problem in the wavelet domain is given by

$$\begin{bmatrix} \widehat{\mathcal{W}_{2D}^M(L_1)} \\ \widehat{\mathcal{W}_{2D}^M(L_2)} \\ \widehat{\mathcal{W}_{2D}^M(L_3)} \end{bmatrix} = \begin{bmatrix} \mathcal{W}_{2D}^M(\hat{R}_{0,0}) & \mathcal{W}_{2D}^M(\hat{R}_{1,1}) & \mathcal{W}_{2D}^M(\hat{R}_{0,1}) \\ \mathcal{W}_{2D}^M(\hat{R}_{1,1}) & \mathcal{W}_{2D}^M(\hat{R}_{0,0}) & \mathcal{W}_{2D}^M(\hat{R}_{1,0}) \\ \mathcal{W}_{2D}^M(\hat{R}_{0,1}) & \mathcal{W}_{2D}^M(\hat{R}_{1,0}) & \mathcal{W}_{2D}^M(\hat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{W}_{2D}^M(\hat{R}_{1,0}) \\ \mathcal{W}_{2D}^M(\hat{R}_{0,1}) \\ \mathcal{W}_{2D}^M(\hat{R}_{1,1}) \end{bmatrix}$$

where $\mathcal{W}_{2D}^M$ denotes the two-dimensional discrete interval wavelet transform at resolution level M.

- **Appendix A3.** The following spatial functional multiple regression model with regression operators depending on space and with functional response and regressors is introduced in this appendix

$$Z(i,j) = \mathcal{L}_1(i,j)X_1 + \ldots + \mathcal{L}_q(i,j)X_q + \varepsilon(i,j), \tag{4.3}$$

where the error term $\varepsilon(i,j)$ is assumed to be a SARH(1) process. The regression operators $\mathcal{L}_i$, $i = 1, \ldots, q$, are estimated from the equation system

$$\begin{aligned}
\widehat{R}_1(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{11} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{12} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{1q}, \\
\widehat{R}_2(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{21} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{22} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{2q}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\widehat{R}_q(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{q1} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{q2} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{qq},
\end{aligned} \tag{4.4}$$

for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$. The projection into a suitable orthogonal or biorthogonal basis leads to the following finite-dimensional inversion of (A3.23):

$$\begin{aligned}
&\begin{bmatrix} \Phi_M^*(i,j)L_1(i,j)\Phi_M(i,j) \\ \ldots \\ \Phi_M^*(i,j)L_q(i,j)\Phi_M(i,j) \end{bmatrix} \\
&\quad = \begin{bmatrix} \mathbf{C}\left(\hat{R}_{1,1}\right) & \mathbf{C}\left(\hat{R}_{1,2}\right) & \ldots & \mathbf{C}\left(\hat{R}_{1,q}\right) \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{C}\left(\hat{R}_{q,1}\right) & \mathbf{C}\left(\hat{R}_{q,2}\right) & \ldots & \mathbf{C}\left(\hat{R}_{q,q}\right) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}\left(\hat{R}_1(i,j)\right) \\ \ldots \\ \mathbf{C}\left(\hat{R}_q(i,j)\right) \end{bmatrix},
\end{aligned} \tag{4.5}$$

where, as before, for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, $\mathbf{C}(\mathcal{K}(i,j))$ represents the coordinate matrix of operator $\mathcal{K}(i,j)$ with respect to the basis $\left\{\widehat{\phi}_l^{\widehat{R}_0(i,j)} \otimes \widehat{\phi}_m^{\widehat{R}_0(i,j)},\ l,m = 1,\ldots M(i,j)\right\}$. Functional parameter estimation of the corresponding Hilbert-valued residuals, after fitting model (A3.21), is then achieved through the application of the SARH(1) projection estimation methodology derived in the previous appendices.

- **Appendix A4.** The spatial functional classification methodology adopted here is based on the spatial functional mixed effect model (3.3). Specifically, the functional design matrices of the fixed and random effect curves are assumed to be the

identity matrices in the respective separable Hilbert spaces. Additionally, it is assumed that the fixed effect curves display a linear spatially heterogeneous relationship with a set the functional regressors, represented in terms of a spatially dependent multiple regression model like the one introduced in Appendix A3. The random effect curves satisfy a SARH(1) equation, and the Hilbert-valued error term is Hilbert-valued strong Gaussian white noise. After application of the functional multiple regression estimation methodology derived in Appendix A3, and the SARH(1) estimation methodology implemented in Appendix A1-A2, the proposed spatial functional classification methodology is applied. Specifically, fixed effect curves are classified according to the Hilbert-Schmidt distances between their estimated vectors of regression operators. While random effect curves are classified by computation of a wavelet approximation of the empirical SARH(1) variogram given in terms of the estimated SARH(1) parameters. Summarizing, by the application of the proposed functional classification methodology, local homogeneous patterns in space are detected in the functional linear relationship of the fixed effect curves with the regressors, as well as in the spatial local variation of the random effect curves.

- **Appendix A5.** Inspired by the work of Gray, Zhang and Woodward (1989) and Woodward, Cheng and Gray (1998), the model class introduced in Appendix 5 extends to the spatial and Hilbert-valued contexts the family of autoregressive Gegenbauer processes in time. The methodological approach adopted extends the one presented in the cited papers for the introduction of temporal autoregressive Gegenbauer processes. Specifically, the methodological extension presented in this appendix allows the introduction of the following families of spatial processes:

  i) *Spatial Gegenbauer processes*, for $(t_1, t_2) \in \mathbb{Z}^2$,

$$\nabla_{u_1}^{d_1} \circ \nabla_{u_2}^{d_2} X_{t_1,t_2} = \left(I - 2u_1 B_1 + B_1^2\right)^{d_1} \circ \left(I - 2u_2 B_2 + B_2^2\right)^{d_2} X_{t_1,t_2} = \varepsilon_{t_1,t_2}, \quad (4.6)$$

where B_i, for $i = 1, 2$, denotes the backward-shift operators for each spatial coordinate, that is, $B_1 X_{t_1,t_2} = X_{t_1-1,t_2}$, and $B_2 X_{t_1,t_2} = X_{t_1,t_2-1}$, and ε_{t_1,t_2} is white noise.

ii) *Spatial autoregressive Gegenbauer process*, for $(t_1, t_2) \in \mathbb{Z}^2$,

$$\phi(B_1, B_2)(I - 2u_1 B_1 + B_1^2)^{d_1}(I - 2u_2 B_2 + B_2^2)^{d_2}(Y_{t_1,t_2} - \mu) = \varepsilon_{t_1,t_2}.$$

where ϕ is the spatial autoregressive operator.

iii) *Spatial autoregressive Gegenbauer process with multiple singularities*,

$$\phi(B_1, B_2) \prod_{j=1}^{k-1} \left(I - 2B_1 \cos(\mu_1^j) + B_1^2\right)^{d_{j1}} \prod_{j=1}^{k-1} \left(I - 2B_2 \cos(\mu_2^j) + B_2^2\right)^{d_{j2}} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}.$$

iv) *Hilbert-valued spatial autoregressive Gegenbauer process*, for $(i, j) \in \mathbb{Z}^2$,

$$\phi(B_1, B_2)(I - 2M_1 B_1 + B_1^2)^{d_1}(I - 2M_2 B_2 + B_2^2)^{d_2}(Y_{i,j} - R) = \varepsilon(i, j),$$

where M_i, $i = 1, 2$, are bounded linear operators on $\mathcal{H}$, and R, $Y_{i,j}$, and $\varepsilon(i, j) \in \mathcal{H}$.

- **Appendix A6** Goicoa, Ugarte, Etxeberria and Militno, (2012) develop a comparative study between CAR and P-spline modeling through a mixed-effect approach for smoothing and stabilization of mortality risks in disease mapping. In Appendix A6, we also adopt a mixed-effect framework in the Hilbert-valued context to stabilize standardized mortality ratios in disease mapping. The RKHS-based penalized estimation methodology of ARH(1) processes derived in this appendix considers the following minimization problem:

$$\min_{F(\cdot)\in\mathcal{H}(Z)} \frac{1}{T} \sum_{t=1}^{T} \|\beta(\cdot) + u_t(\cdot) - F(Z_1, \ldots, Z_t)\|^2_{\mathcal{H}_0(Z)} + \gamma \|\mathbf{\Phi}_{\mathcal{H}_1(Z)} (F(Z_1, \ldots, Z_t))\|^2_{\mathcal{H}_1(Z)}, \tag{4.7}$$

where $\mathcal{H}(Z)$ denotes the RKHS generated by the kernel $\mathcal{K}_Z$ of the auto-covariance operator $R_0^Z = E[(Z_t - \beta) \otimes (Z_t - \beta)] = E[(Z_1 - \beta) \otimes (Z_1 - \beta)]$, for $t = 1, \ldots, T$, with T denoting

the number of observed times. Here, the Hilbert-valued process u satisfies the ARH(1) equation

$$u_t = \mathcal{A}(u_{t-1}) + \nu_t, \quad t \in \mathbb{Z},$$

with ν_t denoting a Hilbert-valued Gaussian process in the strong sense representing the functional innovations. The H-valued observation model is defined in a mixed effect framework as follows:

$$Z_t(i) = \beta(i) + u_t(i) + \varepsilon_t(i), \quad i = 1, \ldots, l, \quad t = 1, \ldots T, \tag{4.8}$$

where $\beta, Z_t, u_t, \varepsilon_t \in H$, for $t = 1, \ldots, T$, $H = L^2(D)$, $D \subset \mathbb{R}^2$. The functional fixed effect parameter β_i does not depend on time, while the random effect parameter u_t is an ARH(1) process. The zero-mean H-valued Gaussian process ε_t satisfies

$$E[\varepsilon_t(i)\varepsilon_s(j)] = \delta(t-s)\delta(i-j), \quad \forall i, j = 1, \ldots, l, \quad t, s \in \mathbb{R}_+.$$

The index i refers to specific spatial coordinates $\mathbf{x}_i = (x_{i1}, x_{i2}) \in D$ associated with the i region studied, for $i = 1, \ldots, l$. Note that we are assuming that the support of the spatial functions in H includes the locations $\mathbf{x}_i = (x_{i1}, x_{i2})$, $i = 1, \ldots, l$, whose coordinates define the centroid associated with each one of the small areas under study. In addition, the two norms involved in equation (A6.3) are respectively given in term of the spaces $\mathcal{H}_0(Z)$ and $\mathcal{H}_1(Z)$, whose direct sum defines the RKHS $\mathcal{H}(Z) = \mathcal{H}_0(Z) \oplus \mathcal{H}_1(Z)$ of Z. It is well-known (see, for example, Bosq, (2000)), that $\mathcal{H}(Z)$ can be isometrically identified with the closed subspace $H(Z)$ of $\mathcal{L}_H^2(\Omega, \mathcal{A}, P)$ generated by the zero-mean Gaussian Hilbert-valued random variables $Z_t - \beta$, $t \in \mathbb{R}_+$, and their limits in the mean-square sense, where $(\Omega, \mathcal{A}, P)$ denotes, as usual, the basic probability space. Thus, $H(Z) \equiv \mathcal{H}(Z)$. The inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}(Z)}$ in $\mathcal{H}(Z)$ is defined from the closed bilinear form generated by operator $[R_0^Z]^{-1}$, the inverse of operator R_0^Z, that is,

$$\langle f, g \rangle_{\mathcal{H}(Z)} = [R_0^Z]^{-1}(f)(g), \quad \forall f, g \in \mathcal{H}(Z). \tag{4.9}$$

Appendix A7. In the Appendix A7, a generalization of an invariance principle for Robbins-Monro type processes, formulated in terms of Hilbert-valued martingale differences, is applied to provide the convergence to the gaussian distribution of the maximum-likelihood estimator of the auto-covariance operator

$$(R_\nu) = E[\nu_t \otimes \nu_t], \quad t \in \mathbb{Z},$$

of the innovation process ν involved in the definition of the state space equation of an ARH(1) process, given by

$$Z_t = Y_t(\cdot) - \mu(\cdot) = \mathcal{A}\left(Z_{t-1}\right)(\cdot) + \nu_t(\cdot), \quad t \in \mathbb{Z}. \tag{4.10}$$

Here, as before, $\mu, Y_t, \nu_t \in H$, $t \in \mathbb{Z}$, with H being a separable Hilbert space. The innovation process ν is assumed to be a zero-mean H-valued martingale difference process. The infinite-dimensional parameter $\mathcal{A}$ lying in the space of bounded linear operators on H is the autocovariance operator. The functional parameter vector $(\mathcal{A}, R_\nu) \in \mathcal{L}(H) \times \mathcal{S}(H)$, with $\mathcal{S}(H)$ denoting the space of Hilbert-Schmidt operators on H. Appendix A7 provides sufficient conditions for the asymptotic normal distribution of the maximum likelihood estimator $\widehat{\Theta} = \widehat{R}_\nu = \frac{1}{T}\sum_{i=1}^{T} \nu_i \otimes \nu_i$, of R_ν, i.e.,

$$\sqrt{T}\left(\widehat{\Theta} - \Theta\right) \underset{d}{\longrightarrow} Y \sim \mathcal{N}\left(0, \mathcal{R}_{T_\nu}\right), \quad T \to \infty,$$

where, as before, $\Theta = R_\nu$, and $\mathcal{R}_{T_\nu}$ is the covariance operator of the sufficient statistics $T_\nu = \nu \otimes \nu$, for the introduced Hilbert-valued exponential family. Note that $\{\nu_t \otimes \nu_t\}_{t\in\mathbb{N}}$ defines an $\mathcal{S}(H)$-valued martingale difference sequence under the conditions assumed.

- **Appendix A8.**

In Ruiz-Medina and Fernandez-Pascual (2010) a Hilbert-valued context is adopted for regularization of the least-squares linear estimation problem for the solution to a fractional

pseudodifferential equation with H-valued Gaussian innovations. The fractal order of such a solution defines the suitable order of the dual fractional Sobolev space involved in the definition of the biothorgonal Riesz bases for regularization. This Appendix A8 extends the above referred regularization methodology in terms of fractional Sobolev spaces of variable order. The integral form of the state equation considered in our formulation of the functional least-squares estimation problem for multifractional Gaussian random fields is given by

$$\mathcal{Y}_{\beta(\mathbf{z})}(\mathbf{z}) = A\mathcal{S}_{\gamma(\cdot)}(\mathbf{z}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\lambda}\rangle} p_A(\mathbf{z},\boldsymbol{\lambda}) \widehat{\mathcal{S}}_{\gamma(\cdot)}(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{4.11}$$

for all $\mathbf{z} \in S \subseteq \mathbb{R}^n$, where $\gamma(\cdot)$ denotes the variable order of the fractional Sobolev space related to the RKHS of the multifractional random signal $\mathcal{S}_{\gamma(\cdot)}$. The Hilbert-valued process $\mathcal{Y}_{\beta(\cdot)}$ denotes the output of a linear system given in terms of a pseudodifferential operator of variable order $A = \beta(\mathbf{z}) - \gamma(\cdot)$, with random input $\mathcal{S}_{\gamma(\cdot)}$. The functional least-squares estimator $\widehat{\mathcal{Y}}_{\beta(\cdot)} = K\mathcal{X}_{\alpha(\cdot)}(\cdot)$ of $\mathcal{Y}_{\beta(\cdot)}$ is defined for K being the solution to the minimization problem

$$\mathrm{MSE}(K'\phi) := \mathcal{E}(K'\phi) = E[\mathcal{Y}_{\beta(\cdot)}(\phi) - \mathcal{X}_{\alpha(\cdot)}(K'\phi)]^2, \tag{4.12}$$

for all $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$, where K' is the adjoint of the operator K. Here, the H-valued observation process $\mathcal{X}_{\alpha(\cdot)}$ satisfies

$$\mathcal{X}_{\alpha(\cdot)}(\varphi) = S_{\gamma(\cdot)}(\varphi) + \mathcal{N}_{\theta(\cdot)}(\varphi), \quad \forall \varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}})$$

(respectively,

$$\mathcal{X}_{\alpha(\cdot)}(\varphi) = \mathcal{Y}_{\beta(\cdot)}(\varphi) + \mathcal{N}_{\theta(\cdot)}(\varphi), \quad \varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}}),$$

in the inverse formulation). The observation noise $\mathcal{N}_{\theta(\cdot)}$ is a generalized random field with minimum multifractional singularity order $-\theta(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$, and $-\alpha(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$

represents the minimum multifractional singularity order of the observation random field $\mathcal{X}$ defining $\mathcal{X}_{\alpha(\cdot)}$ on $H^{\alpha(\cdot)}(\mathbb{R}^n)$, with $\mathcal{B}^\infty(\mathbb{R}^n)$ being the space of all C^∞-functions on $\mathbb{R}^n$, whose derivatives of all orders are bounded, and $H^{\alpha(\cdot)}(\mathbb{R}^n)$ denotes the fractional Sobolev space of variable order $\alpha(\cdot)$ on $\mathbb{R}^n$. Depending on the conditions assumed on the functional parameter space where the variable order of differentiation (in the weak-sense) $\alpha(\cdot)$ lies, different projection methodologies can be applied, respectively based on the eigenvectors of the autocovariance operator of $\mathcal{X}_{\alpha(\cdot)}$, when $\min_{\mathbf{x}\in\mathbb{R}^n}\alpha(\mathbf{x}) > n/2$, and based on dual Riesz bases, when the equivalence of the norm of the RKHS of $\mathcal{X}_{\alpha(\cdot)}$ with the fractional Sobolev space of variable order $\alpha(\cdot)$ holds. Both projection methodologies lead to a suitable regularization or continuous inversion of the estimation problem for the computation of the functional least-squares estimator $\widehat{\mathcal{Y}}_{\beta(\cdot)} = K\mathcal{X}_{\alpha(\cdot)}(\cdot)$ defined from (4.12).

Chapter 5

Results

- **Appendix A1.** The SARH(1) extrapolator is implemented by projection into the eigenvector system of the empirical auto-covariance operator of the SARH(1) process. Under suitable conditions, the empirical eigenvector system constitutes a consistent estimator of the true auto-covariance eigenvectors, hence, this basis provides an optimal dimension reduction in the sense of the percentage of explained variance. This projection estimation methodology has been tested with a real-data example where spatial functional extrapolation of ocean surface temperature curves is achieved. Weather stations usually present a non-regular distribution in space. That is, they are irregularly distributed. Hence, spatial interpolation techniques should be applied to implement SARH(1) extrapolator defined on a spatial regular grid, or, as in our case, areal functional data are constructed by block averaging. This fact constitutes one of the main drawback of the SARH(1) estimation methodology implemented.

- **Appendix A2.** The presence of strong spatial correlations in the data curve analyzed hinders the statistical analysis performed with spatial Markovian models like the SARH(1) model. Appendix A2 partially solves this problem by application of the compactly supported wavelet transform to the analyzed curve data removing strong correlations and

eliminating border effects. However, we have to note that the results of Appendix A1 are improved in the case where the weather stations display high spatial concentration like in the area of Hawaii Ocean, but in other coastal ocean areas like the Gulf of Mexico with a more regular and sparse distribution of weather stations (see also east and west coast of Australia), the empirical auto-covariance eigenvector based projection methodology outperforms the wavelet transform when lengthening of the support of the analyzed functional data is considered.

- **Appendix A3.** The spatial heterogeneous functional multiple regression model with SARH(1) error term introduced in this appendix provides a flexible framework for the statistical analysis of weak-dependent spatial correlated curve data. In particular, the proposed model allows a richer and more informative analysis than the classical linear models used in the empirical studies developed from firm panel data. This fact has been illustrated with a real-data example where a panel data constituted by SMEs Spanish companies from different industry sectors is analyzed during a temporal period. The suitability of the formulated model family and projection methodologies is tested by cross validation.

- **Appendix A4.** The spatial functional mixed effect framework introduced in this appendix allows spatial heterogeneities in trends and mean-quadratic functional local variation. Both aspects are taken into account for detection of local homogeneous patterns in fixed and random functional parameters after model fitting. The proposed classification methodology allows to reduce the problem of inference on the introduced class of spatially heterogeneous functional linear models to the spatial homogeneous case over the clusters distinguished by both, fixed effect and random effect, classification procedures. The interest of the proposed spatial functional classification procedure is illustrated with

a real-data example where Spanish companies SMEs analyzed in a panel data are grouped according to their homogeneity in the linear relationship between indebtedness curves and firm factors within an industry sector, and according to their homogeneity in relation to the mean-quadratic functional local variability of their indebtedness curves.

- **Appendix A5.** The previous model families introduced in a spatial functional framework are weak dependent in space. However, in this appendix the spatial autoregressive Gegenbauer processes studied allows, in addition, to represent long-range dependence behaviors in space, extending the classical case of an isolated spectral singularity at zero frequency. This more flexible class, that includes as particular case weak-dependence autoregressive behaviors in space, open a new line of investigation for the statistical analysis of strong spatially correlated curves.

- **Appendix A6.** A RKHS based penalized estimation methodology is derived for ARH(1) processes. The resulting spatial functional smoothing performed at each year analyzed on mortality risk maps interpolates the local variation of standardized mortality ratio maps, and the CAR or spline smoothing with a closer behavior of CAR smoothing. This approach then has the advantage of avoiding false negatives, but the proportion of false positives is larger than in the CAR and spline based approaches.

- **Appendix A7.** The asymptotic properties of maximum likelihood functional parameter estimators of ARH(1) processes have not been derived until now, due to the lack of a explicit finite expression for the probability density functional characterizing the infinite-dimensional random variable. This Appendix constitutes a first attempt for the special case of exponential family studied in a Hilbert-valued framework beyond the Hilbert-valued Gaussian case.

- **Appendix A8.** Fractional pseudodifferential operators of variable order define high lo-

cally singular filters that must be regularized in terms of suitable bases with variable order of differentiation, in order to obtain a continuous inversion. The regularity of the exponent function defining the variable order of differentiation also constitutes a key aspect in the implementation of the least-squares functional estimator in this context. Both aspect have been conjugated in this appendix for defining an almost continuous solution to the least squares estimation problem by application of suitable numerical projection methods.

Chapter 6

Conclusions

- **Appendix A1 and Appendix A2.** Summarizing from Appendix A1 and A2, we conclude that the SARH(1) extrapolator is very sensitive to the bases used for projection, as well as an additional error, due to spatial interpolation or averaging by blocks, should be incorporated when the sample stations are irregularly distributed in space. Hence, we have to apply some cross-validation methodologies, or to solve by some criteria the model selection problem associated with the choice of a suitable basis for projection.

- **Appendix A3.** As in Appendices A1 and A2, the sensitiveness to the orthogonal basis used for projection must be taken into account, as well as the sensitiveness to the truncation order selected depending on the local regularity and moment conditions of the functions conforming the basis chosen, and of the curve data analyzed. The computational cost also depends on the basis selected regarding the quality of the approximation obtained for a given truncation level, since heterogeneous spatial behaviors should be approximated in terms of a common orthogonal basis to obtain a real dimension reduction of the functional parameter estimation problem. In that sense, it should be advisable the use of unconditional bases like certain wavelet bases providing a sparse and optimal decomposition of functions in several Hilbert spaces (e.g., fractional Besov spaces).

- **Appendix A4.** The kernel-based estimation of the auto-covariance and cross-covariance operators in terms of an unconditional basis applied here outperforms the estimation results obtained in Appendix 3 in terms of empirical auto-covariance operator eigenvectors, and biorthogonal bases associated with the RKHS.

- **Appendix A5.** Inference on this class of processes Gegenbauer constitutes a difficult task. Several frameworks could be adopted. However, we have chosen the minimum contrast estimation methodology since it can be formulated in the spectral domain leading to an easy treatment of these processes in the weak-sense in terms of suitable spectral weight functions. Thus, the contrast function and contrast process are formulated in terms of the unbiased tapered periodogram (see Espejo *et al.*, 2014, and the following section on actual reaerch lines.

- **Appendix A6.** The penalized ARH(1) estimation approach presented can be improved by incorporating additional functional parameters in the observation model, corresponding to the spatial random functional effect and the spatiotemporal interaction random functional effect.

- **Appendix A7.** Maximum likelihood estimation in the context of ARH(1) processes beyond the Gaussian case should be addressed in the weak-sense in terms of suitable test functions (or, equivalently, absolute convergent series).

- **Appendix A8.** Several open problems must be addressed in this framework, in particular, in relation to the definition of variable order pure point spectra of symmetric integral covariance operators with continuous kernel having compact support. This aspect opens a new line of research where new Gaussian Hilbert-valued infinite-dimensional distributions can be properly introduced in terms of infinite products of variable order one-dimensional Gaussian measures.

Chapter 7

Conclusiones

- **Apéndice A1 and Appendix A2.**

  Resumiendo el Apéndice A1 y A2, llegamos a la conclusión de que el extrapolador SARH(1) es muy sensible a las bases utilizadas para la proyección, así como se tiene en la implementación un error adicional, debido a la interpolación espacial o al promedio por bloques que se debe realizar cuando las estaciones espaciales están irregularmente distribuidas. Por lo tanto, se deben aplicar metodologías de validación cruzada, o bien, aplicar criterios apropiados de selección de modelos para determinar la base más adecuada, de acuerdo a las características de regularidad local y momentos, así como a la distribución espacial de las curvas, o datos funcionales analizados.

- **Apéndice A3.**

  Como en los Apéndices A1 y A2, un aspecto esencial del análisis estadístico, desarrollado en el marco de los modelos espaciales de regresión múltiple funcional introducidos, es la influencia de la base ortogonal utilizada para la proyección, así como la sensibilidad al orden de truncamiento, que dependerá, en parte, de la base seleccionada y de las características de las curvas analizadas, así como del grado de correlación espacial presente. La calidad de la aproximación o estimaciones derivadas para un nivel de truncamiento, en

presencia de comportamientos espaciales heterogéneos, debe ser modulada de acuerdo al coste computacional adicional que supone la consideración de sistemas de autovectores que dependen del espacio. En ese sentido, es aconsejable el uso de bases incondicionales, tales como ciertas bases de wavelets que proporcionan una descomposición óptima de funciones en diversos espacios de Hilbert, con mínimo núnero de proyecciones asociadas (este es el caso de bases de wavelets en determinados espacios de Besov fraccionarios).

- **ApéndiceA4.**

  La estimación tipo núcleo de los operadores de auto-covarianza y covarianza cruzada, en términos de bases incondicionales, efectuada en este apéndice, mejora los resultados de estimación obtenidos en el Apéndice A3 en términos de autovectores empíricos del operador de auto-covarianza, y de las bases biortogonales asociadas al espacio del núcleo reproductor.

- **Apéndice A5.**

  La inferencia a partir de la clase de procesos de Gegenbauer con dinámica autorregresiva presenta una serie de dificultades, dadas las múltiples singularidades espectrales de dichos procesos. Numerosos marcos pueden ser adoptados en el desarrollo de dicha inferencia. En nuestro caso, hemos optado por la metodología de estimación de mínimo contraste, ya que puede ser formulada en el dominio espectral, lo que permite un fácil tratamiento de estos procesos en el sentido débil, en términos de funciones de peso espectrales adecuadas. Por tanto, la función de contraste y el proceso de contraste son formulados en términos del periodograma insesgado *tapered* (véase Espejo *et al.*, 2014, y la siguiente sección sobre líneas de investigación actuales.)

- **Apéndice A6.**

  La estimación ARH(1) penalizada presentada, basada en la geometrí a del espacio del

núcleo reproductor, puede ser mejorada mediante la incorporación de parámetros funcionales adicionales en el modelo de observación, correspondientes al efecto funcional aleatorio espacial y al efecto funcional aleatorio de interacción espacio-temporal.

- **Apéndice A7.**

  Un marco apropiado para el desarrollo de la estimación mediante máxima verosimilitud en el contexto de procesos ARH(1), en el caso no Gaussiano, se obtiene cuando se considera su formulación en sentido débil, en términos de funciones de contraste adecuadas, que garantizan la convergencia de las series infinito-dimensionales involucradas, o bien, de las integrales impropias que se requieren para la definición de los correspondientes funcionales de densidad de probabilidad en espacios de distribuciones *tempered*.

- **Apéndice A8.**

  Se encuentran pendientes de abordar diversos problemas abiertos en este contexto, en particular, en relación con la definición de un espectro puntual de orden variable, dependiente del espacio, para operadores de covarianza integrales simétricos con núcleos continuos que tienen soporte compacto y orden de reglaridad local cambiante en el espacio. Este aspecto abre una nueva línea de investigación donde las nuevas distribuciones infinito-dimensionales Hilbert-valuadas Gaussianas asociadas pueden introducirse adecuadamente en términos de productos infinitos de variables de Gaussianas independientes con varianzas heterogéneas espacialmente.

Chapter 8

Actual Research Lines

8.1 Objective

In connection with Appendix A5 was pending the estimation of parameter involved in the spatial Gegenbauer model. In Espejo *et al.* (2014) the minimum contrast methodology has been adopted for such estimation, obtaining the process and the corresponding contrast function. The consistency and the asymptotic normality of the minimum contrast estimator designed in the formulated model has been tested. Following, a summary of the methodology used and results obtained are given in the next sections.

8.2 Methodology

Let $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ be a non-random real-valued function, usually referred as the *contrast function*, given by

$$K(\boldsymbol{\theta_0}, \boldsymbol{\theta}) := \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \frac{\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta_0})}{\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})} \, d\boldsymbol{\lambda},$$

and let the *contrast field* be

$$U(\boldsymbol{\theta}) := -\int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda},$$

where $\boldsymbol{\theta_0} = (d_{10}, d_{20})$ is the true parameter value.

and the corresponding empirical *contrast field* is

$$\hat{U}_T^*(\boldsymbol{\theta}) = -\int_{[-\pi,\pi]^2} I_T^*(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\ d\boldsymbol{\lambda}.$$

The conditions for consistency and asymptotic normality are verified by the following conditions **A1-A9** basing at work introduced by Anh, Leonenko, and Sakhno (2004), to satisfy the consistency is necessary to satisfy the conditions **A1-A6** and and additionally the conditions **A7-A9** for the asymptotic normality.

A1. Let $Y_{t_1,t_2},\ \boldsymbol{t} = (t_1,t_2) \in \mathbb{Z}^2$, be a real-valued measurable stationary Gaussian random field with zero mean and a spectral density $f(\boldsymbol{\lambda},\boldsymbol{\theta})$, where $\boldsymbol{\lambda} = (\lambda_1,\lambda_2) \in [-\pi,\pi]^2$, $\boldsymbol{\theta} \in \Theta$, and Θ is a compact set. Assume that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, where $\boldsymbol{\theta}_0$ is the true value of the parameter vector $\boldsymbol{\theta}$.

A2. If $\boldsymbol{\theta_1} \neq \boldsymbol{\theta_2}$ then $f(\boldsymbol{\lambda},\boldsymbol{\theta_1}) \neq f(\boldsymbol{\lambda},\boldsymbol{\theta_2})$ for almost all $\boldsymbol{\lambda} \in [-\pi,\pi]^2$ with respect to the Lebesgue measure.

A3. There exists a nonnegative function $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi,\pi]^2$, such that

1. $w(\boldsymbol{\lambda})$ is symmetric about $(0,0)$, i.e. $w(\boldsymbol{\lambda}) = w(-\boldsymbol{\lambda})$;
2. $w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda},\boldsymbol{\theta}) \in L_1\left([-\pi,\pi]^2\right)$ for all $\boldsymbol{\theta} \in \Theta$.

A4. The derivatives $\nabla_{\boldsymbol{\theta}} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})$ exist and it is legitimate to differentiate under the integral sign in equation (A9*), i.e.

$$\nabla_{\boldsymbol{\theta}} \int_{[-\pi,\pi]^2} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda})\ d\boldsymbol{\lambda} = \int_{[-\pi,\pi]^2} \nabla_{\boldsymbol{\theta}} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda})\ d\boldsymbol{\lambda} = 0.$$

A5. For all $\boldsymbol{\theta} \in \Theta$ the function $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi,\pi]^2$, satisfies

$$f(\boldsymbol{\lambda},\boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) \in L_1\left([-\pi,\pi]^2\right) \cap L_2\left([-\pi,\pi]^2\right).$$

A6. There exists a function $\upsilon(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, such that

1. the function $h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \upsilon(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is uniformly continuous on $[-\pi, \pi]^2 \times \Theta$;
2. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) / \upsilon(\boldsymbol{\lambda}) \in L_1([-\pi, \pi]^2) \cap L_2([-\pi, \pi]^2)$.

A7. The function $\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is twice differentiable on Θ and

1. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1([-\pi, \pi]^2) \bigcap L_2([-\pi, \pi]^2)$, for all i, j, and $\boldsymbol{\theta} \in \Theta$;
2. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_k([-\pi, \pi]^2)$, for all i, $\boldsymbol{\theta} \in \Theta$, and $k \geq 1$.

A8. The matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))$ and $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))$ with the elements defined by (A9*) and (A9*) are positive definite.

A9. The spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, the weight function $w(\boldsymbol{\lambda})$, and the function $\frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ are such that for all i and $\boldsymbol{\theta} \in \Theta$:

$$T \int_{[-\pi,\pi]^2} (EI_T^*(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda} \longrightarrow 0, \quad \text{as} \quad T \longrightarrow \infty.$$

8.3 Result

Through the following theorem the result of consistency is derived by the method of minimum contrast estimation,

Theorem 1. *Let Y_{t_1,t_2}, $(t_1, t_2) \in \mathbb{Z}^2$, be a stationary Gegenbauer random field which spectral density satisfies equation* (A9*). *If $\hat{U}_T(\boldsymbol{\theta})$ is the empirical contrast field defined by equation* (A9*), *then*

- *Y_{t_1,t_2} satisfies the conditions* **A1-A6** *in the Appendix;*
- *the minimum contrast estimator $\hat{\boldsymbol{\theta}}_T = (\hat{d}_1, \hat{d}_2) = arg \min_{\boldsymbol{\theta} \in \Theta} \hat{U}_T(\boldsymbol{\theta}) \in \Theta$ is a consistent estimator of the parameter vector $\boldsymbol{\theta}$. That is, there is a convergence in P_0 probability:*

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{P_0} \boldsymbol{\theta_0}, \quad T \longrightarrow \infty;$$

- $\hat{\sigma}_T^2 \xrightarrow{P_0} \sigma^2(\boldsymbol{\theta_0})$, $T \longrightarrow \infty$, *where the variance estimator* $\hat{\sigma}_T^2$ *is given by*

$$\hat{\sigma}_T^2 = \int_{[-\pi,\pi]^2} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}.$$

Similarly, the asymptotic normality is derived by the following theorem

Theorem 2. *If* Y_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$, *is a stationary Gegenbauer random field which spectral density satisfies equation* (A9*) *with* $(d_1,d_2) \in (0,1/4)^2$, *then*

- Y_{t_1,t_2} *satisfies the conditions* **A1-A9** *in the Appendix;*

- *the adjusted MCE defined by* (A9*) *is asymptotically normal. That is,*

$$T(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_2(0, \mathbf{S}^{-1}(\boldsymbol{\theta}_0) A(\boldsymbol{\theta}_0) \mathbf{S}^{-1}(\boldsymbol{\theta}_0)), \quad T \longrightarrow \infty,$$

where the entries of the matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))$ *and* $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))$ *are*

$$\begin{aligned} s_{ij}(\boldsymbol{\theta}) &= \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\, d\boldsymbol{\lambda} \\ &= \sigma^2(\boldsymbol{\theta}) \int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda}) \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) - \frac{1}{\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})} \frac{\partial}{\partial\theta_i} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) \right] d\boldsymbol{\lambda}, \quad (8.1) \end{aligned}$$

$$\begin{aligned} a_{ij}(\boldsymbol{\theta}) &= 8\pi^2 \int_{[-\pi,\pi]^2} f^2(\boldsymbol{\lambda},\boldsymbol{\theta}) w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \log\left(\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right) \frac{\partial}{\partial\theta_j} \log\left(\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\, d\boldsymbol{\lambda} \\ &= 8\pi^2 \sigma^2(\boldsymbol{\theta}) \int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial\theta_i} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\, d\boldsymbol{\lambda}. \quad (8.2) \end{aligned}$$

Part II

Apendix

Appendix A1

Spatial autoregressive functional plug-in prediction of ocean surface temperature

Ruiz-Medina, M. D. and Espejo, R. M. (2012).

Spatial autoregressive functional plug-in prediction of ocean surface temperature.

Stochastic Environmental Research and Risk Assessment, 26, 335–344.

DOI: 10.1007/s00477-012-0559-z.

Abstract

This paper addresses the problem of spatial functional extrapolation in the framework of spatial autoregressive Hilbertian processes of order one (SARH(1) processes) introduced in Ruiz-Medina (2011). Moment-based estimators of the operators involved in the state equation of these processes are computed by projection into a suitable orthogonal basis. Specifically, the eigenfunction basis diagonalizing the autocovariance operator is considered. An estimation algorithm is designed for the implementation of the resulting SARH(1)-plug-in projection extrapolator from temporal curves irregularly distributed in space. Its performance is illustrated with a real-data example, where the problem of spatial functional extrapolation of ocean surface temperature profiles is addressed. This problem is crucial in the assessment of climate change anomalies. The data are collected from the public oceanographic bio-optical database: *The World-wide Ocean Optics Database (WOOD).* Cross Validation (CV)

procedures are applied for the evaluation of the estimation results derived.

A1.1 Introduction

There exists a large number of contributions in the current literature in relation to the statistical analysis of temporal correlated functional random variables (see Bosq, 2000, 2010; Bosq and Blanke, 2007; Ferraty and Vieu, 2006; Ramsay and Silverman, 2005; Ruiz-Medina and Salmerón, 2010; Salmerón and Ruiz-Medina, 2009, among others). However, the field of *Spatial Functional Statistics* still requires further development. We mention a few recent papers regarding the statistical inference from spatial correlated functional random variables: Guillas and Lai (2010) propose a new class of spatial functional regression models based on bivariate splines, in terms of which the surface defining the explanatory random variables is approximated. Such an approximation allows the construction of least squares estimators of the regression function with or without a penalization term. In Baladandayuthapani *et al.* (2008), a fully Bayesian and Markov Chain Monte Carlo based approach is derived for the analysis of the spatial correlation between functional data arising from different spatial locations of biological structures called colonic crypts (i.e., it allows the analysis of crypt signaling phenomenon). In the nonparametric context, the statistical properties of kernel-based density estimators, formulated in the context of spatial functional random variables, are studied in Basse, Diop and Dabo-Niang (2008).

In a more applied framework, a new regionalization method for spatially dependent functional data, based on local variogram models, is derived in Romano, Balzanella and Verde (2010), for its application to environmental data. Nerini Monestiez and Manté, (2010) consider functional statistical tools for the analysis of spatial correlated Oceanology temporal curve data (see also Monestiez and Nerini, 2008). As spatial extrapolator, they propose a finite functional linear combination of the observed temporal curves, weighted by bounded integral operators. This

extrapolator is computed by minimizing the functional mean square error, which involves the spatial interaction between temporal curves.

Under some regularity conditions of the sample curves, an ordinary infinite-dimensional kriging system is established. In practice, numerical projection methods are applied, performing standard co-kriging in terms of the corresponding projections into a selected basis. A similar spatial functional extrapolation technique is considered in Giraldo, Delicado and Mateu (2010), where the weights defining the spatial extrapolator are temporal functions, instead of integral operators, minimizing the functional variance. Projection into a set of functional bases is also performed for implementation of the spatial functional predictor (see also Delicado *et al.* 2010). The main difference between the two last cited approaches, and the one presented here relies on the absence of the assumption on the existence of an underlying spatial functional Markovian model. Note that, in this paper, a spatial unilateral dynamics is assumed, according to the SARH(1) state equation, since we are motivated by environmental applications involving irregular sampling of spatial diffusion Markovian processes (see, for example, Tandeo, Ailliot, and Autret, 2011, in relation to sea surface temperature applications).

In Ruiz-Medina (2011,2012a), the construction of SARH(1) processes from the spatial functional sampling of two-parameter diffusion processes, in the class introduced by Nualart and Sanz-Solé (1979), is illustrated, with some specific examples, including the two-parameter Ornstein-Uhlenbeck process. Specifically, in Ruiz-Medina (2011), spatial autoregressive Hilbertian models of order p (SARH(p) models), and spatial moving average Hilbertian models of order q (SMAH(q) models) are introduced. Their structural properties are derived. In particular, conditions for the existence of a unique spatial stationary solution to the SARH(1) equation, admitting a Spatial Moving Average Hilbertian representation of infinite order (SMAH(∞) representation) are obtained. In Ruiz-Medina (2012a), the generation of SARH(1) processes from the tensorial product of Autoregressive Hilbertian processes of order one (ARH(1) processes)

is illustrated considering the tensorial product of functional observations of the one-parameter Ornstein-Uhlenbeck process. In this paper, the implementation of SARH(1) extrapolator is achieved from projection into the eigenfunction basis associated with the spectral decomposition of SARH(1) parameters (see also Ruiz-Medina and Salmerón, 2010, and Salmerón and Ruiz-Medina, 2009, in relation to the plug-in predictor for Autoregressive Hilbertian processes of order p, ARH(p)-plug-in predictor, computed from the spectral decomposition of the autocorrelation operators). A comparative study between this projection methodology and the compactly supported wavelet transform is also developed. Summarizing, it is showed that the compactly supported wavelet transform outperforms the projection into the eigenfunction system of the SARH(1)parameters in presence of strong correlations, caused, for example, by high-concentration-level of spatial stations. While the projection into the eigenfunction system of the SARH(1) parameters leads to a regularization of the associated estimation integral equation system, removing the ill-posed nature of this problem. Additionally, border effects are suitably processed with the interval discrete wavelet transform, without lengthening the support of the analyzed functional data. As we show in the real-data example analyzed in Section A1.4, the projection methodology applied in this paper, based on the eigenfunction basis of the autocovariance operator, outperforms the projection methodologies presented in Ruiz-Medina (2012a), in relation to dimension reduction to attach a prescribed level of explaining variability, in absence of strong correlations.

In practice, several additional problems arise in the implementation of the above-referred SARH(1) estimation methodology from spatial functional data. This fact motivates the design of the estimation algorithm proposed for SARH(1) extrapolation, from irregularly spaced functional data, in Section A1.3.1. Note that the spatial Hilbert-valued process context chosen allows the global prediction of curves or surfaces in space, reproducing information on local variability features that is lost in the classical discrete approaches, based, for example, on kriging inter-

polator or Generalized Linear Mixed Models (GLMM models). As commented, the connection of SARH(1) process values with data resulting from the spatial functional sampling of diffusion processes, in the class introduced in Nualart and Sanz-Solé (1979), makes SARH(1)-based extrapolation specially attractive for addressing environmental applications involving this class of spatial Markovian processes (see Ruiz-Medina, 2011, 2012a, and Tandeo, Ailliot, and Autret, 2011).

The outline of the paper is the following. Section A1.2 provides the basic definitions and previous results on SARH(1) processes. Section A1.3 describes the SARH(1) projection estimation methodology. The estimation algorithm proposed for the implementation of SARH(1) plug-in predictor in the case of irregularly spaced functional data is formulated in Section A1.3.1, considering the eigenfunction system of the autocovariance operator. In Section A1.4, mean annual ocean surface daily temperature profiles, collected at irregularly spaced weather stations of the south coast of the U.S., in the latitude-longitude interval $[20, 50] \times [-89, -58]$, are analyzed through 12 years, corresponding to the period 1997-2008. The performance of the proposed spatial functional estimation algorithm is evaluated by applying 10-fold Cross Validation (10-fold CV) procedures. Concluding remarks on the quality of the analysis performed, its applicability, as well as on further development needed, for improving the estimation methodology proposed, are conducted in Section A1.5.

A1.2 Statistical modeling and preliminary results

The basic definitions and results related to SARH(1) processes are now summarized (see Ruiz-Medina, 2011). Let $\mathcal{H}$ be a separable Hilbert space with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and the norm $\| \cdot \|_{\mathcal{H}}$. Given a basic probability space $(\Omega, \mathcal{A}, P)$, denote by $L^2_{\mathcal{H}}(\Omega, \mathcal{A}, P)$ the Hilbert space of classes of random variables with values in $\mathcal{H}$ and with $E\|X\|^2_{\mathcal{H}} < \infty$.

Definition 1. *A spatial functional process* $\mathbf{Y}_{SARH} = \{Y_{ij},\ (i,j) \in \mathbb{Z}^2\}$, *with values in a separa-*

ble Hilbert space $\mathcal{H}$, is said to be a unilateral SARH(1) process if it is stationary and it satisfies the following equation

$$Y_{i,j} = R + L_1(Y_{i-1,j}) + L_2(Y_{i,j-1}) + L_3(Y_{i-1,j-1}) + \epsilon_{i,j}, \tag{A1.1}$$

where $R \in \mathcal{H}$ and, for $k = 1, 2, 3$, $L_k \in \mathcal{L}(\mathcal{H})$, the space of linear bounded operators. The spatial functional innovation process also satisfies $E\|\epsilon_{i,j}\|^2_{\mathcal{H}} = \sigma^2$, and $E[\epsilon_{i,j} \otimes \epsilon_{i,j}] = C_{\epsilon_{0,0}}$, that is, it has finite functional variance, and both; its functional variance and autocovariance operator do not depend on the spatial location (i, j), for every $(i, j) \in \mathbb{Z}^2$.

Remark 1. *Note that, in the above definition, the order one of the SARH(1) process family introduced refers to the fact that the functional value $Y_{i,j}$ interacts in space with the values $Y_{i-1,j}$, $Y_{i,j-1}$ and $Y_{i-1,j-1}$, respectively corresponding to one spatial lag at coordinate i, j, or at both spatial coordinates i and j. Since an unilateral dynamics is considered, only negative spatial lags are involved in such a definition.*

SMAH(∞) representation

Let us now introduce the conditions that ensure the existence and uniqueness of a spatial stationary functional solution to the SARH(1) equation, admitting a SMAH(∞) representation (see Ruiz-Medina, 2011).

A.1 For $i = 1, 2, 3$, operator $L_i \in \mathcal{L}(\mathcal{H})$ is assumed to admit a spectral decomposition in terms of the eigenvalue sequence $\{\lambda_k(L_i),\ k \in \mathbb{N}\}$, and the biorthogonal systems of left, $\{\psi_k,\ k \in \mathbb{N}\}$, and right, $\{\phi_k,\ k \in \mathbb{N}\}$, eigenvectors satisfying

$$\begin{aligned} L_i(\psi_k) &= \lambda_k(L_i)\psi_k \\ L_i^*(\phi_k) &= \lambda_k(L_i)\phi_k \\ \langle \phi_k, \psi_l \rangle_{\mathcal{H}} &= \delta_{k,l}, \quad k, l \in \mathbb{N}, \end{aligned}$$

where L_i^* denotes the adjoint of L_i, and $\delta_{k,l}$ the Kronecker-delta function. Under **A.1**, for $i = 1, 2, 3$, L_i admits the following spectral-kernel-based representation

$$L_i(g)(f) = \sum_{k\in\mathbb{N}} \lambda_k(L_i)\psi_k(f)\phi_k(g), \quad \forall f, g \in \mathcal{H}. \tag{A1.2}$$

Remark 2. *As commented in Ruiz-Medina (2012a), examples of operators satisfying Assumption **A.1** can be found in the context of non-symmetric compact operators, which often arise in the definition of stochastic differential systems modeling the relationship between two environmental phenomena. Hydrological processes involved in dynamic flow equations constitute an interesting example, where aquifer transmissivity, porosity, storage coefficients, positioning of withdrawals are estimated from related magnitudes like piezometric data.*

Proposition 1. *(see Ruiz-Medina, 2011) Assume that **A.1** holds. For each $k \in \mathbb{N}$, none of the roots of $\Phi(z_1, z_2) = 0 = 1 - \lambda_k(L_1)z_1 - \lambda_k(L_2)z_2 - \lambda_k(L_3)z_1z_2$ lie within the closed unit polydisc ($|z_1| \leq 1, |z_2| \leq 1$) if and only if, for each $k \in \mathbb{N}$,*

(i) $|\lambda_k(L_i)| < 1$, *for* $i = 1, 2, 3$,

(ii) $\left(1 + [\lambda_k(L_1)]^2 - [\lambda_k(L_2)]^2 - [\lambda_k(L_3)]^2\right)^2 - 4(\lambda_k(L_1) + \lambda_k(L_2)\lambda_k(L_3))^2 > 0$,

(iii) $1 - [\lambda_k(L_2)]^2 > |\lambda_k(L_1) + \lambda_k(L_2)\lambda_k(L_3)|$.

Under (i)-(iii), the solution to equation (A1.1) is stationary.

The conditions needed for the SMAH(∞) representation of process $\mathbf{Y}_{SARH}$ in equation (A1.1) are established in the following result.

Proposition 2. *(see Ruiz-Medina, 2011) Let $\mathbf{Y}_{SARH}$ be a SARH(1) process, as given in Definition 1, satisfying **A.1**. Assume that the conditions considered in Proposition 1 hold. Then, equation (A1.1) admits a unique stationary solution given by*

$$Y_{i,j} = R + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} L_1^k L_2^l L_3^r(\epsilon_{i-k-r,j-l-r}), \tag{A1.3}$$

where L_i, $i = 1, 2, 3$, are defined in equation (A1.1).

A1.3 SARH(1) functional prediction

This section introduces the elements involved in the computation of SARH(1)-plug-in projection extrapolator from spatial functional data on a regular grid. The irregularly-spaced data case will be addressed in the estimation algorithm proposed in the next section.

The spatial functional second-order structure of SARH(1) processes is defined in terms of the following covariance operators, which collect the spatial interaction between functional data obeying the functional state equation (A1.1): for $Z_{i,j} = Y_{i,j} - R$, $(i, j) \in \mathbb{Z}^2$,

$$\begin{aligned} R_{Z_{i,j}Z_{i,j}} &= R_{0,0} = E\left[Z_{i,j} \otimes Z_{i,j}\right] \\ R_{Z_{i+1,j}Z_{i,j}} &= R_{1,0} = E\left[Z_{i+1,j} \otimes Z_{i,j}\right] \\ R_{Z_{i,j+1}Z_{i,j}} &= R_{0,1} = E\left[Z_{i,j+1} \otimes Z_{i,j}\right] \\ R_{Z_{i+1,j+1}Z_{i,j}} &= R_{1,1} = E\left[Z_{i+1,j+1} \otimes Z_{i,j}\right]. \end{aligned} \tag{A1.4}$$

Remark 3. *Note that the tensorial product $\otimes$ of two functions in $\mathcal{H}$ defines a Hilbert-Schmidt operator on $\mathcal{H}$, given by*

$$f \otimes g(h) = \langle f, h^* \rangle_{\mathcal{H}}\, g, \quad f, g \in \mathcal{H},\ h \in \mathcal{H}^* \equiv \mathcal{H},$$

where h^* denotes the dual element of h defined from Riesz Representation Theorem. From now on, $\mathcal{S}$ will denote the space of Hilbert-Schmidt operators on $\mathcal{H}$ identified with the self-tensorial product of $\mathcal{H}$, i.e., $\mathcal{S} \equiv \mathcal{H} \otimes \mathcal{H}$. Also, as before, by $\mathcal{L}(\mathcal{H})$ we will denote the space of bounded linear operators on $\mathcal{H}$.

The following empirical covariance operators

$$\hat{R}_{0,0} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} Z_{i,j} \otimes Z_{i,j}$$
$$\hat{R}_{1,0} = \frac{1}{(M-1)N}\sum_{i=1}^{M-1}\sum_{j=1}^{N} Z_{i+1,j} \otimes Z_{i,j}$$
$$\hat{R}_{0,1} = \frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N-1} Z_{i,j+1} \otimes Z_{i,j}$$
$$\hat{R}_{1,1} = \frac{1}{(M-1)(N-1)}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1} Z_{i+1,j+1} \otimes Z_{i,j}, \tag{A1.5}$$

provide moment-based estimators of the spatial autocovariance and cross-covariance operators (A1.4), from the functional observations

$$\{Z_{i,j} = Y_{i,j} - R,\ (i,j) \in \{1,\ldots,M\} \times \{1,\ldots,N\}\},$$

with R and $\mathbf{Y}_{SARH} = \{Y_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ given in equation (A1.1). From this equation, the autocorrelation operators L_i, $i = 1,2,3$, satisfy the functional linear equation system

$$\begin{aligned} R_{1,0} &= L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1}, \\ R_{0,1} &= L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0} \\ R_{1,1} &= L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0}. \end{aligned} \tag{A1.6}$$

Replacing in (A1.6) the autocovariance and cross-covariance operators by their empirical versions in (A1.5), the respective estimators $\widehat{L}_i$, $i = 1,2,3$, of L_i, $i = 1,2,3$, are obtained as the solution to the corresponding functional linear equation system. Thus, formally, $\widehat{L}_i$, $i = 1,2,3$, are defined by

$$\begin{bmatrix} \widehat{L}_1 \\ \widehat{L}_2 \\ \widehat{L}_3 \end{bmatrix} = \begin{bmatrix} \hat{R}_{0,0} & \hat{R}_{1,1} & \hat{R}_{0,1} \\ \hat{R}_{1,1} & \hat{R}_{0,0} & \hat{R}_{1,0} \\ \hat{R}_{0,1} & \hat{R}_{1,0} & \hat{R}_{0,0} \end{bmatrix}^{-1} \begin{bmatrix} \hat{R}_{1,0} \\ \hat{R}_{0,1} \\ \hat{R}_{1,1} \end{bmatrix}. \tag{A1.7}$$

The SARH(1) extrapolator is then approximated in terms of the plug-in estimator:

$$\begin{aligned} E\left[Z_{i,j} \mid Z_{k,l},\ k<i,\ \text{or } l<j,\ \text{or } k<i \text{ and } l<j\right] &= L_1 Z_{i-1,j} + L_2 Z_{i,j-1} + L_3 Z_{i-1,j-1} \\ &\simeq \widehat{L_1} Z_{i-1,j} + \widehat{L_2} Z_{i,j-1} + \widehat{L_3} Z_{i-1,j-1}. \end{aligned} \tag{A1.8}$$

In the computation of an explicit vector-valued functional solution to equation system (A1.7), numerical projection methods are applied in terms of a suitable orthonormal basis of $\mathcal{H}$. Specifically, given an orthonormal system $\{\phi_k,\ k \in \mathbb{N}\}$ of $\mathcal{H}$, we obtain, by projection, the following infinite-dimensional system, for $k, l, m \in \mathbb{N}$,

$$\begin{aligned} \lambda_{k,m}(\widehat{R}_{1,0}) &= \sum_{l\geq 0} \lambda_{k,l}(L_1)\lambda_{l,m}(\hat{R}_{0,0}) + \lambda_{k,l}(L_2)\lambda_{l,m}(\hat{R}_{1,1}) + \lambda_{k,l}(L_3)\lambda_{l,m}(\hat{R}_{0,1}) \\ \lambda_{k,m}(\widehat{R}_{0,1}) &= \sum_{l\geq 0} \lambda_{k,l}(L_1)\lambda_{l,m}(\hat{R}_{1,1}) + \lambda_{k,l}(L_2)\lambda_{l,m}(\hat{R}_{0,0}) + \lambda_{k,l}(L_3)\lambda_{l,m}(\hat{R}_{1,0}) \\ \lambda_{k,m}(\widehat{R}_{1,1}) &= \sum_{l\geq 0} \lambda_{k,l}(L_1)\lambda_{l,m}(\hat{R}_{0,1}) + \lambda_{k,l}(L_2)\lambda_{l,m}(\hat{R}_{1,0}) + \lambda_{k,l}(L_3)\lambda_{l,m}(\hat{R}_{0,0}), \end{aligned} \tag{A1.9}$$

where, for a Hilbert-Schmidt operator K on $\mathcal{H}$,

$$\lambda_{l,m}(K) = K(\phi_l)(\phi_m) = \langle K(\phi_l), \phi_m \rangle_{\mathcal{H}}\,.$$

Note that, in equation (A1.9), these coordinate matrices are evaluated for $K = L_i$, $i = 1, 2, 3$, and $K = \hat{R}_{i,j}$, $i, j = 0, 1$.

After truncation, equation (A1.9) can be rewritten in matrix form as follows:

$$\begin{bmatrix} \widehat{\Lambda}(L_1) \\ \widehat{\Lambda}(L_2) \\ \widehat{\Lambda}(L_3) \end{bmatrix} = \begin{bmatrix} \Lambda(\hat{R}_{0,0}) & \Lambda(\hat{R}_{1,1}) & \Lambda(\hat{R}_{0,1}) \\ \Lambda(\hat{R}_{1,1}) & \Lambda(\hat{R}_{0,0}) & \Lambda(\hat{R}_{1,0}) \\ \Lambda(\hat{R}_{0,1}) & \Lambda(\hat{R}_{1,0}) & \Lambda(\hat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \Lambda(\hat{R}_{1,0}) \\ \Lambda(\hat{R}_{0,1}) \\ \Lambda(\hat{R}_{1,1}) \end{bmatrix}, \tag{A1.10}$$

where, for a Hilbert-Schmidt operator K on $\mathcal{H}$, $\Lambda(K)$ represents the matrix whose entries are the coordinates of operator K with respect to the tensorial product basis

$$\{\phi_l \otimes \phi_m,\ l, m = 1, \ldots L\}\,,$$

i.e., the matrix with entry (i,j), given by $\lambda_{i,j}(K)$, for $i,j=1,\dots L$. Also, $\widehat{\Lambda}(L_k)$ denotes the estimate of matrix $\Lambda(L_k)$, for $k=1,2,3$.

Note that, in the particular case where the orthonormal system $\{\phi_j,\ j\in\mathbb{N}\}$ of $\mathcal{H}$ coincides with the eigenfunction basis associated with the covariance operator $R_{0,0}$, the corresponding projection of equation system (A1.6) into such a basis leads to a sparse infinite-dimensional system. Similarly, when the empirical version $\{\widehat{\phi}_j,\ j\in\mathbb{N}\}$ of $\{\phi_j,\ j\in\mathbb{N}\}$ is considered, given by the eigenfunction system of $\hat{R}_{0,0}$, the projected equation system (A1.10) has block-matrix diagonal constituted by diagonal matrices. That is, in equation (A1.10), matrix $\Lambda(\hat{R}_{0,0})$ is diagonal with entries given by the eigenvalues $\widehat{\lambda}_j$, $j\geq 1$, of $\hat{R}_{0,0}$, respectively associated with the eigenvectors $\widehat{\phi}_j$, $j\geq 1$, that satisfy

$$\hat{R}_{0,0}\widehat{\phi}_j=\widehat{\lambda}_j\widehat{\phi}_j,\quad j\geq 1.$$

The empirical autocovariance eigenfunction system will be considered in Section A1.4, in the implementation of the SARH(1)-plug-in extrapolator in the real-data example considered. The main advantage of this projection methodology lies in its dimension reduction. This dimension reduction is higher than, for example, the one achieved with the projection methodologies applied in Ruiz-Medina (2012a), where the eigenfunction system associated with operators L_k, $k=1,2,3$, and the compactly supported wavelet transform are respectively considered. Thus, a lower number L of terms than in the other two referred projection methodologies (based on the autocorrelation operator eigenfunction system and compactly supported wavelet transform) is needed to attach a prescribed level of explained variability.

Finally, we remark the fact that the consistency of the empirical autocovariance operator, as well as of its eigenvalues and eigenvectors will allow us the investigation of the asymptotic properties of the derived SARH(1)-plug-in extrapolator (see, for example, Bosq, 2000; Bosq and Blanke, 2007, and Horváth and Reeder, 2012). While, for the SARH(1) parameter eigenfunction system, or the compactly supported wavelet transform, consistency results are not still available

in the SARH(1) context. Thus, from a theoretical point of view, asymptotic inference can easily be addressed by projection into the autocovariance eigenfunction system.

A1.3.1 Implementation of SARH(1)-plug-in extrapolator from irregularly spaced functional data

In this section, an estimation algorithm is proposed for the implementation of the SARH(1)-plug-in projection extrapolator from irregularly spaced functional data. The main steps are the following:

Step 1 A number $M \times N$ of nodes or areal data is fixed.

Step 2 To distribute in a suitable way the number of areal data fixed, an initial computational grid, with a lower number of nodes than the number of areal data established in the previous step, is considered.

Step 3 In the boxes of the computational grid where a large number of spatial stations is localized, a subgrid is considered, for getting an homogeneous level of integration/aggregation per block, increasing the number of rows and/or columns, until the number of nodes of the global grid reaches the number of areal data fixed in Step 1.

Step 4 Functional data are spatially averaged over each box of the grid fitted in Step 3, and the spatially averaged functional values are assigned to the respective nodes located at the lower left corners of the non-empty boxes.

Step 5 Interpolation is performed over the empty nodes.

Step 6 The empirical covariance operators (A1.5) are computed from the spatial averaged functional data defined in the previous steps according to the lattice configuration fitted.

Step 7 The orthogonal eigenvectors of the empirical autocovariance operator are then obtained.

Step 8 The truncation order L is selected (e.g. by cross validation).

Step 9 The computation of the finite-dimensional equation system (A1.10) is achieved by applying numerical inversion methods.

Step 10 The SARH(1)-plug-in extrapolator is derived from the previous step considering equation (A1.8).

Step 11 Cross validation procedures are applied for the evaluation of the performance of the estimation methodology implemented in Steps 1-10.

A1.4 Spatial functional prediction of ocean surface temperature

The ocean temperature signal provides a baseline temperature reference that can be used, for example, to investigate urban heat island effects, and look for anomalies in the weather station record. In particular, urban heat island effect decreases the differences between ocean and air surface temperatures at the end of the day and during the night, reducing heat fluxes from the ocean into the air. Hence, coastal ocean surface temperature increases. This fact can be appreciated by comparing the slope functions (derivatives) associated with coastal and ocean daily temperature curves (see Figure A1.3 below). In our case, the temperature profiles collected at the weather stations in the south coast of the U.S. reflect an increase of coastal ocean surface temperature, with respect to the baseline temperature reference provided by the ocean surface temperature signal far from the coast (Hawaii ocean). This fact motivates the derivation of suitable spatial extrapolation techniques for prediction of daily ocean surface temperature curves. In particular, in this section, we have concentrated in the problem of spatial functional extrapolation from the observation of temperature curves provided by irregularly spatial distributed weather station located at the south coast of the U.S, in the latitude-longitude interval $[20, 50] \times [-89, -58]$), available in the public oceanographic bio-optical database, *The World-*

wide Ocean Optics Database (WOOD). Mean annual daily temperature profiles corresponding to 12 years in the period 1997-2008 are computed, to construct our functional data sets in this real-data example. The data size considered, number of observations per each average daily temperature profile, has been 200.

In the finite-dimensional approximation of the autocorrelation operator estimators, the truncation order $L = 15$ is fixed for all the years considered, since a stabilization of the $L^{\infty}-$norm of the $10-$fold CV mean absolute error curves is obtained in most of the years studied. The consideration of a common finite dimension Hilbert space, corresponding to $L = 15$, allows us to compare the effect, on the performance of the estimation methodology proposed, of the spatial distribution of weather stations, as well as of the spatial averaging regions selected for the construction of a lattice configuration.

Validation results are displayed in Table A1.1, and Figure A1.1, after applying 10-fold CV, which leads to a suitable reduction of CV error variability. Specifically, the $L^{\infty}-$norms of the 10-fold CV mean absolute error curves are displayed in Table A1.1. The behavior of these curves is represented in Figure A1.1, for certain nodes of the validation sample. Edge effects have been removed by considering the empirical orthogonal eigenfunction bases associated with the computation of the empirical autocovariance operator $\hat{R}_{0,0}$ from temperature profiles with 256 observations. The results displayed illustrate the fact that the aggregation and interpolation levels associated with the 6×25 spatial grids fitted for every year, in the period $1997 - 2008$, are suitable for most of these years (see Figure A1.2). Additionally, since in the subtropical region studied, the temperature profiles display slow variation through space of their global scale and local pattern properties, the loose of information, due to the averaging of temperature profiles in a small region, is minimum.

In the implementation of the estimation algorithm described in Section A1.3.1, the number of areal data fixed in Step 1 has been 150, since a 6×25 spatial grid is fitted at every year

studied. In most of the years, weather stations are more concentrated in latitude (x-axis of graphs in Figure A1.2) than in longitude (y-axis of graphs in Figure A1.2). Therefore, from a 6×6 initial regular computational grid, we construct a final 6×25 spatial grid, splitting regions with a large number of weather stations. This procedure is repeated for each year, subdividing different boxes of the initial computational grid, according to the spatial distribution of weather stations. The best performance corresponds to the years 1997 (65 weather stations), 1998 (73 weather stations), 2000 (65 weather stations), 2001 (93 weather stations) and 2007 (42 weather stations), followed by the years 2002 (114 weather stations) and 2003 (101 weather stations). The worst performance is displayed by the year 2008 (15 weather stations). The above results illustrate the fact that the best performance of the proposed estimation algorithm is achieved at the years where weather stations are sparsely distributed in space. The absence of high concentration of weather stations decreases the number of temperature profiles aggregated at each box of the final 6×25 computational grid, reducing variance and increasing the level of explaining variability, i.e., the L^{∞}-norms of the $10-$fold CV mean absolute error curves decrease. This is the case, for example, of year 2007, where the number of weather stations is not too large. However, their sparse spatial distribution leads to a low aggregation level of temperature profiles at each box of the final 6×25 spatial grid, decreasing variance of areal data, and increasing velocity decay of empirical eigenvalues, i.e., a higher level of explained variability is reached with $L = 15$. In addition, the sparse spatial distribution of weather stations in the boxes of the final 6×25 grid fitted to this year, and the slow variation of temperature profiles through space also contribute to the low values of the $L^{\infty}-$norms of the $10-$fold CV mean absolute error curves at the interpolated empty nodes. On the contrary, when the aggregation level per box increases, due to the high concentration of weather station locations, the variance of the resulting areal data increases, and a slower decay of the corresponding empirical eigenvalues holds. Thus, a lower level of explained variability, with $L = 15$ truncation order, is reached (e.g. years 1999

Table A1.1: *L^∞-norms of the $10-$fold CV mean absolute error curves with $L = 15$*

Nodes	(2,2)	(3,2)	(2,3)	(3,3)	(2,4)	(3,4)	(2,5)	(3,5)
1997	0.0183	0.0129	0.0155	0.0147	0.0146	0.0156	0.0151	0.0153
1998	0.0634	0.0458	0.0432	0.0294	0.0602	0.0670	0.0619	0.0544
1999	0.1929	0.2198	0.1662	0.2360	0.1752	0.2653	0.2346	0.2984
2000	0.0397	0.0467	0.0267	0.0300	0.0281	0.0257	0.0391	0.0316
2001	0.0082	0.0286	0.0134	0.0187	0.0389	0.0484	0.0481	0.0543
2002	0.0699	0.0642	0.1753	0.0601	0.0415	0.0624	0.0578	0.0299
2003	0.0183	0.0129	0.0155	0.0147	0.0146	0.0156	0.0151	0.0153
2004	0.0634	0.0458	0.0432	0.0294	0.0602	0.0670	0.0619	0.0544
2005	0.1929	0.2198	0.1662	0.2360	0.1752	0.2653	0.2346	0.2984
2006	0.0397	0.0467	0.0267	0.0300	0.0281	0.0257	0.0391	0.0316
2007	0.0082	0.0286	0.0134	0.0187	0.0389	0.0484	0.0481	0.0543
2008	0.0699	0.0642	0.1753	0.0601	0.0415	0.0624	0.0578	0.0299

and 2006). Finally, we point out that, in the year 2008, a small number of weather stations is available, with a clear directional diagonal distribution, increasing the interpolation error, which is reflected in the higher values of the L^∞-norms of the $10-$fold CV mean absolute error curves.

Through the 12 years studied, we finally compare the slope functions of the U.S. coastal ocean surface temperature curves and of the Hawaii ocean temperature curves (the baseline temperature reference), both extrapolated to a 6×25 computational grid from the daily temperature profiles observed at weather stations of the south coast of the U.S. and of the Hawaii island. Figure A1.3 below shows the mean, over the 6×25 nodes of the computational grid, of the difference: coastal temperature slope function minus ocean temperature slope function. It can be appreciated along the years studied a slightly increasing of coastal temperature curve values at the end of the day, i.e., a slower decreasing of these coastal temperatures at the end of the day, that leads to bigger mean slope function differences, through the 12 years studied, between the U.S. south coastal and the Hawaii ocean temperature curves, specially in the last years (see Figure A1.3).

Figure A1.1: Years 1997-2008. Mean absolute error curves on the spatial nodes (2,3) and (3,3) of the validation sample, after applying 10−fold CV, with $L = 15$

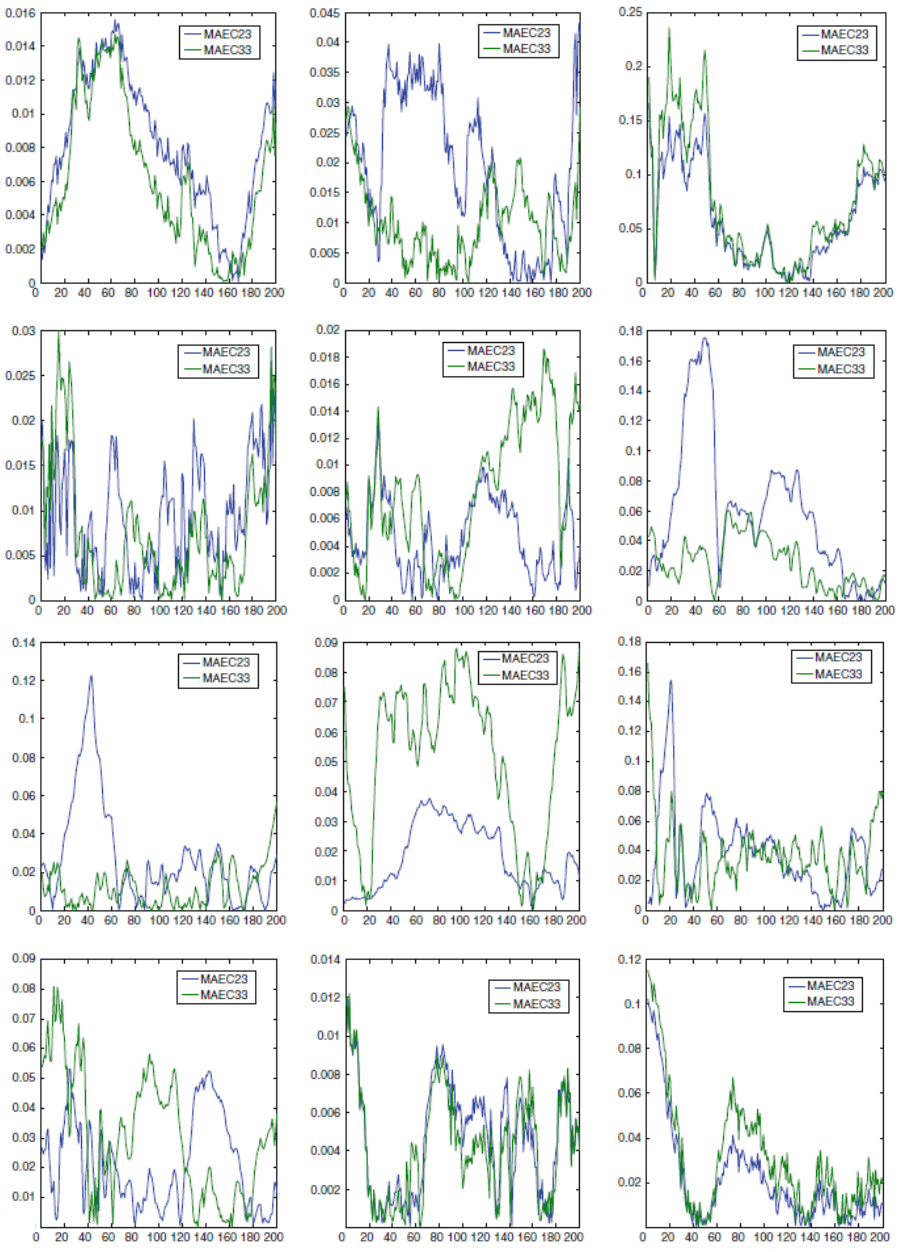


Figure A1.2: Spatial distribution of weather stations: from year 1997 (top-left) to year 2008 (right), going from left-hand-side to right-hand-side, and from top to bottom

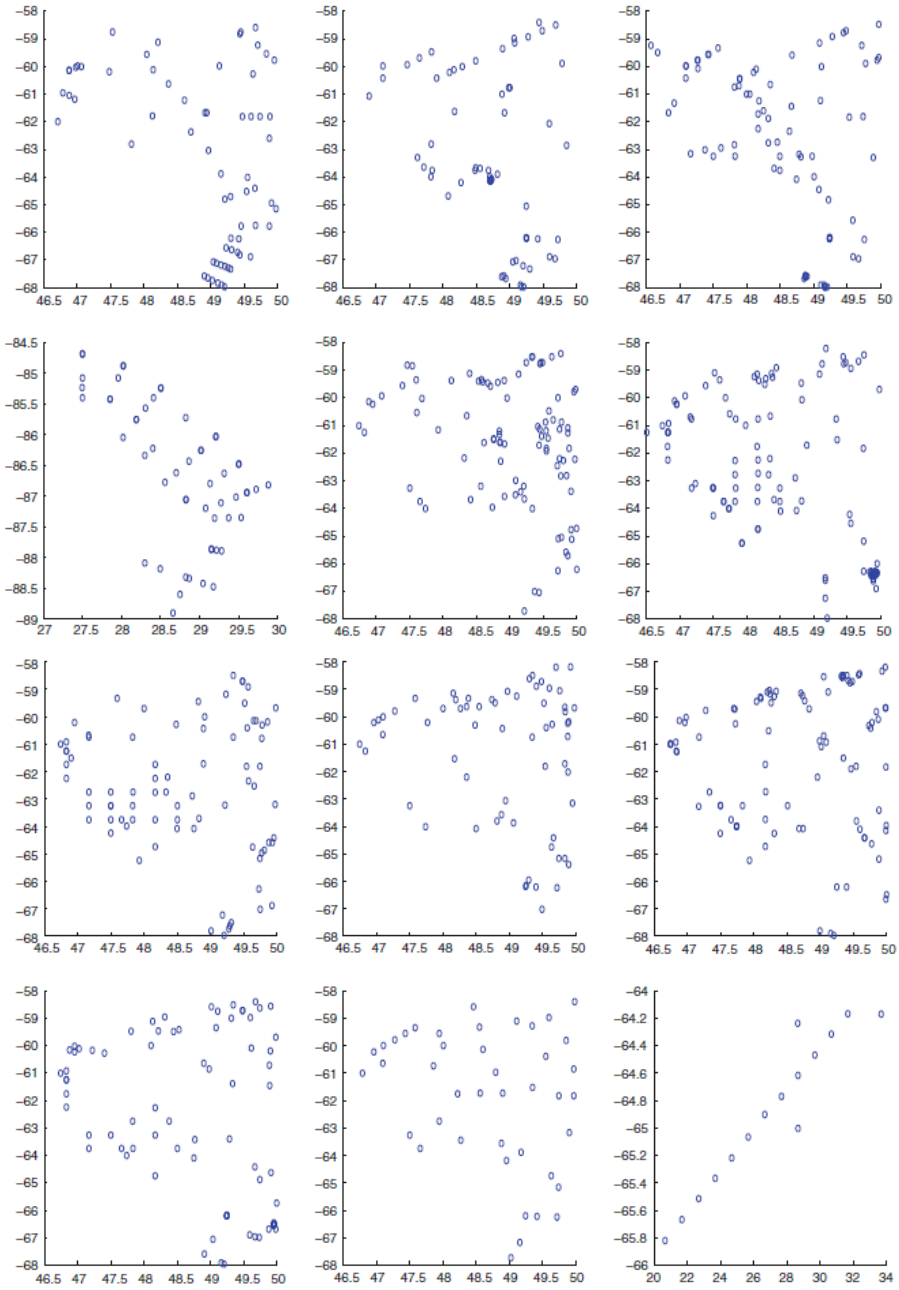
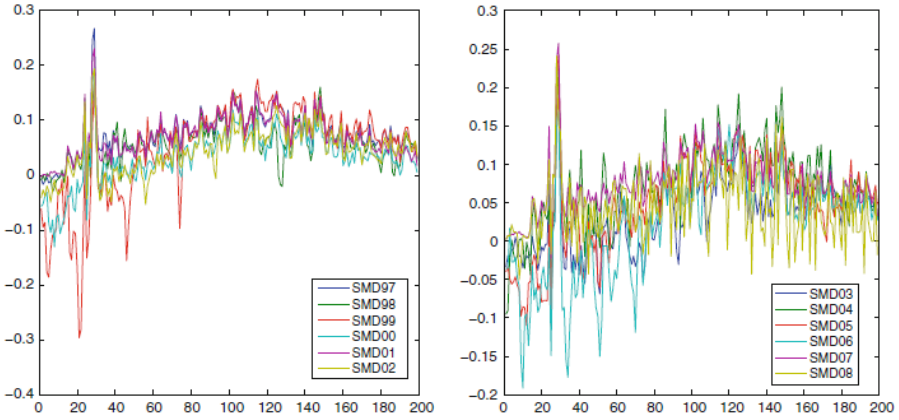


Figure A1.3: Mean over the 6×25 nodes of the spatial computational grid considered of the difference between slope functions associated with extrapolated coastal and ocean surface daily temperature curves. Years 1997-2002 at the left, and 2003-2008 at the right, where notation SMD means, Slope Mean Difference



A1.5 Evaluation of the proposed spatial functional estimation algorithm

This paper provides a methodological approach for the implementation of SARH(1) plug-in extrapolator from irregularly spaced functional data (see Section A1.3.1). The projection methodology applied is based on the spectral decomposition of the autocovariance operator. We now refer to the key points that must be taken into account in the evaluation of the proposed spatial functional estimation methodology, with reference to other related approaches.

The number of areal data fixed in Step 1 of the estimation algorithm, described in Section A1.3.1, must allow an equilibrium between the aggregation level per block and the interpolation level. The averaging regions selected, in the construction of a lattice configuration of areal data, must minimize the loose of information. This selection problem is not always solvable. It depends on the distribution of spatial stations and on the variability displayed by the functional data analyzed at the temporal micro- and macro-scale levels. In particular, from the results

derived in the previous section, it can be appreciated that the effect of the averaging region is minimized by the slow variation of the local and global temporal patterns of the curve data analyzed. This is, in general, the case of temperature profiles in tropical and subtropical regions.

When weather stations display high spatial concentration, the compactly supported wavelet transform is more suitable than the eigenfunction system-based approach, since it weakens strong spatial correlations (see Ruiz-Medina, 2012a, and Ruiz-Medina and Espejo, 2011). In the analysis performed in Section A1.4, border effects have been eliminated by extending the temporal support of the functions in the space $\mathcal{H}$, where the SARH(1) process takes its values. However, the compactly supported wavelet transform eliminates border effects without lengthening this temporal support.

In the case where the eigenvalues of

$$\begin{bmatrix} \hat{R}_{0,0} & \hat{R}_{1,1} & \hat{R}_{0,1} \\ \hat{R}_{1,1} & \hat{R}_{0,0} & \hat{R}_{1,0} \\ \hat{R}_{0,1} & \hat{R}_{1,0} & \hat{R}_{0,0} \end{bmatrix},$$

display a very fast decay to zero, the projection methodology, based on the infinite-dimensional SARH(1) parameter eigenfunction system (see Ruiz-Medina, 2012a) outperforms the autocovariance eigenfunction-system-based projection methodology, since it provides a regularization of the ill-posed parameter estimation problem. To reach a prescribed level of explaining variability, in absence of high concentration of spatial stations, the autocovariance eigenfunction system provides an optimal SARH(1)-plug-in projection extrapolator, in the sense of the dimension reduction achieved.

Further developments are needed in the optimal design of Steps 1-5 of the proposed estimation algorithm in Section A1.3.1 to minimizing the effect of the averaging region, in the construction of an areal data lattice configuration, when the functional data analyzed display high spatial variation in their local and/or global temporal patterns through space.

Appendix A2

Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming

Ruiz-Medina, M. D. and Espejo, R. M. (2013a).

Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming.

International Journal of Applied Earth Observation and Geoinformation, 22, (1):27–39.

DOI 10.1016/j.jag.2012.01.021

Abstract

The aim of this paper is to derive spatiotemporal extrapolation maps of ocean surface temperature to investigate two global warming effects: On the one hand, the reduction of daily heat fluxes from the sea into the air at the end of the day and during the night, in tropical regions. On the other hand, the strengthening of ocean current flows, due to the increase of ocean surface minimum daily temperature differences between two connected ocean regions. These maps are constructed from the spatial functional time series framework. Specifically, the spatial functional extrapolation of ocean surface temperature from Hawaii ocean to the Gulf of México reflects an increase of Hawaii ocean surface temperature in the last 15 years, caused by the reduction of daily heat fluxes

from the sea into the air. Furthermore, for the two connected regions of Indian Ocean, and the eastern coast of Australia, the spatial functional extrapolation results derived show more pronounced differences between ocean surface minimum daily temperatures in the year 2003 than in the years 1995-1997. Thus, a strengthening of the flow of the East Australian Current is appreciated.

A2.1 Introduction

It is well-known that advantages and limitations are present in the application of pure data synthesis and model/data integration approaches. For example, pure data synthesis methods are more sensitive (i.e. less robust) to the characteristics of measurement devices, or to blending procedures for data sets from different sensors, etc. Model/data integration can be performed following different approaches. In particular, physical ocean models can provide ocean surface temperature estimates that are consistent with the dynamical and thermodynamical constrains. However, a model error term reflecting unresolved physics is needed (also in relation to sub-griding parameterizations). Thus, physical modeling must be combined with stochastic modeling and statistical estimation.

In this paper, physical and stochastic modeling are integrated, and data-driven methodologies are applied to obtain a compromise between pure data synthesis and physical stochastic model fitting. Specifically, spatial functional time series modeling (see Ruiz-Medina, 2011) is applied for stochastic spatiotemporal extrapolation of ocean surface temperature. Under suitable conditions, this framework provides a statistical approximation of the usual physical equations governing sea surface temperature dynamics. In particular, recent studies of mid-latitude sea surface temperature dynamics are developed in terms of the thermal feedback between a two level atmospheric model and a dynamically passive slab ocean (see, for example, Nilsson, 2001). In this context, within the atmosphere coupled to a slab ocean modeling, the long-term evolution of sea surface temperature field (driven by air-sea transfer) is governed by an advective-diffusive

equation, which can be applied to a broad class of atmospheric models. This fact constitutes our main motivation to consider the selected spatial functional time series framework intimately connected with the theory of diffusion equations (see Nualart and Sanz, 1979; Leonenko and Ruiz-Medina, 2006; Ruiz-Medina, 2011).

The main drawback of the applied framework is that functional data must be located on a regular spatial grid. Spatial averaging over the boxes of a computational grid is then performed to obtain a spatial lattice configuration. The grid box size is adjusted to the spatial distribution of weather stations. Spatial Autoregressive Hilbertian (SARH) model fitting (see Ruiz-Medina, 2012a) is then performed for spatial functional extrapolation of ocean surface temperature profiles from the coastal weather stations to the deep ocean. The separable Hilbert-valued context is crucial in the introduction of suitable orthogonal function bases allowing the implementation of numerical projection methods for dimension reduction in the high-dimensional data context. For example, in the implementation of the SARH(1) extrapolation methodology, the projection into the autocovariance eigenfunction system is considered in Ruiz-Medina and Espejo (2012), while the projection into the eigenfunction system of the SARH(1) parameters is applied in Ruiz-Medina (2012a). In the mentioned approaches, in terms of projections, the operators defining the parameters of SARH(1) state equation are estimated without assuming any parametric form for them. In that sense, we will refer in the following to the non-parametric SARH(1) model fitting.

A crucial problem in the SARH(1) extrapolation implementation by projection is the selection of a suitable basis. Specifically, it is well-known that the projection into the autocovariance eigenfunction system leads to the maximum dimension reduction for a prescribed level of explaining variability, in absence of high correlated curves (see Ruiz-Medina and Espejo, 2012). While the eigenfunction system providing the spectral decomposition of the operators defining the infinite-dimensional parameters of the SARH(1) state equation leads to a regularization of

the associated moment-based projected equation system, when its determinant is close to zero (see Ruiz-Medina, 2012a). The projection methodology presented here, based on the discrete interval wavelet transform, removes strong spatial correlations between ocean surface temperature curves, due, for example, to high concentration level of weather stations. That is, in the wavelet domain, the projected covariance operators are defined in terms of infinite-dimensional sparse matrices. Hence, this domain allows to fit a spatial Markovian dependence to the functional data analyzed. Note that, in the approach presented, an underlying Markovian state space equation is assumed, where each functional value of the spatial process of interest is generated from the negative spatial lags of order one in the vertical, horizontal or diagonal directions (see also Remark 4 below).

The limitations of the linear modeling framework are well known, but allows, as commented before, the representation of some features that are commonly present in the interaction between sea surface temperature anomalies and the atmosphere. In particular, in relation to our objective of detecting anomalies in the velocity decay of the daily ocean surface temperature curves at the end of the day and during the night, it is essential to have a spatial extrapolation method that reflects the global interaction between different locations of the daily temperature curves. The same global prediction in space of daily temperature curves is needed in the detection of increase daily functional trends in ocean surface temperature. This global daily temporal information, that is needed to be spatially extrapolated, constitutes the motivation of the presented stochastic spatial functional extrapolation methodology, based on Spatial Autoregressive Hilbertian models. Note that classical statistical methodologies, based, for example, in kriging interpolator or Generalized Linear Mixed models do not provide spatial extrapolation of daily temperature functional evolution, as well as information on spatial propagation of micro-scale temporal features is lost. As commented before, in our approach, spatial Markovian interaction is assumed according to the spatial unilateral dynamics displayed by Spatial

Autoregressive Hilbertian processes of order one (SARH(1) processes), see equation (A2.1) below, in contrast with the spatial functional extrapolation methodologies recently developed in absence of an underlying state-space model (see, for example, Delicado *et al.*, 2010; Giraldo, Delicado and Mateu, 2010; Monestiez and Nerini, 2008; Nerini, Monestiez and Manté, 2010). One of the main reasons for considering a state-space-based spatial functional framework is the integration of the physical law in the stochastic modeling approach, allowing the derivation of spatial functional extrapolators, requiring the incorporation of a spatial dynamics or association. In our case, SARH(1) processes can be constructed from irregular spatial functional sampling of two-parameter diffusion processes, in the class introduced by Nualart and Sanz-Solé (1979), including the two-parameter Ornstein-Uhlenbeck process (see Ruiz-Medina, 2012a). Note that the so-called linear Gaussian state-space model framework has been extensively studied in the literature (see e.g. Durbin and Koopman, 2001, and references therein). For example, Ide *et al.* (1997) proposed unified notations for state-space models and data assimilation in oceanography and meteorology, which also were partially adopted by Tandeo, Ailliot, and Autret (2011).

Returning to our initial goal, global warming alters air temperature faster, decreasing the differences between ocean and air surface temperatures. Thus, heat fluxes from the sea into the air are reduced. In particular, coastal ocean surface temperature increases. The designed functional spatial stochastic inter/extrapolation methodology is applied to the construction of mean annual daily ocean surface temperature maps to investigate global warming effects. Specifically, as commented, two global warming effects are investigated. First, the damping of sea-air temperature differences is studied over the connected mid-latitude ocean and coastal regions of Hawaii ocean and the Gulf of México. Additionally, an increase trend in ocean surface temperature can be appreciated along the time, due to global warming. This effect is studied in relation to the increase of Australian ocean surface temperatures, and the pronounced differences between the western and eastern Australian sea temperatures, jointly with the fact that the flow of the East

Australian Current has strengthened in the past years, altering marine biodiversity.

A2.2 Areas studied and data

The performance of the proposed spatial functional time series methodology for stochastic inter/extrapolation of the mid-latitude ocean surface temperature dynamics is tested, considering the spatiotemporal extrapolation problem associated with the construction of a daily ocean surface temperature map from Hawaii ocean (latitude-longitude interval $[16, 32] \times [-170, -140]$) to the Gulf of México (latitude-longitude interval $[16, 32] \times [-100, -80]$), and from Indian Ocean (latitude-longitude interval $[-34, -15] \times [75.4, 99.5]$) to the Pacific Ocean at the eastern coast of Australia (latitude-longitude interval $[-20, -4] \times [136, 153]$).

We briefly comment why we have selected these regions in our study. Extreme temperature series of the Gulf of México reveals the presence of periodicities similar to those found in meteorological and solar activity phenomena (see, for example, Maravilla, Mendoza and Jáureguib, 2008, in relation to Maximum Entropy based analysis). This fact suggests that the solar activity signals are possibly present in the minimum extreme temperature records of this Mexican region. Since the Gulf of México constitutes the most plausible source location of the anomalous waters observed in the Hawaii ocean surface temperature, both regions are analyzed in this paper. In particular, the derived mean annual daily temperature maps reveal the damping through time of sea-air temperature differences, due to global warming. Additionally, the influence of Indian Ocean sea-surface temperature variability on winter rainfall across eastern Australia is also well-known. Therefore, we also study these regions connected by the ocean surface temperature maps obtained by SARH(1)-based stochastic extrapolation. In particular, we investigate the possible increase of western and eastern ocean Australian temperature differences, due to the fact that eastern Australian waters warming faster than Indian Ocean.

A2.3 Computational methodology

The details on the implementation of the spatial functional time series model framework for stochastic inter/extrapolation of temperature profiles are now provided. First, we introduce the class of SARH(1) models fitted.

Definition 2. *A spatial functional process* $\mathbf{Y}_{SARH} = \{Y_{ij},\ (i,j) \in \mathbb{Z}^2\}$, *with values in a separable Hilbert space* $\mathcal{H}$, *is said to be a unilateral SARH(1) process if it is stationary and it satisfies the following equation:*

$$Y_{i,j} = R + L_1(Y_{i-1,j}) + L_2(Y_{i,j-1}) + L_3(Y_{i-1,j-1}) + \epsilon_{i,j}, \tag{A2.1}$$

where $R \in \mathcal{H}$ *and, for* $i = 1, 2, 3$, $L_i \in \mathcal{L}(\mathcal{H})$, *the space of linear bounded operators. The spatial functional innovation process* ϵ *is assumed to be a spatial functional martingale difference, uncorrelated with the functional random initial conditions,* Y_{00}, Y_{10}, Y_{01}, *and satisfying* $E\|\epsilon_{i,j}\|^2_{\mathcal{H}} = E\|\epsilon_{1,1}\|^2_{\mathcal{H}} = \sigma^2$, *for all* $(i,j) \in \mathbb{Z}^2$, *and* $E[\epsilon_{i,j} \otimes \epsilon_{k,l}] = E[\epsilon_{|i-k|,|j-l|} \otimes \epsilon_{0,0}] = C_{\epsilon_{|i-k|,|j-l|}\epsilon_{0,0}}$, *for all* (i,j) *and* (k,l) *in* $\mathbb{Z}^2$. *Here ,* $\otimes$ *denotes the tensorial product of two function in* $\mathcal{H}$, *that defines a Hilbert-Schmidt operator on* $\mathcal{H}$ *as follows: For two functions* $f, g \in \mathcal{H}$,

$$f \otimes g(h) = \langle f, h^* \rangle_{\mathcal{H}}\, g, \quad h \in \mathcal{H}^*,$$

where $h^* \in \mathcal{H}$ *is the dual element of* h, *defined from Riesz Representation theorem, and* $\mathcal{H}^*$ *denotes the dual Hilbert space of* $\mathcal{H}$.

Remark 4. *Note that, in the above definition, the order one of the SARH(1) process family introduced refers to the fact that the functional value* $Y_{i,j}$ *interacts in space with the values* $Y_{i-1,j}$, $Y_{i,j-1}$ *and* $Y_{i-1,j-1}$, *respectively corresponding to one negative spatial lag at coordinate* i, j, *or at both spatial coordinates* i *and* j.

The weak spatial dependence scheme displayed by the functional innovation process ϵ is introduced in the following definition (see Ruiz-Medina, 2011).

Definition 3. *Let* $\{X_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *and* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *be two-parameter stochastic sequences. Then,* $\{X_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *is a spatial or two-parameter martingale difference sequence with respect to the sequence* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *if the conditional expectation*

$$E^{\mathcal{B}_{\mathcal{P}_{(n,m)}}}(X_{n,m}) = 0,$$

where

$$\mathcal{B}_{\mathcal{P}_{(n,m)}} = \sigma\left(Z_{i,j},\ (i,j) \in \mathcal{P}_{(m,n)}\right) = \sigma\left(Z_{i,j},\ i < n \quad \text{or} \quad j < m\right)$$

denotes the $\sigma-$*algebra generated by the past values of the stochastic sequence* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$.

From Definition 2, and, in particular, from model (A2.1), the following linear equation system is satisfied by operators L_i, $i = 1, 2, 3$,

$$\begin{aligned} R_{1,0} &= L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1} \\ R_{0,1} &= L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0} \\ R_{1,1} &= L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0}, \end{aligned} \tag{A2.2}$$

where

$$\begin{aligned} R_{0,0} &= R_{Z_{i,j}Z_{i,j}} = E\left[Z_{i,j} \otimes Z_{i,j}\right] \\ R_{1,0} &= R_{Z_{i+1,j}Z_{i,j}} = E\left[Z_{i+1,j} \otimes Z_{i,j}\right] \\ R_{0,1} &= R_{Z_{i,j+1}Z_{i,j}} = E\left[Z_{i,j+1} \otimes Z_{i,j}\right] \\ R_{1,1} &= R_{Z_{i+1,j+1}Z_{i,j}} = E\left[Z_{i+1,j+1} \otimes Z_{i,j}\right], \end{aligned} \tag{A2.3}$$

with $Z_{i,j} = Y_{i,j} - R$, for all $(i,j) \in \mathbb{Z}^2$.

Remark 5. *Note that, under very general conditions (see, for example, Bosq, 2000; Bosq and Blanke, 2007, and Horváth and Reeder, 2012), the convergence in probability of the empirical auto-covariance and cross-covariance operators to the theoretical ones holds.*

The spatial dependence structure displayed by functional process Z is defined in terms of operators in equation A2.3. Operators L_i, $i = 1, 2, 3$, are involved in the definition of the correlation operator of the spatial functional process Z. Therefore, we will refer to them as autocorrelation operators (see Ruiz-Medina, 2011, on SMAH(∞) representation of SARH(1) processes). To illustrate the spatial dependence structure displayed by process Z, in Figure A2.1, the above introduced covariance and autocorrelation operators are showed, in the case where Z is generated from the spatial functional sampling of the two-parameter Ornstein-Uhlenbeck process (see Ruiz-Medina, 2012a). Specifically, in the top panels of this figure, the following vectorized empirical covariance operators can be found, for $N = M = 10$ (see Ruiz-Medina, 2012a):

$$\begin{aligned}\hat{R}^V_{0,0} &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Vec}(Z_{i,j})\otimes\mathrm{Vec}(Z_{i,j})\\ \hat{R}^V_{1,1} &= \frac{1}{(N-1)^2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}\mathrm{Vec}(Z_{i+1,j+1})\otimes\mathrm{Vec}(Z_{i,j}),\end{aligned} \tag{A2.4}$$

where, for each $i, j = 1, \ldots, N$, and $\{Z_{i,j} = Z_{i,j}(x_k, x_l),\ k, l = 1, \ldots, M\}$,

$$\mathrm{Vec}(Z_{i,j}) = \left(Z_{i,j}(x_1, x_2), \ldots, Z_{i,j}(x_1, x_M), Z_{i,j}(x_2, x_1), \ldots, Z_{i,j}(x_M, x_M)\right)^T, \tag{A2.5}$$

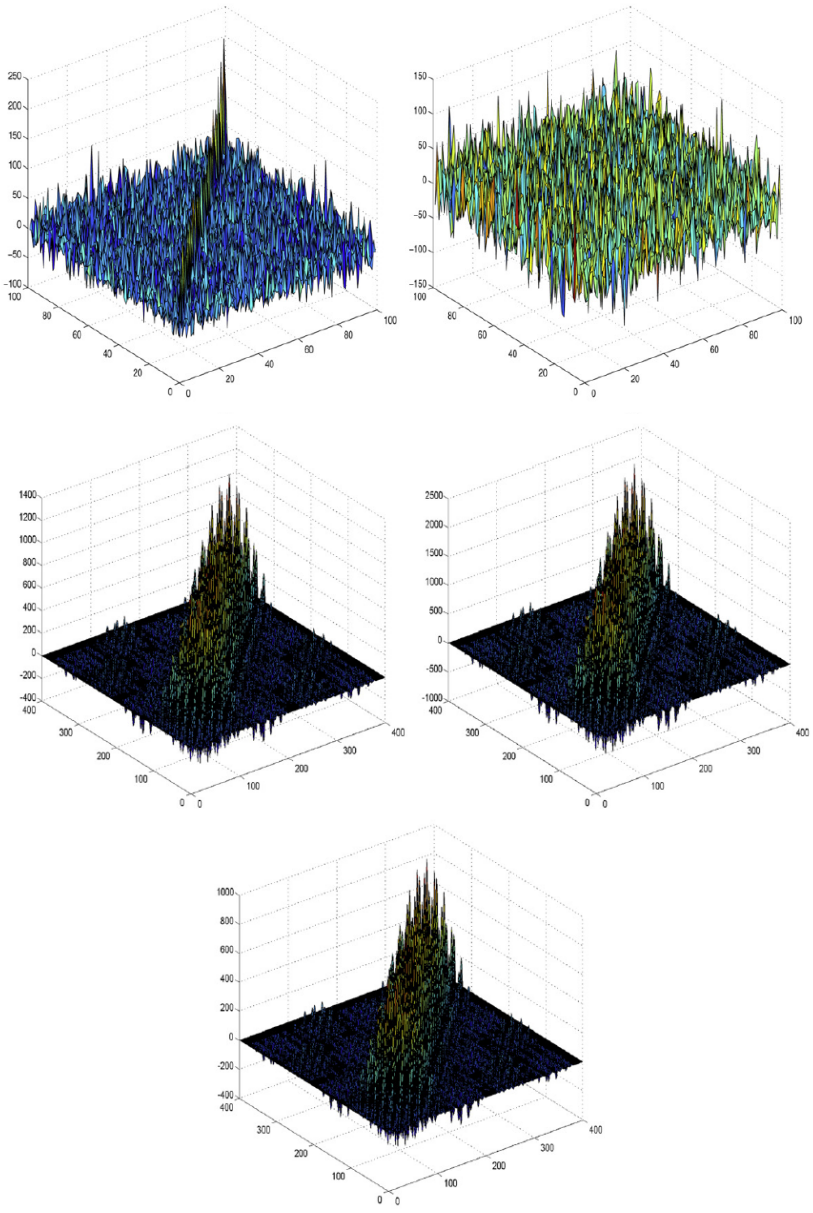
with ^{T} being the transpose. Additionally, for $i = 1, 2, 3$, denote by l_i the spectral kernel of L_i, in the center and in the bottom of Figure A2.1, the surfaces displayed are given from interpolation of

$$\begin{pmatrix} l_i(x_1,x_1;x_1,x_1), & \ldots, & l_i(x_1,x_1;x_1,x_M), & \ldots, & l_i(x_1,x_1;x_M,x_M)\\ l_i(x_1,x_2;x_1,x_1), & \ldots, & l_i(x_1,x_2;x_1,x_M), & \ldots, & l_i(x_1,x_2;x_M,x_M)\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ l_i(x_M,x_M;x_1,x_1), & \ldots, & l_i(x_M,x_M;x_1,x_M), & \ldots, & l_i(x_M,x_M;x_M,x_M)\end{pmatrix},$$

where the values $N = 10$ and $M = 20$ have been considered (see also Ruiz-Medina, 2012a).

Remark 6. *The cyclic nature of equation (A2.2) arises from the expression of the functional values $Z_{i+1,j}$, $Z_{i,j+1}$ and $Z_{i+1,j+1}$ in equation (A2.3), in terms of the functional linear combi-*

Figure A2.1: Auto-covariance operators (top-left), and one-lag-diagonal covariance operators (top-right), for $N = M = 10$, and operators L_1 (center-left), L_2 (center-right) and L_3 (bottom), from the spatial functional sampling of O-U, for M = 20.



nation of their negative spatial lags of order one, involving operators L_i, $i = 1, 2, 3$, *according to equation (A2.1), and from the spatial invariance of the second order moments of the functional innovation process* ϵ *in Definition 2.*

The coefficients of the linear equation system (A2.2) are the autocovariance, and cross covariance operators introduced in equation (A2.3), i.e., this system is defined in terms of infinite-dimensional coefficients. From now on, we will then refer to it as the functional linear equation system satisfied by the operators L_i, $i = 1, 2, 3$. In practice, model fitting is achieved by solving such a functional linear equation system (A2.2), in terms of the empirical covariance operators

$$\begin{aligned}
\hat{R}_{0,0} &= \frac{1}{NN} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_{i,j} \otimes Z_{i,j} \\
\hat{R}_{1,0} &= \frac{1}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=1}^{N} Z_{i+1,j} \otimes Z_{i,j} \\
\hat{R}_{0,1} &= \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N-1} Z_{i,j+1} \otimes Z_{i,j} \\
\hat{R}_{1,1} &= \frac{1}{(N-1)(N-1)} \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} Z_{i+1,j+1} \otimes Z_{i,j},
\end{aligned} \tag{A2.6}$$

where $\{Z_{i,j},\ i = 1, \ldots, N,\ j = 1, \ldots, N\}$ denotes the functional sample located at the spatial nodes of a regular grid.

The first problem we have to address, previously to our spatial functional time series model fitting, is the construction of a functional sample on a regular spatial grid from the observed non-equally spaced mean annual daily temperature profiles. Specifically, weather station location concentration at deep ocean regions hinders the analysis, due to the presence of redundant sample information, as well as border effects. The interval wavelet transform is applied to eliminate these two effects. It is well known that this transform provides a sparse and more efficient border representation of smooth functions (see, for example, Cohen, Daubechies and Vial, 1994; Meyer, 1991; Chyzak *et al.*, 2001). Thus, the main steps of the presented stochastic

inter/extrapolation methodology are the following:

Data-driven spatial block design. Mean annual daily temperature profiles are averaged over the spatial boxes of a computational grid to construct a functional sample on a spatial regular grid. The spatially averaged values over each box are assigned to its lower left corner. In the design procedure of a lattice configuration, the resolution of this grid is not fixed, and depends on the spatial density of temperature profiles over each spatial block, splitting boxes with a high-density level of temperature profiles (see Ruiz-Medina and Espejo, 2012, for more details on the design of the final spatial computational grid).

Spatial functional time series model fitting. A SARH(1) model is fitted to the generated functional sample of spatial aggregated temperature profiles on a regular grid. Specifically, the empirical covariance operators $\hat{R}_{0,0}$, $\hat{R}_{1,0}$, $\hat{R}_{0,1}$ and $\hat{R}_{1,1}$ are computed. To obtain an explicit solution to the functional empirical equation system constructed from (A2.2), numerical projection methods, on a suitable orthonormal basis, are applied. Specifically, we consider the discrete interval wavelet transform, based on Daubechies wavelet functions, at a specific resolution level M. The following projection wavelet estimates are then obtained:

$$\begin{bmatrix} \widehat{\mathcal{W}_{2D}^M}(L_1) \\ \widehat{\mathcal{W}_{2D}^M}(L_2) \\ \widehat{\mathcal{W}_{2D}^M}(L_3) \end{bmatrix} = \begin{bmatrix} \mathcal{W}_{2D}^M(\hat{R}_{0,0}) & \mathcal{W}_{2D}^M(\hat{R}_{1,1}) & \mathcal{W}_{2D}^M(\hat{R}_{0,1}) \\ \mathcal{W}_{2D}^M(\hat{R}_{1,1}) & \mathcal{W}_{2D}^M(\hat{R}_{0,0}) & \mathcal{W}_{2D}^M(\hat{R}_{1,0}) \\ \mathcal{W}_{2D}^M(\hat{R}_{0,1}) & \mathcal{W}_{2D}^M(\hat{R}_{1,0}) & \mathcal{W}_{2D}^M(\hat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \mathcal{W}_{2D}^M(\hat{R}_{1,0}) \\ \mathcal{W}_{2D}^M(\hat{R}_{0,1}) \\ \mathcal{W}_{2D}^M(\hat{R}_{1,1}) \end{bmatrix}, \quad \text{(A2.7)}$$

where $\mathcal{W}_{2D}^M$ denotes the two-dimensional discrete interval wavelet transform at resolution level M.

Stochastic inter/extrapolation. This step is performed in terms of the inverse of the two-dimensional discrete interval wavelet transforms, $\widehat{\mathcal{W}_{2D}^M}(L_i)$, $i = 1, 2, 3$, computed in the

previous step. Thus, the SARH(1)-based stochastic inter/extrapolator is defined as:

$$E\left[Z_{i,j} \mid Z_{k,l},\ k<i,\ \text{or } l<j,\ \text{or } k<i \text{ and } l<j\right] = \\ = \widehat{L_1} Z_{i-1,j} + \widehat{L_2} Z_{i,j-1} + \widehat{L_3} Z_{i-1,j-1} \qquad \text{(A2.8)}$$

where $\widehat{L_i}$, $i = 1, 2, 3$, represent the respective inverses of the two-dimensional discrete interval wavelet transforms $\widehat{\mathcal{W}_{2D}^{M}}(L_i)$, $i = 1, 2, 3$, at resolution level M.

It is well-known that for signals with finite support, the interval wavelet transform provides a suitable processing of edges, avoiding border distortion. Specifically, special filters are designed to replace the original filters at the signal borders, that is, they are adapted to interval borders, and they do not require signal values outside the interval. The boundary wavelets possess vanishing moments that allow to exploit the signal smoothness near the borders. In our case, the Hilbert space $\mathcal{H}$ is constituted by functions defined on a finite temporal interval. To derive the wavelet-based system (A2.7) of projected equations, the values of the empirical covariance operators at the horizontal and vertical directions are transformed using suitable boundary filter assignment rules to produce sharper edges with less artifacts than the ones produced by classical wavelet transform. On the other hand, in the computation of SARH(1)-plug-in extrapolator (A2.8), the inverse interval wavelet transform is applied, which requires information about the position of the two-dimensional interval boundaries. The information needed for each wavelet coefficient then consists of its position and value, in order to derive an almost perfect reconstruction. To extract edge information simple edge detection schemes have been applied (see, for example, Lee and Kassim, 2007).

Remark 7. *The above described stochastic spatial functional extrapolation algorithm is based on the assumption that the process of interest* Y_{SARH} *satisfies the state equation (A2.1). Specifically, in the estimation of operators* L_i, $i = 1, 2, 3$, *we have applied the spatial unilateral dynamics*

reflected in (A2.1). Furthermore, in the computation of the spatial extrapolator (A2.8), we have considered as the vector of explanatory variables the functional values of process $Z = Y_{SARH} - R$ *at the negative spatial lags of order one involved in the recursive generation scheme (A2.1). While in the spatial classical or functional kriging, an underlying spatial state equation is not assumed. However, as in the classical case, we can consider this spatial functional extrapolator as a special case of spatial functional regression estimator, obtained from the above-referred vector of spatial functional explanatory variables. That is, in our case, for each spatial location* (i, j)*, the spatial functional response is given by* $Z_{i,j}$*, and the vector of spatial functional explanatory variables has components:* $Z_{i-1,j}$*,* $Z_{i,j-1}$ *and* $Z_{i-1,j-1}$.

A2.4 Results

The objective of this section is to investigate how global warming affects the usual influence of ocean-air surface temperature differences in regional daily climate variability. Specifically, we study the effect of global warming on daily heat fluxes from the sea into the air. As commented before, Hawaii ocean and the Gulf of México will be connected through a mean annual daily ocean surface temperature map derived by application of the proposed SARH(1)-based estimation methodology. Additionally, mean annual daily ocean surface temperature is spatially functional extrapolated from the Indian ocean to the Pacific ocean (eastern of Australia coast). The suitability of the SARH(1) spatial dynamics, fitted by projection into an interval wavelet basis of Daubechies type, is tested by applying 10-fold Cross Validation (10- fold CV) procedures. The results are compared to those ones obtained by considering the projection into the eigenfunction systems involved in the spectral decomposition of SARH(1) parameters, L_i, $i = 1, 2, 3$ (see Ruiz-Medina, 2012a).

A2.5 Validation of the model

Ocean surface daily temperature profiles are collected over the Gulf of México, the Hawaii ocean, Indian Ocean, and the western Australia coast, from the public oceanographic bio-optical database, *The World-wide Ocean Optics Database (WOOD)*. The years 2000 (in the Gulf of México and Hawaii ocean), and 1995 (in the western of Australia coast and Indian Ocean) have been selected for illustration purposes, to measuring the spatial stochastic extrapolation capability of the proposed methodology, in terms of cross validation procedures, in comparison with the projection methodology based on the eigenfunction systems of the SARH(1) parameters. Note that, in the implementation of the proposed estimation algorithm, the truncation order M involved in the wavelet-based finite-dimensional approximation (A2.7) of the functional equation system (A2.2) is selected by cross validation. In all the spatial regions analyzed, Gulf of México, Hawaii ocean, western coast of Australia, and Indian Ocean, a data size of 200 daily temperature observations is considered, for each temperature profile. This data size substantially limits the number of available weather stations in the WOOD database. In the tropical and subtropical regions studied, small changes in temporal global and local patterns of ocean surface daily temperature profiles can be appreciated between different months. Thus, in these regions, the derived extrapolation results would not be specially affected by temporal scale changes. However, spatial functional extrapolation results will be more sensitive to changes in time scale if spatial block averaging precedes the temporal one, since sampled weather stations substantially change through months, specially, as commented before, when daily temperature records of size 200 are required.

The value $M = 7$ resolution levels has been established in the finite-dimensional approximation of space $\mathcal{H}$, based on the discrete interval wavelet transform projection methodology applied. Thus, $128 \times 128 = 16384$ wavelet functions are involved in the finite-dimensional approximation (A2.7) of the functional equation system (A2.2), from the tensorial product of one-dimensional

interval wavelet functions. A 6×25 spatial computational grid has been fitted in the four analyzed regions. For each box of the 6×25 computational grid, the areal functional data obtained after spatial averaging daily temperature profiles over such a box is assigned to its lower left corner. Spatiotemporal interpolation is then performed over the empty nodes.

The $10-$fold CV procedure is applied to validate the SARH(1) model fitted, as well as the projection methodology considered in this fitting. The $10-$fold CV mean absolute error curves over the spatial nodes of the validation sample are shown in Figure A2.2, for the Gulf of México, in Figure A2.3, for the Hawaii ocean, in Figure A2.4, for the western coast of Australia, and in Figure A2.5, for the Indian Ocean. This kind of k-fold CV allows the reduction of CV error variability. The validation sample is located at a 3×5 sub-grid, randomly selected from the global 6×25 computational grid, i.e., at each iteration of the $10-$fold CV procedure, this sub-grid can move its location from the left to the right, and from the top to the bottom of such a 6×25 spatial grid. The learning sample then consists of the remaining 135 nodes to fill the complementary space of the 6×25 spatial grid outside the sub-grid occupied by the validation sample, at each iteration of the $10-$fold CV procedure applied. Table A2.1 shows the L^{∞}-norms of the $10-$fold CV mean absolute error curves at each node of the 3×5 validation sample, removing the first row and column. These validation results are compared to those ones obtained by applying projection into the empirical eigenfunction system of the infinite-dimensional parameters L_i, $i = 1, 2, 3$, see Table A2.2, where the truncation order $T = 20$ has been considered, i.e., the projection into the finite dimensional space $\widetilde{\mathcal{H}}$ generated by 20 eigenfunctions has been considered, which is equivalent to consider 200 eigenfunctions of the tensorial product basis, in the finite-dimensional approximation of the estimation equation system (A2.2). It can be appreciated that the wavelet-based projection clearly outperforms the SARH(1) parameter eigenfunction projection, since larger dimension, due to its sparsity, than in the eigenfunction approach can be considered in the finite-dimensional approximation of the

Table A2.1: *Two-dimensional discrete interval wavelet projection: L^∞-norm of the 10−fold CV mean absolute error curves, for $M = 7$ resolution levels*

	Gulf of México	Hawaii ocean	Australia	Indian Ocean
(2,2)	0.0246	0.3864	0.1305	0.0487
(3,2)	0.0275	0.4315	0.1728	0.0401
(2,3)	0.0226	0.4082	0.1285	0.0456
(3,3)	0.0287	0.4475	0.1587	0.0387
(2,4)	0.0286	0.4070	0.1370	0.0367
(3,4)	0.0416	0.4429	0.1643	0.0372
(2,5)	0.0333	0.3838	0.1502	0.0574
(3,5)	0.0434	0.4185	0.1734	0.0428

space $\mathcal{H}$. Note that, in the SARH(1) parameter eigenfunction projection, the consideration of higher truncation orders do not improve the performance of the projection methodology due to the redundancy in the information provided by the terms of the finite dimensional expansion of the ocean surface temperature curves, in terms of the eigenfunctions of operators L_i, $i = 1, 2, 3$. However, since the wavelet transform provides sparse covariance matrices in the definition of equation system (A2.7), redundancy is removed weakening strong correlations, specially in the region of Hawaii ocean, where it can be appreciated, see Figure A2.6, a high concentration level of weather stations. On the other hand, it can also be appreciated the improvement obtained with the discrete interval wavelet transform in the Indian Ocean region, due to the suitable processing of border effects.

In all the figures below the following abbreviation is considered: MAECij denotes the 10−fold CV mean absolute error curve over the spatial node (i,j) of the validation sample (e.g., MAEC22 means mean absolute error curve over the spatial node (2,2) of the validation sample).

The spatial distribution of weather stations in the four regions analyzed can be found in Figure A2.6 below.

Figure A2.2: Gulf of México, year 2000. Mean absolute error curve (MAEC) over the spatial nodes of the validation sample in the application of a 10-fold CV with $M = 7$ resolution levels

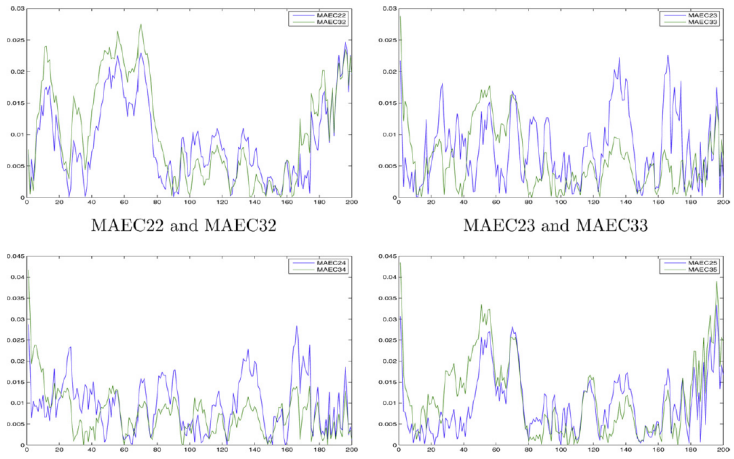


Figure A2.3: Hawaii Ocean, year 2000. Mean absolute error curve (MAEC) over the spatial nodes of the validation sample in the application of a 10-fold CV with $M = 7$ resolution levels

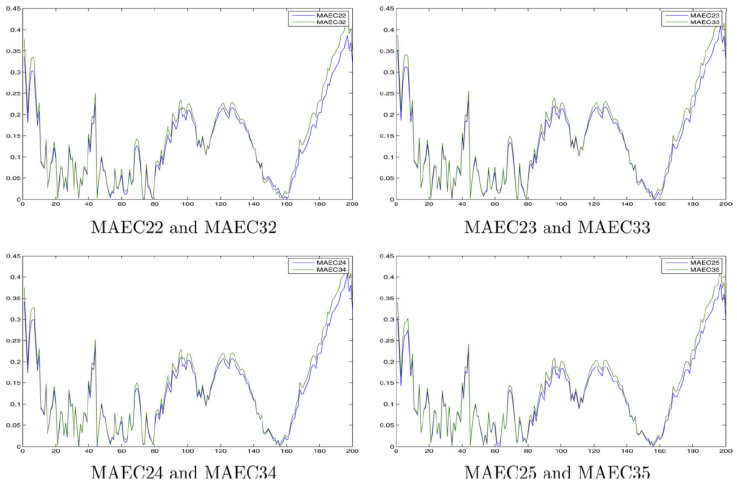


Figure A2.4: Western Australia, year 1995. Mean absolute error curve (MAEC) over the spatial nodes of the validation sample in the application of a 10-fold CV with $M = 7$ resolution levels

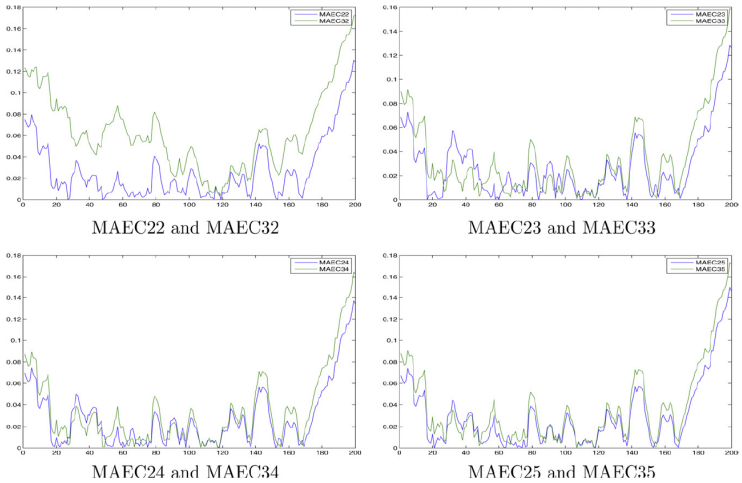


Figure A2.5: Indian Ocean, year 1995. Mean absolute error curve (MAEC) over the spatial nodes of the validation sample in the application of a 10-fold CV with $M = 7$ resolution levels.

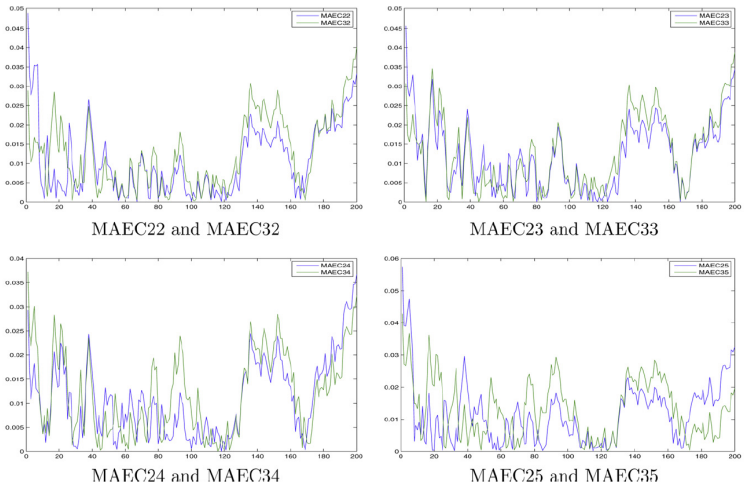
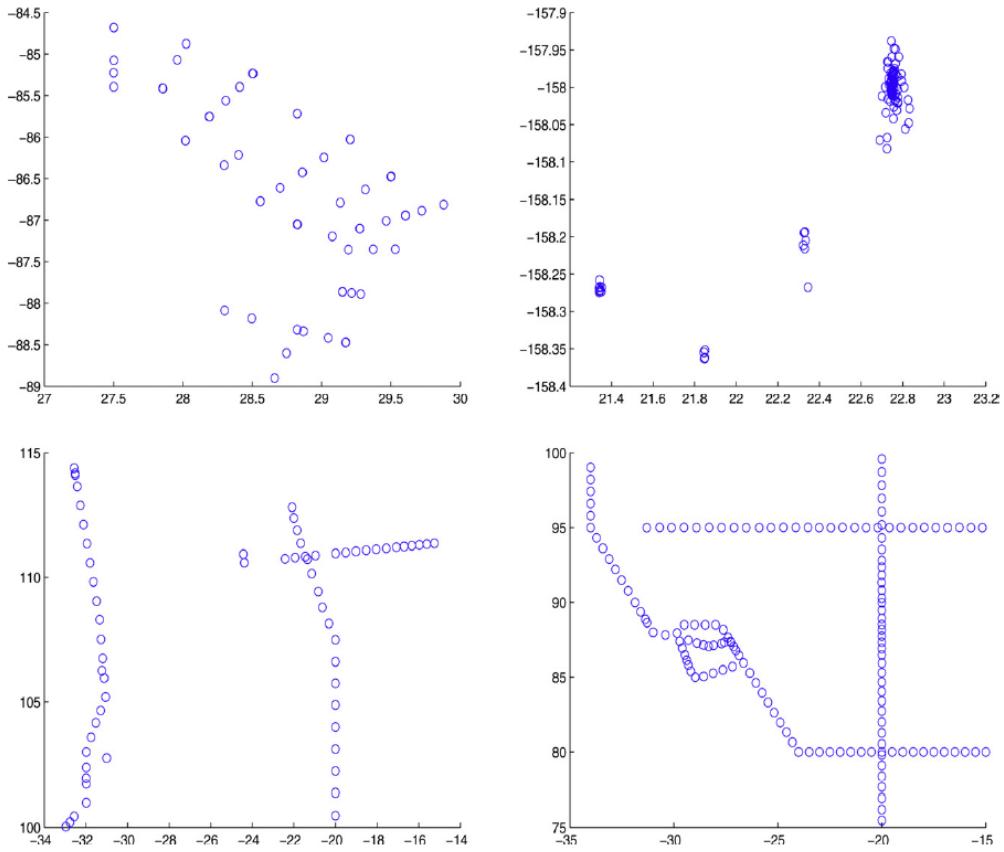


Table A2.2: *SARH(1)-parameter eigenfunction projection: L^∞-norm of the 10−fold CV mean absolute error curves, for $T = 20$*

	Gulf of México	Hawaii ocean	Australia	Indian Ocean
(2,2)	0.0769	0.5785	0.6047	0.3114
(3,2)	0.1098	0.7183	0.6495	0.2649
(2,3)	0.0920	0.6211	0.5835	0.3802
(3,3)	0.0786	0.6785	0.6975	0.2912
(2,4)	0.1059	0.6327	0.9539	0.3008
(3,4)	0.0879	0.6845	0.9624	0.2192
(2,5)	0.1270	0.6125	0.5647	0.2939
(3,5)	0.1342	0.6621	0.6426	0.2420

Figure A2.6: Spatial distribution of weather stations at the Gulf of México (top-left), Hawaii Ocean (top-right), Western coast of Australia (bottom-left) and Indian Ocean (bottom right).



A2.6 Temperature map

Global warming changes the air temperature more quickly, and induces a decreased temperature differences between sea and air, which is expected to reduce the heat fluxes from the sea into the air at night. In this section, this global warming effect is investigated, in the Hawaii ocean and the coast of the Gulf of México, in terms of the daily evolution of ocean surface temperature from the mid-day to the night, reflected in the spatial extrapolation temperature maps, derived with the application of the proposed functional estimation methodology. The strengthening of ocean current flows is also investigated through spatial ocean surface temperature maps, reflecting minimum daily temperature differences between the Indian Ocean and the eastern coast of Australia, in the Pacific Ocean.

First, mean annual daily temperature maps are obtained covering the spatial evolution of ocean surface temperature from Hawaii to the Gulf of México. Specifically, the sea-surface temperature values are extrapolated from Hawaii ocean (left-hand side of the graphs below in Figures A2.7 and A2.8) to the coastal ocean of the Gulf of México (right-hand side of the graphs below in Figures A2.7 and A2.8). Within the period of the last 15 years, the years 1995 (see Figure A2.7) and 2006 (see Figure A2.8) have been selected to detect a possible global warming effect through this period, reflected, as commented before, in a reduction of the air-sea temperature differences, and higher coastal ocean surface temperatures at night. These years have been chosen, since they display the largest numbers of weather stations having available high-dimensional daily temperature profiles. It can be appreciated that sea-air temperature differences are smaller during the year 2006 (see Figure A2.8) than in the year 1995 (see Figure A2.7), leading to a reduction of heat fluxes from the sea into the coast at night. Specifically, regarding minimum daily temperatures, the temperature of the coastal ocean surface ranges between 8.5C and 11C in the year 1995 (see right-hand-side of bottom-right panel of Figure A2.7), while, during the year 2006, it ranges between 11C and 12.5C (see right-hand-side of

bottom-right panel of A2.8). Thus, an increase of the coastal ocean surface minimum daily temperature is experimented at the Gulf of México within this period. Note that the graphs in Figures A2.7 and A2.8 display the ocean surface temperature from Hawaii to the Gulf of México at the same daily time intervals from mid-day to night. In the second panel (top-right panel) of these figures coastal ocean surface temperature begins to go down. The two central panels reflect heat fluxes from the sea into the air, due to the difference between coastal and ocean surface temperatures. In the last two panels, an equilibrium between these temperatures is reached. A faster drop in coastal ocean surface temperature range can be appreciated during the year 1995 with respect to the year 2006, as it is reflected in Figures A2.7 and A2.8. Specifically, in Figure A2.8, coastal temperature range, from midday to the afternoon, goes from $[26\text{C}, 27\text{C}]$ (right-hand-side of top-left panel) to the range $[25.5\text{C}, 26.5\text{C}]$ (right-hand-side of top-right panel), while during the year 1995 (see Figure A2.7), it goes from $[25\text{C}, 27.5\text{C}]$ (right-hand-side of top-left panel) to $[24\text{C}, 25.5\text{C}]$ (right-hand-side of top-right panel). An equilibrium or homogeneity between coastal (right-hand-side of each panel) and ocean (left-hand-side of each panel) is first reached in the year 2006 (see bottom-left panels of Figures A2.7 and A2.8), since, as commented, in the year 2006 a slower decay of coastal temperatures than in the year 1995 is displayed. Finally, in the last panel (bottom-right panel) of Figures A2.7 and A2.8, it can be appreciated an increase of the temperature range of ocean surface in the year 2006. In this year (see left-hand-side of bottom-right panel of Figure A2.8), the ocean surface temperature range is $[9.5\text{C}, 13\text{C}]$, while, in the year 1995 (see left-hand-side of bottom-right panel of Figure A2.7), has been $[8.5\text{C}, 11.5\text{C}]$. Therefore, we can conclude that the increase of coastal ocean surface temperature reduces heat fluxes from the ocean into the coast, and, hence, an slightly increase of ocean surface temperature has been appreciated in the last 15 years. Note that, here, heat island effect is also present.

Since 1910-1929, ocean surface temperatures around Australia have warmed 0.7°C. Marine biodiversity is changing in south-east Australia in response to warming temperatures and a

Figure A2.7: Hawaii Ocean and Gulf of México. Mean annual daily temperature map (from midday to night), year 1995.

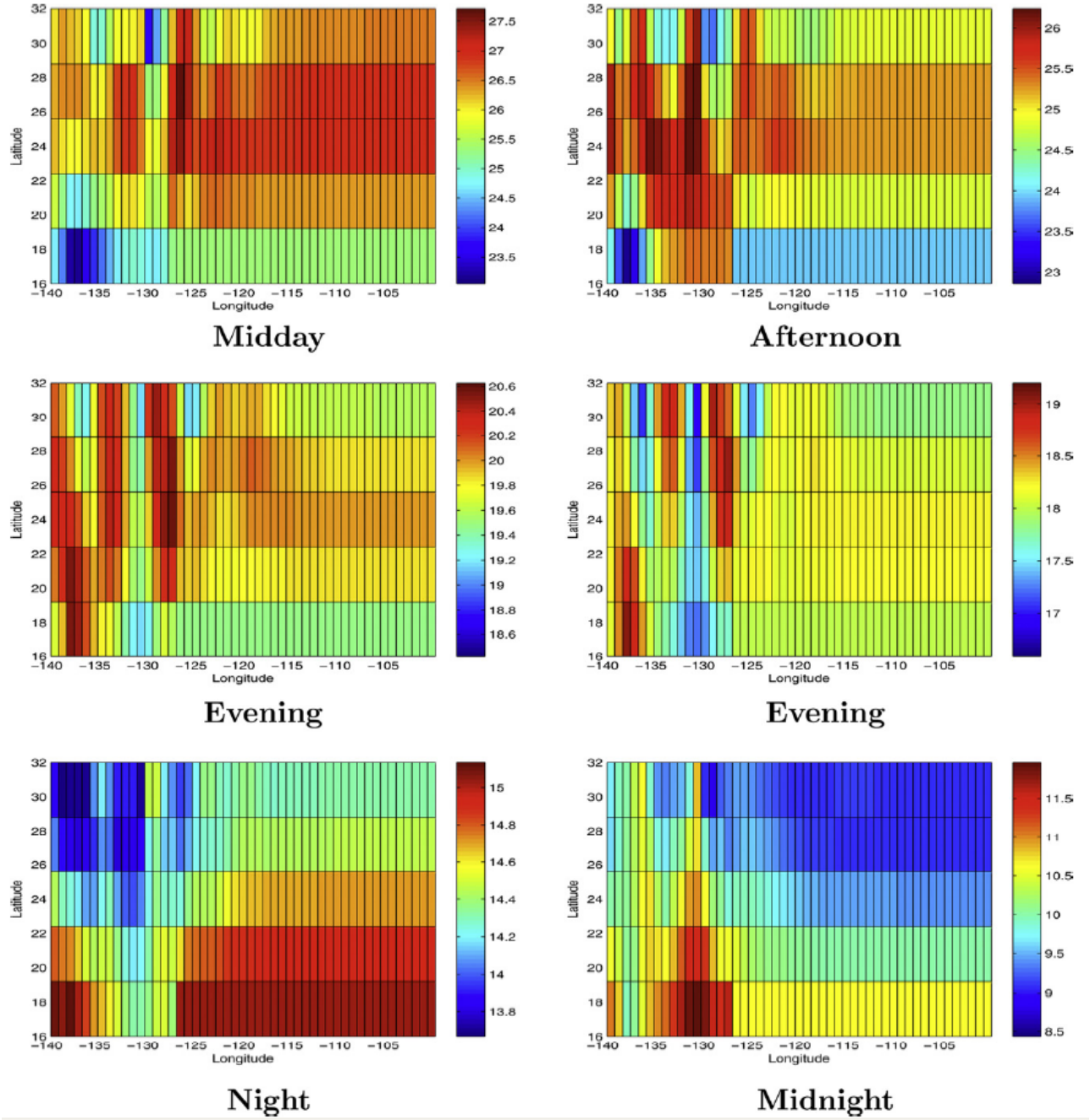
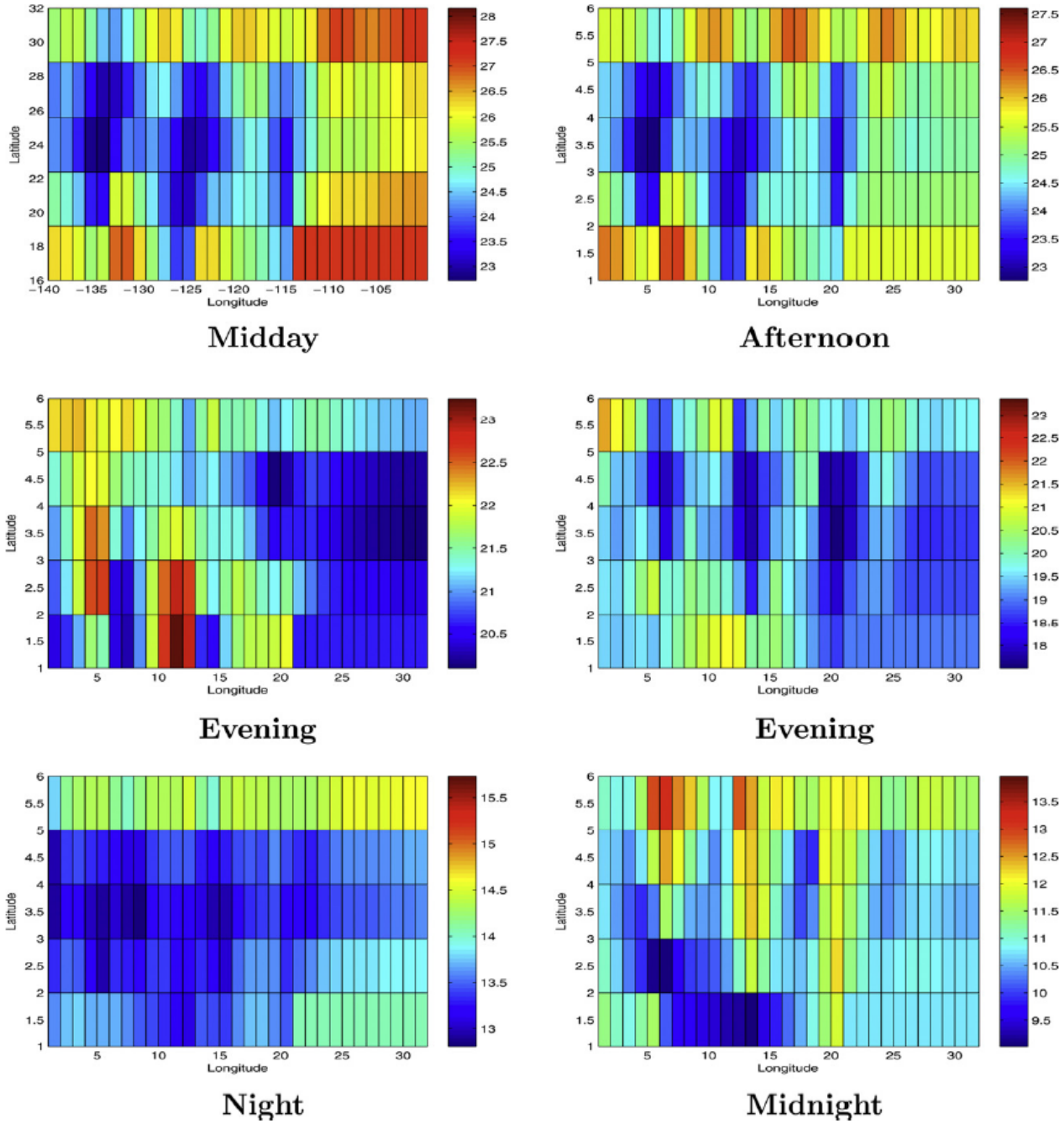


Figure A2.8: Hawaii Ocean and Gulf of México. Mean annual daily temperature map (from midday to night), year 2006.



stronger East Australian Current. In particular, we concentrate our study in ocean surface minimum daily temperature differences between Indian Ocean and the eastern coast of Australia, to detect one of the main factors of strengthening the flow of the East Australian Current. Due to the existing limitations in the availability of WOOD public oceanographic bio-optical database, in relation to weather stations with high-dimensional temperature series in Indian Ocean and the eastern coastal ocean of Australia, we have selected the year 1995, in the Indian Ocean, and the year 1997, in the eastern coastal ocean of Australia, to be compared with the year 2003, where there exists availability in both regions. Specifically, differences between the Indian Ocean surface temperature in 1995, and eastern coastal Australia ocean surface temperatures in 1997, are compared to those ones displayed between these two regions during the year 2003. In the year 1995, daily ocean surface temperature from the evening to the night ranges between 9.5C and 14C at the Indian Ocean (see top panels of Figure A2.9). While, in the eastern coast of Australia, in the year 1997, it ranges between 20.6C and 24.8C (see bottom panels of Figure A2.9). In the year 2003, the Indian Ocean surface temperature range from the evening to the night was $[4\text{C}, 12\text{C}]$ (see top panels of Figure A2.10) and, in the eastern coast of Australia, it was $[22\text{C}, 30\text{C}]$ (see bottom panels of Figure A2.10). Therefore, more pronounced differences between minimum daily temperatures of the Indian Ocean, and the Eastern coast of Australia can be appreciated in the year 2003 than in the years 1995-1997. Some strategies must then be designed in order to get an adaptation to marine climate change, and to increase resilience of Australian marine biodiversity.

A2.7 Discussion and conclusions

A new spatial functional stochastic extrapolation technique is proposed, based on SARH(1) processes. This estimation technique requires to have the data over a spatial regular grid. A lattice configuration is constructed by spatial averaging of mean annual daily temperature

Figure A2.9: Mean annual daily temperature map, at the Indian Ocean, for evening time (top-left) and for night time (top-right), during year 1995, and at the eastern coast of Australia, for evening time (bottom-left) and for night time (bottom-right), during year 1997 (bottom).

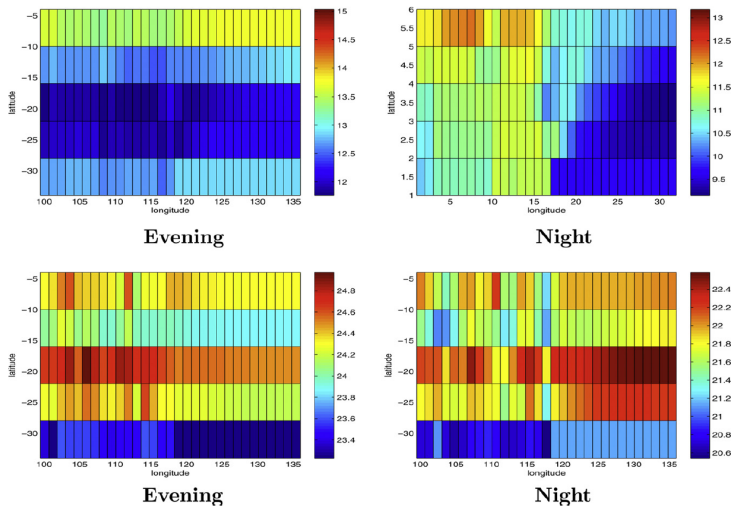
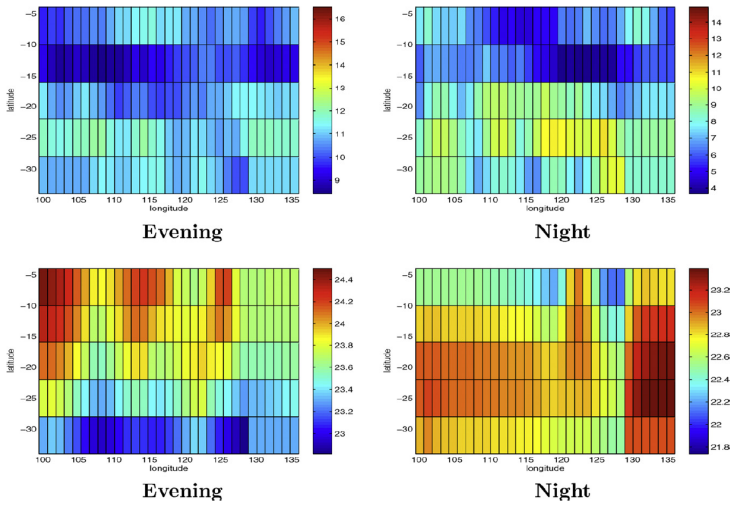


Figure A2.10: Mean annual daily temperature map, at the Indian Ocean, for evening time (top-left) and for night time (top-right), and at the eastern coast of Australia, for evening time (bottom-left) and for night time (bottom-right), during year 2003.



profiles over the boxes of a final spatial grid fitted, after splitting, in an initial computational regular grid, the boxes with the largest numbers of weather stations. An equilibrium between the aggregation level per block, and the global interpolation level must be reached. In the derivation of the final lattice configuration, the distribution of weather stations and the variability displayed by the functional data analyzed at the temporal micro- and macro-scale levels must be taken into account.

Note that the spatial aggregation of martingale difference processes at each grid box is again a spatial martingale difference process with respect to the information provided by the aggregated functional values of the innovation process in the remaining boxes (see Definition 3). Therefore, the SARH(1) fitting, performed to the spatially aggregated temperature profiles, has behind the assumption of an underlying spatial martingale difference process inducing spatial local variability at micro-scale level of daily temperature profiles. Moreover, since a functional statistical test is not yet available to verify if the invariance of the functional second-order moments of process ϵ, assumed in Definition 2, is a suitable assumption for the ocean surface temperature phenomenon investigated, the empirical evidence is only considered as verification tool. Namely, we can conclude, in terms of the 10-fold CV results obtained in Section A2.5, that the SARH(1) fitting, under the second-order moment spatial invariance assumption in Definition 2, is suitable for the areal functional data sets constructed from the observed daily temperature profiles in the four regions analyzed.

The implementation of the SARH(1) extrapolation in terms of the discrete interval wavelet transform allows to overcome the difficulties arising in the projection methodologies, based on the eigenfunction system of the autocovariance operator (see Ruiz-Medina and Espejo, 2012), as well as of the infinite-dimensional parameters defining the SARH(1) state equation (see Ruiz-Medina, 2012a). In particular, in the real-data example analyzed, as commented in Section A2.5, the suitable processing of edge effects by the interval wavelet transform improves the

results derived in the Indian Ocean with the SARH(1) parameter eigenfunction system. Additionally, this transform also weakens spatial correlations leading to better estimation results than the ones obtained with the SARH(1) parameter eigenfunction system, in presence of high spatial concentration of weather stations, removing redundancy in the data. This fact can be appreciated in the results derived in Hawaii Ocean. It is important to note, that the empirical autocovariance function projection (see Ruiz-Medina and Espejo, 2012) leads to the maximum dimension reduction for a given level of explained variability. The projection into the SARH(1) parameter eigenfunction system is advisable for the case of an estimation equation system with determinant close to zero (see Ruiz-Medina, 2012a). However, this projection methodology is not as optimal as the autocovariance eigenfunction system in relation to dimension reduction, but, it also presents the problem that, when the number of eigenfunctions considered for projection increases, redundancy present in the sample projections decreases the quality of the estimations, as it can be appreciated in the real-data example studied in Section A2.5.

Cross validation procedures are applied for studying the performance of the presented spatial functional extrapolation algorithm. This procedures are also applied in the selection of the truncation order. The 10-fold CV mean absolute error curves displayed in Section A2.5 then reflect the lattice configuration error and the truncation error, additionally to the empirical error. We have to note that, in the real-data example analyzed, the spatial averaging region error is minimum, due to the slow variation through space of the local and global temporal patterns of the temperature profiles in the tropical and subtropical regions analyzed. The spatial distribution of weather stations in the Gulf of México and the Indian Ocean better fits the SARH(1) spatial unilateral dynamics. Thus, a better integration between data and methodology is reached in the Gulf of México and the Indian Ocean than in the Hawaii Ocean and the western coast of Australia.

The mean annual daily ocean surface temperature maps derived by spatial functional extrap-

olation from the deep Hawaii ocean to the Gulf of México, allows to observe the spatiotemporal daily evolution of the heat fluxes from the sea into the air. As discussed in Section A2.6, the constructed ocean surface temperature maps show empirical evidence of a decrease of differences between sea and air temperatures at the end of the day through the years analyzed, leading to a slightly increase of ocean surface temperature. Finally, we remark that, from the temperature maps, covering the region from Indian Ocean to the eastern coast of Australia, we can conclude that more pronounced differences between minimum daily temperatures of the Indian Ocean, and the eastern coast of Australia can be appreciated in the year 2003 than in the years 1995-1997, with faster warming of the east, and a strengthening of the flow of the East Australian Current.

Appendix A3

Heterogeneous spatial dynamics regression in a Hilbert-valued context

Ruiz-Medina, M. D., Anh, V. V., Espejo, R. M. and Frías, M. P. (2013).

Heterogeneous spatial dynamical regression in a Hilbert-valued context.

Stochastic Analysis and Applications, 31, 509–527.

Abstract

This paper introduces a Hilbert-valued spatially dynamic regression model. The spatially heterogeneous functional trend is modelled by functional multiple regression, with varying regression operators. The spatial autoregressive Hilbertian model of order one (SARH(1) model, see Ruiz-Medina, 2011) is considered to represent the spatial correlation and dynamics displayed by the functional error term. The RKHS theory is applied in the construction of suitable bases for projection and regularization of the associated estimation problems. The performance of the proposed Hilbert-valued modeling and estimation methodology is illustrated with a real-data example, related to financing decisions from firm panel data.

A3.1 Introduction

The theory of linear models in functional spaces has been extensively developed in the last two decades. In particular, functional linear regression models, FANOVA models, and Hilbert space-valued time series models have been widely investigated (see, for example, Abramovich and Angelini, 2006; Bosq, 2000; Cai and Hall, 2006; Cardot *et al.*, 2003; Cardot, Ferraty and Sarda, 1999, 2003; Cardot and Sarda, 2005; Crambes, Kneip and Sarda, 2009; Fan and Zhang, 2000; James, 2002; James, Wang and Zhu, 2009; Ramsay and Silverman, 2005; Ruiz-Medina and Salmerón, 2010; Spitzner, Marron and Essick, 2003; Zoglat, 2008). Different approaches have been presented in the functional non-parametric framework in Ferraty and Vieu (2006), Hoover *et al.* (1998); Masry (2005); Rachdi and Vieu (2007); Wu, Chiang and Hoover (1998) and the references therein. A spatial functional statistical framework is adopted in Guillas and Lai (2010), based on bivariate splines, in terms of which the surface defining the explanatory random variables is approximated.

The Reproducing Kernel Hilbert Space (RKHS) and Wiener chaos theories reveal as useful tools in functional regression in a Hilbert-valued context (see, for example, Angelini, De Canditiis and Leblanc, 2003; Bosq, 2000; Guo, 2002; Ivanov and Leonenko, 2004; Ivanov *et al.*, 2013; Kadri *et al.*, 2010; Mendelson, 2002, among others). Extensions of the functional linear regression modeling involving varying coefficients are given, for instance, in Sentürk and Müller (2010) and Wu, Fan and Müller (2010).

In the present paper, the spatial autoregressive framework, introduced in Ruiz-Medina (2011), is combined with a Hilbert-valued spatially heterogeneous trend, fitted by functional linear multiple regression with varying regression operators, to model spatial functional dynamic regressions. The estimation method proposed here is based on the functional state equation satisfied by the spatial Hilbert-valued process. This process displays a spatially heterogeneous functional trend, given in terms of a fixed vector of functional covariables, and a spatial functional

autoregressive dynamics, induced by the SARH(1) process which models the functional error term. The RKHS theory is considered in the regularization of the functional estimation problem associated with the regression and SARH(1) operators. Specifically, orthonormal bases of the RKHS of the Hilbert-valued SARH(1) error process are considered to rectify the ill-posed nature of the problem. Alternatively, dual Riesz bases diagonalizing the SARH(1) cross-covariance operators are also investigated. A stable inversion of the functional equation systems satisfied by the regression and SARH(1) operators is then performed. In practice, the eigenvector and eigenvalues systems involved in the spectral decomposition of the empirical auto-covariance operator of the response, in the multiple regression model, and of the auto-covariance operator and cross-covariance operators of the SARH(1) process are used.

In an applied context, empirical research on the capital structure of companies has an emphasis on the examination of factors that influence the financing decisions in firms. Earlier research established a significant connection between financing decisions and the debt level of companies in the different theoretical and empirical studies carried out (see, for example, Bhaird and Lucey, 2010; Chittenden, Hall and Hutchinson, 1996; Degryse, De Goeij and Kappert, 2012; Heyman, Deloof and Ooghe, 2008; Michaelas, Chittenden and Poutzioris, 1999; Sogorb-Mira, 2005; Van der Wijst and Thurik, 1993). More recently, many works paid attention to the geographical region and/or industry sectors as factors of influence in financing decisions. If these factors are confirmed, the capital structure or its firm factor determinants will differ across regions and/or sectors. These studies used panel data models mainly because information is available from many firms for several years. A variety of approaches within the panel data methodology to examine regional or industry effects have been developed. In some works, these factors are introduced by dummy variables (see, for example, Booth *et al.*, 2001; Degryse, De Goeij and Kappert, 2012; Michaelas, Chittenden and Poutzioris, 1999; Rocca, Rocca and Cariola, 2010; Utrero-González, 2007). The dummy variable approach suggests that regional or sectoral impacts are unrelated to

the independent factors. Some other studies apply regression with varying coefficients, and test the stability of parameter estimates across regions or sectors (Bhaird and Lucey, 2010; Booth *et al.*, 2001; Degryse, De Goeij and Kappert, 2012; De Jong, Kabir and Nguyen, 2008; Psillaki and Daskalakis, 2009).

The Hilbert-valued model proposed here is applied to the above described financial context. Specifically, to represent the spatial heterogeneous functional linear relationship between indebtedness curves and company factors, continuously registered through time, a Hilbert-valued spatially heterogeneous multiple regression model is fitted to estimate the trend. Indeed, this heterogeneity is induced by the industry sector and the geographical region, since functional data are grouped at the nodes of a spatial abstract regular grid, constructed from the combination of the categories displayed by these two variables (see, for example, Case, Rosen and Hines, 1993; Conley and Ligon, 2002; Hautsch and Klotz, 1999). A spatial unilateral dynamics of order one is assumed for the functional error term, extending the classical longitudinal studies usually developed from firm panel data, where spatial correlations are ignored. In the real-data example analyzed, 638 Spanish companies are considered; these companies belong to 9 industry sectors and 17 Spanish communities. A 9×17 spatial regular grid is then fitted. The results of the empirical study developed show a good performance of the estimation methodology proposed, supporting the spatial heterogeneous assumption and the SARH(1) dynamics.

The outline of the paper is the following. Section A3.2 provides some key definitions and results related to SARH(1) modeling. In Section A3.3, different numerical projection methods are described for implementation of the SARH(1) prediction. Section A3.4 introduces a new class of spatial functional multiple regression models with varying regression operators, and its moment-based projection estimation. In Section A3.5, a Spanish firm panel data set, constituted by 638 companies from 9 industry sectors and 17 geographical regions, is analyzed during the period 1999-2007. A cross-validation study is developed for assessing model fitting. Conclusions

and final remarks are given in Section A3.6.

A3.2 Preliminaries

In this section, we briefly review the SARH(1) process framework, introduced in Ruiz-Medina (2011, 2012a). The Markov property of the three points (see Nualart and Sanz, 1979) for two-parameter processes is considered in the definition of the H-valued innovation process involved in the construction of SARH(1) models, as given in the following definition.

Definition 4. *Let* $X = \{X(\mathbf{z}) : \ \mathbf{z} \in \mathbb{R}^2\}$ *be a random field defined on a probability space* $(\Omega, \mathcal{A}, P)$. *Random field* X *is said to satisfy the Markov property of the three points if*

$$P\left(X(\mathbf{z})/\mathcal{G}_{\mathbf{y}}\right) = P\left(X(\mathbf{z})/X(y_1, z_2), X(z_1, y_2), X(y_1, y_2)\right), \tag{A3.1}$$

for all $\mathbf{y} = (y_1, y_2)$ *such that* $y_1 \leq z_1$ *and* $y_2 \leq z_2$, *i.e.,* $\mathbf{y} \leq \mathbf{z}$, *where*

$$\mathcal{G}_{\mathbf{y}} = \sigma\left(X(\mathbf{x}),\ \mathbf{x} \leq \mathbf{y}\right).$$

Note that in the above definition the following partial ordering of $\mathbb{R}^2$ is considered:

$$(y_1, y_2) \leq (z_1, z_2) \Leftrightarrow y_1 \leq z_1 \text{ and } y_2 \leq z_2.$$

For a given point $\mathbf{z} = (z_1, z_2) \in \mathbb{R}^2$, the associated *future* and *past* are defined as the complementary regions $\mathcal{F}_{\mathbf{z}} = \{\mathbf{z}' \in \mathbb{R}^2;\ \mathbf{z} \leq \mathbf{z}'\}$ and $\mathcal{P}_{\mathbf{z}} = \{\mathbf{z}' \in \mathbb{R}^2;\ z_1' < z_1 \quad \text{or} \quad z_2' < z_2\}$ respectively associated with the point $\mathbf{z} \in \mathbb{R}^2$.

In the Gaussian case, the Markov property in Definition 4 is equivalent to the existence of real numbers α_1, α_2 and α_3 such that, for all $\mathbf{y} \leq \mathbf{z}$,

$$E\left[X(\mathbf{z})/\mathcal{G}_{\mathbf{y}}\right] = \alpha_1 X(y_1, z_2) + \alpha_2 X(z_1, y_2) + \alpha_3 X(y_1, y_2). \tag{A3.2}$$

The following concept of spatial or two-parameter martingale difference sequence $\{X_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ is considered.

Definition 5. *Let* $\{X_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *and* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *be two-parameter stochastic sequences. Then,* $\{X_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *is a spatial or two-parameter martingale difference sequence with respect to the sequence* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$ *if the conditional expectation*

$$E^{\mathcal{B}_{\mathcal{P}_{(n,m)}}}(X_{n,m}) = 0,$$

where $\mathcal{P}_{(n,m)}$ *denotes the* past *region of point* (n,m)*, and*

$$\mathcal{B}_{\mathcal{P}_{(n,m)}} = \sigma\left(Z_{i,j},\ (i,j) \in \mathcal{P}_{(m,n)}\right) = \sigma\left(Z_{i,j},\ i < n \quad \text{or} \quad j < m\right)$$

denotes the $\sigma-$*algebra generated by the past values of the stochastic sequence* $\{Z_{i,j},\ (i,j) \in \mathbb{Z}^2\}$.

The spatial functional innovation process of SARH(1) processes (see Ruiz-Medina, 2011) is assumed to be a two-parameter martingale difference. Specifically, the following definition provides the concept of spatial autoregressive Hilbertian processes adopted in Ruiz-Medina (2011):

Definition 6. *(Ruiz-Medina (2011)) A spatial functional process*

$$\mathbf{Y}_{SARH} = \{Y_{ij},\ (i,j) \in \mathbb{Z}^2\},$$

with values in a separable Hilbert space $\mathcal{H}$*, is said to be a unilateral SARH(1) process if it is stationary and satisfies the following equation:*

$$Y_{i,j} = R + L_1(Y_{i-1,j}) + L_2(Y_{i,j-1}) + L_3(Y_{i-1,j-1}) + \epsilon_{i,j}, \tag{A3.3}$$

where $R \in \mathcal{H}$ *and, for* $i = 1,2,3$, $L_i \in \mathcal{L}(\mathcal{H})$*, the space of linear bounded operators on* $\mathcal{H}$*. The spatial functional innovation process satisfies the two-parameter matingale difference property given in Definition 5. Moreover,* $E\|\epsilon_{i,j}\|^2_{\mathcal{H}} = \sigma^2$*, and* $E[\epsilon_{i,j} \otimes \epsilon_{i,j}] = C_{\epsilon_{0,0}}$*, that is, it has finite functional variance, and it is weak-sense stationary in space.*

Let us now introduce the conditions that ensure the existence and uniqueness of a spatial stationary functional solution to the SARH(1) equation admitting a SMAH(∞) representation (see Ruiz-Medina, 2011). Specifically, in Ruiz-Medina (2011), operators L_i, $i = 1, 2, 3$, are assumed to belong to the class of compact possibly non-symmetric operators, admitting the following spectral decomposition, in terms of the eigenvalues $\{\lambda_k(L_i),\ k \in \mathbb{N}\}$, and the biorthogonal systems of left, $\{\psi_k,\ k \in \mathbb{N}\}$, and right, $\{\phi_k,\ k \in \mathbb{N}\}$, eigenvectors,

$$\begin{aligned} L_i(\psi_k) &= \lambda_k(L_i)\psi_k \\ L_i^*(\phi_k) &= \lambda_k(L_i)\phi_k \\ \langle \phi_k, \psi_l \rangle_{\mathcal{H}} &= \delta_{k,l}, \quad k, l \in \mathbb{N}, \end{aligned} \tag{A3.4}$$

where L_i^* denotes the adjoint of L_i, and $\delta_{k,l}$ the Kronecker-delta function. Under (A3.4), the following spectral kernel-based representation of L_i, for $i = 1, 2, 3$, is obtained:

$$L_i(g)(f) = \sum_{k \in \mathbb{N}} \lambda_k(L_i)\psi_k(f)\phi_k(g), \quad \forall f, g \in \mathcal{H}. \tag{A3.5}$$

The conditions needed for the SMAH(∞) representation of process $\mathbf{Y}_{SARH}$ in equation (A3.3) are established in the following result.

Proposition 3. *(Ruiz-Medina (2011)) Assume that (A3.4) holds. For each $k \in \mathbb{N}$, none of the roots of $\Phi(z_1, z_2) = 0 = 1 - \lambda_k(L_1)z_1 - \lambda_k(L_2)z_2 - \lambda_k(L_3)z_1z_2$ lie within the closed unit polydisc ($|z_1| \leq 1, |z_2| \leq 1$) if and only if, for each $k \in \mathbb{N}$,*

(i) $|\lambda_k(L_i)| < 1$, *for* $i = 1, 2, 3$,

(ii) $\left(1 + [\lambda_k(L_1)]^2 - [\lambda_k(L_2)]^2 - [\lambda_k(L_3)]^2\right)^2 - 4(\lambda_k(L_1) + \lambda_k(L_2)\lambda_k(L_3))^2 > 0$

(iii) $1 - [\lambda_k(L_2)]^2 > |\lambda_k(L_1) + \lambda_k(L_2)\lambda_k(L_3)|$.

Under (i)-(iii), the solution to equation (A3.3) is stationary.

The SMAH(∞) representation of process $\mathbf{Y}_{SARH}$ is now provided.

Proposition 4. *(Ruiz-Medina (2011)) Let $\mathbf{Y}_{SARH}$ be a SARH(1) process, as given in Definition 6, with L_i, $i = 1, 2, 3$, admitting the spectral decomposition (A3.4). Assume that the conditions considered in Proposition 3 hold. Then, equation (A3.3) admits a unique stationary solution given by*

$$Y_{i,j} = R + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} L_1^k L_2^l L_3^r(\epsilon_{i-k-r,j-l-r}), \tag{A3.6}$$

where L_i, $i = 1, 2, 3$, are defined in equation (A3.3).

A3.3 SARH(1) functional prediction

The second-order spatial functional structure of Y is characterized in terms of the following covariance operators:

$$\begin{aligned} R_{0,0} &= R_{Z_{i,j}Z_{i,j}} = E\left[Z_{i,j} \otimes Z_{i,j}\right] \\ R_{1,0} &= R_{Z_{i+1,j}Z_{i,j}} = E\left[Z_{i+1,j} \otimes Z_{i,j}\right] \\ R_{0,1} &= R_{Z_{i,j+1}Z_{i,j}} = E\left[Z_{i,j+1} \otimes Z_{i,j}\right] \\ R_{1,1} &= R_{Z_{i+1,j+1}Z_{i,j}} = E\left[Z_{i+1,j+1} \otimes Z_{i,j}\right], \end{aligned} \tag{A3.7}$$

with $Z_{i,j} = Y_{i,j} - R$, for all $(i, j) \in \mathbb{Z}^2$.

Replacing in equation (A3.7), $Z_{i+1,j}$, $Z_{i,j+1}$, and $Z_{i+1,j+1}$ by their expressions, according to the functional state equation (A3.3), we obtain the following functional linear system, satisfied by the autocorrelation operators L_i, $i = 1, 2, 3$:

$$\begin{aligned} R_{1,0} &= L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1} \\ R_{0,1} &= L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0} \\ R_{1,1} &= L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0}. \end{aligned} \tag{A3.8}$$

System (A3.8) can be rewritten as a kernel-based equation system, in terms of the kernels defining the spectral decomposition of operators $R_{0,0}$, $R_{1,0}$, $R_{0,1}$ and $R_{1,1}$, from their eigenvectors and eigenvalues. Specifically, since $R_{i,j}$, $i,j \in \{0,1\}$, are Hilbert-Schmidt operators, they admit the following series representation:

$$\begin{aligned}
R_{0,0}(\varphi)(\varsigma) &= \sum_{k=1}^{\infty} \lambda_k(R_{0,0})[\phi_k^{R_{0,0}}(\varphi)\phi_k^{R_{0,0}}(\varsigma)], \quad \varphi,\varsigma \in \mathcal{D}(R_{0,0}) \\
R_{1,0}(\varphi)(\varsigma) &= \sum_{k=1}^{\infty} \lambda_k(R_{1,0})[\phi_k^{R_{1,0}}(\varsigma)\psi_k^{R_{1,0}}(\varphi)], \quad \varphi,\varsigma \in \mathcal{D}(R_{1,0}) \\
R_{0,1}(\varphi)(\varsigma) &= \sum_{k=1}^{\infty} \lambda_k(R_{0,1})[\phi_k^{R_{0,1}}(\varsigma)\psi_k^{R_{0,1}}(\varphi)], \quad \varphi,\varsigma \in \mathcal{D}(R_{0,1}) \\
R_{1,1}(\varphi)(\varsigma) &= \sum_{k=1}^{\infty} \lambda_k(R_{1,1})[\phi_k^{R_{1,1}}(\varsigma)\psi_k^{R_{1,1}}(\varphi)], \quad \varphi,\varsigma \in \mathcal{D}(R_{1,1}),
\end{aligned} \tag{A3.9}$$

where, for $i,j \in \{0,1\}$, with $i \neq j$, or $i = j = 1$, $\{\psi_k^{R_{i,j}},\ k \geq 1\}$ and $\{\phi_k^{R_{i,j}},\ k \geq 1\}$ respectively denote the right and left eigenvector systems of $R_{i,j}$, and, in the symmetric case, i.e., for the case $i = j = 0$, $\{\phi_k^{R_{0,0}},\ k \geq 1\}$ represents the eigenvector system of $R_{0,0}$, since, in this case, $\phi_k = \psi_k$, for $k \geq 1$. Also, $\{\lambda_k(R_{i,j}),\ k \geq 1\}$ denotes the eigenvalue system of operator $R_{i,j}$, for $i,j \in \{0,1\}$. Here, as usual, $\mathcal{D}(A)$, denotes the domain of operator A.

Now, we consider a kernel-based representation of operators L_i, $i = 1,2,3$, in terms of their projections $L_i(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})$, $k,l \geq 1$, into the basis $\{\phi_k^{R_{0,0}} \otimes \phi_l^{R_{0,0}},\ k,l \geq 1\}$, where $f \otimes g$ denotes the Hilbert-Schmidt operator defined by

$$f \otimes g(\varphi) = f \langle g, \varphi \rangle_H, \quad \forall \varphi \in H, \quad f,g \in H.$$

That is,

$$L_i \underset{H}{=} \sum_{k,l} L_i(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})[\phi_k^{R_{0,0}} \otimes \phi_l^{R_{0,0}}]. \tag{A3.10}$$

By substituting the series representations (A3.9) and (A3.10) into the equation system (A3.8),

we obtain, for $\varphi \in \mathcal{D}(R_{0,0}) \cap \mathcal{D}(R_{1,0}) \cap \mathcal{D}(R_{0,1}) \cap \mathcal{D}(R_{1,1})$, the following system of equations:

$$
\begin{aligned}
f_1 &= \sum_{k=1}^{\infty} \left\langle \lambda_k(R_{1,0})[\psi_k^{R_{1,0}} \otimes \phi_k^{R_{1,0}}], \varphi \right\rangle_H \\
&= \sum_{k,l,p} L_1(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{0,0}} \right\rangle_H \phi_p^{R_{0,0}}(\varphi)\lambda_p(R_{0,0}) \\
&+ \sum_{k,l,p} L_2(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{1,1}} \right\rangle_H \lambda_p(R_{1,1})\psi_p^{R_{1,1}}(\varphi) \\
&+ \sum_{k,l,p} L_3(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{0,1}} \right\rangle_H \lambda_p(R_{0,1})\psi_p^{R_{0,1}}(\varphi) \\
f_2 &= \sum_{k=1}^{\infty} \left\langle \lambda_k(R_{0,1})[\psi_k^{R_{0,1}} \otimes \phi_k^{R_{0,1}}], \varphi \right\rangle_H \\
&= \sum_{k,l,p} L_1(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{1,1}} \right\rangle_H \lambda_p(R_{1,1})\psi_p^{R_{1,1}}(\varphi) \\
&+ \sum_{k,l,p} L_2(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{0,0}} \right\rangle_H \lambda_p(R_{0,0})\phi_p^{R_{0,0}}(\varphi) \\
&+ \sum_{k,l,p} L_3(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{1,0}} \right\rangle_H \lambda_p(R_{1,0})\psi_p^{R_{1,0}}(\varphi) \\
f_3 &= \sum_{k=1}^{\infty} \left\langle \lambda_k(R_{1,1})[\psi_k^{R_{1,1}} \otimes \phi_k^{R_{1,1}}], \varphi \right\rangle_H \\
&= \sum_{k,l,p} L_1(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{0,1}} \right\rangle_H \lambda_p(R_{0,1})\psi_p^{R_{0,1}}(\varphi) \\
&+ \sum_{k,l,p} L_2(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{1,0}} \right\rangle_H \lambda_p(R_{1,0})\psi_p^{R_{1,0}}(\varphi) \\
&+ \sum_{k,l,p} L_3(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})\phi_l^{R_{0,0}} \left\langle \phi_k^{R_{0,0}}, \phi_p^{R_{0,0}} \right\rangle_H \lambda_p(R_{0,0})\phi_p^{R_{0,0}}(\varphi), \qquad \text{(A3.11)}
\end{aligned}
$$

where convergence is given in the $\mathcal{H}$-norm, that is, for a sequence $f_n \in \mathcal{H}$, $n \geq 1$, convergence in $\mathcal{H}$-norm to $f \in \mathcal{H}$ means, as usual, that $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$. In the particular case where $\phi_k^{R_{i,j}} = \phi_k$, $k \geq 1$, for $i, j \in \{0, 1\}$, equation (A3.11) can be rewritten as

$$
\begin{aligned}
f_1 &= \sum_{k=1}^{\infty} \lambda_k(R_{1,0})\psi_k^{R_{1,0}}(\varphi)\phi_k = \sum_{k,l} \lambda_k(R_{0,0})L_1(\phi_k)(\phi_l)\phi_k(\varphi)\phi_l \\
&+ \sum_{k,l} \lambda_k(R_{1,1})L_2(\phi_k)(\phi_l)\psi_k^{R_{1,1}}(\varphi)\phi_l + \sum_{k,l} \lambda_k(R_{0,1})L_3(\phi_k)(\phi_l)\psi_k^{R_{0,1}}(\varphi)\phi_l \\
f_2 &= \sum_{k=1}^{\infty} \lambda_k(R_{0,1})\psi_k^{R_{0,1}}(\varphi)\phi_k = \sum_{k,l} \lambda_k(R_{1,1})L_1(\phi_k)(\phi_l)\psi_k^{R_{1,1}}(\varphi)\phi_l \\
&+ \sum_{k,l} \lambda_k(R_{0,0})L_2(\phi_k)(\phi_l)\phi_k(\varphi)\phi_l + \sum_{k,l} \lambda_k(R_{1,0})L_3(\phi_k)(\phi_l)\psi_k^{R_{1,0}}(\varphi)\phi_l \\
f_3 &= \sum_{k=1}^{\infty} \lambda_k(R_{1,1})\psi_k^{R_{1,1}}(\varphi)\phi_k = \sum_{k,l} \lambda_k(R_{0,1})L_1(\phi_k)(\phi_l)\psi_k^{R_{0,1}}(\varphi)\phi_l \\
&+ \sum_{k,l} \lambda_k(R_{1,0})L_2(\phi_k)(\phi_l)\psi_k^{R_{1,0}}(\varphi)\phi_l + \sum_{k,l} \lambda_k(R_{0,0})L_3(\phi_k)(\phi_l)\phi_k(\varphi)\phi_l. \qquad \text{(A3.12)}
\end{aligned}
$$

Multiplying both sides of identity (A3.12) by ϕ_p, and applying the orthogonality of the eigenfunction system $\{\phi_k,\ k \geq 1\}$, we obtain

$$
\begin{aligned}
f_1(\phi_p) &= \lambda_p(R_{1,0})\psi_p^{R_{1,0}}(\varphi) = \sum_{k=1}^{\infty} \lambda_k(R_{0,0})L_1(\phi_k)(\phi_p)\phi_k(\varphi) \\
&+ \sum_{k} \lambda_k(R_{1,1})L_2(\phi_k)(\phi_p)\psi_k^{R_{1,1}}(\varphi) \\
&+ \sum_{k} \lambda_k(R_{0,1})L_3(\phi_k)(\phi_p)\psi_k^{R_{0,1}}(\varphi) \\
f_2(\phi_p) &= \lambda_p(R_{0,1})\psi_p^{R_{0,1}}(\varphi) = \sum_{k=1}^{\infty} \lambda_k(R_{1,1})L_1(\phi_k)(\phi_p)\psi_k^{R_{1,1}}(\varphi) \\
&+ \sum_{k} \lambda_k(R_{0,0})L_2(\phi_k)(\phi_p)\phi_k(\varphi) \\
&+ \sum_{k,l} \lambda_k(R_{1,0})L_3(\phi_k)(\phi_p)\psi_k^{R_{1,0}}(\varphi) \\
f_3(\phi_p) &= \lambda_p(R_{1,1})\psi_p^{R_{1,1}}(\varphi) = \sum_{k=1}^{\infty} \lambda_k(R_{0,1})L_1(\phi_k)(\phi_p)\psi_k^{R_{0,1}}(\varphi) \\
&+ \sum_{k} \lambda_k(R_{1,0})L_2(\phi_k)(\phi_p)\psi_k^{R_{1,0}}(\varphi) \\
&+ \sum_{k} \lambda_k(R_{0,0})L_3(\phi_k)(\phi_p)\phi_k(\varphi). \qquad \text{(A3.13)}
\end{aligned}
$$

In particular, considering in (A3.13), $\varphi = \phi_p$, and applying that the right and left eigenfunction systems are biorthonormal, as well as the common eigenfunction system $\{\phi_k,\ k \geq 1\}$ itself, one can gets, for each $p \geq 1$,

$$\begin{aligned} f_1(\phi_p) &= \lambda_p(R_{1,0}) = \lambda_p(R_{0,0})L_1(\phi_p)(\phi_p) + \lambda_p(R_{1,1})L_2(\phi_p)(\phi_p) + \lambda_p(R_{0,1})L_3(\phi_p)(\phi_p) \\ f_2(\phi_p) &= \lambda_p(R_{0,1}) = \lambda_p(R_{1,1})L_1(\phi_p)(\phi_p) + \lambda_p(R_{0,0})L_2(\phi_p)(\phi_p) + \lambda_p(R_{1,0})L_3(\phi_p)(\phi_p) \\ f_3(\phi_p) &= \lambda_p(R_{1,1}) = \lambda_p(R_{0,1})L_1(\phi_p)(\phi_p) + \lambda_p(R_{1,0})L_2(\phi_p)(\phi_p) + \lambda_p(R_{0,0})L_3(\phi_p)(\phi_p). \end{aligned} \tag{A3.14}$$

Thus, the diagonal elements $L_i(\phi_p)(\phi_p)$, $p \geq 1$, of the infinite-dimensional coordinate matrix of L_i, $i = 1, 2, 3$, with respect to the basis $\{\phi_k,\ k \geq 1\}$ can be computed as the solutions of the scalar linear system (A3.14), for each $p \geq 1$.

Furthermore, the truncation at term M leads to the following finite-dimensional approximation of system (A3.13):

$$\begin{aligned} \mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) &= \mathbf{L}_1\mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi) + \mathbf{L}_2\mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) + \mathbf{L}_3\mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) \\ \mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) &= \mathbf{L}_1\mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) + \mathbf{L}_2\mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi) + \mathbf{L}_3\mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) \\ \mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) &= \mathbf{L}_1\mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) + \mathbf{L}_2\mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) + \mathbf{L}_3\mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi), \end{aligned} \tag{A3.15}$$

where, for $i, j \in \{0, 1\}$, $\mathbf{\Lambda}(R_{i,j})$ denotes the diagonal $M \times M$ matrix with entries the first M eigenvalues of $R_{i,j}$, arranged in decreasing order of magnitude. Moreover, for $i = 1, 2, 3$, the $M \times M$ matrix $\mathbf{L}_i$ contains the coordinates of operator L_i with respect to the basis $\{\phi_k,\ k = 1, \ldots, M\}$. Finally, by $\boldsymbol{\phi}(\varphi)$, the $M \times 1$ vector with entries $\phi_k(\varphi)$, $k = 1, \ldots, M$, is denoted, and, for $i, j \in \{0, 1\}$, by $\boldsymbol{\psi}^{R_{i,j}}(\varphi)$, the $M \times 1$ vector with entries $\psi_k^{R_{i,j}}(\varphi)$, $k = 1, \ldots, M$, is represented.

Thus, for every $\varphi \in \mathcal{D}(R_{0,0}) \cap \mathcal{D}(R_{1,0}) \cap \mathcal{D}(R_{0,1}) \cap \mathcal{D}(R_{1,1})$,

$$\begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \mathbf{L}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi) & \mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) & \mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) \\ \mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) & \mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi) & \mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) \\ \mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) & \mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) & \mathbf{\Lambda}(R_{0,0})\boldsymbol{\phi}(\varphi) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Lambda}(R_{1,0})\boldsymbol{\psi}^{R_{1,0}}(\varphi) \\ \mathbf{\Lambda}(R_{0,1})\boldsymbol{\psi}^{R_{0,1}}(\varphi) \\ \mathbf{\Lambda}(R_{1,1})\boldsymbol{\psi}^{R_{1,1}}(\varphi) \end{bmatrix}.$$

Additionally, system (A3.8) can be rewritten in terms of the coordinates of all the operators involved with respect to a common orthonormal basis of $\mathcal{H}$, for example, we can consider basis $\{\phi_k^{R_{0,0}},\ k \geq 1\}$ diagonalizing the auto-covariance operator $R_{0,0}$. In this case, replacing in (A3.8), $R_{i,j}$ by $R_{i,j}(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})$, for $k, l \geq 1$, and $i, j \in \{0, 1\}$, and considering a fixed truncation level M, we have

$$\begin{bmatrix} \Phi_M^* L_1 \Phi_M \\ \Phi_M^* L_2 \Phi_M \\ \Phi_M^* L_3 \Phi_M \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}(R_{0,0}) & \mathbf{C}(R_{1,1}) & \mathbf{C}(R_{0,1}) \\ \mathbf{C}(R_{1,1}) & \mathbf{\Lambda}(R_{0,0}) & \mathbf{C}(R_{1,0}) \\ \mathbf{C}(R_{0,1}) & \mathbf{C}(R_{1,0}) & \mathbf{\Lambda}(R_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}(R_{1,0}) \\ \mathbf{C}(R_{0,1}) \\ \mathbf{C}(R_{1,1}) \end{bmatrix}, \tag{A3.16}$$

where $\mathbf{C}(R_{i,j})$ denotes the matrix of coordinates of $R_{i,j}$ with respect to the truncated basis $\{\phi_k^{R_{0,0}},\ k = 1, \dots M\}$, that is, the matrix with entries $R_{i,j}(\phi_k^{R_{0,0}})(\phi_l^{R_{0,0}})$, for $k, l \in \{1, \dots M\}$. Here, for $i = 1, 2, 3$, $\Phi_M^* L_i \Phi_M$ denotes the $M \times M$ matrix with elements the projections of L_i with respect to the basis $\{\phi_k^{R_{0,0}} \otimes \phi_k^{R_{0,0}},\ k = 1, \dots, M\}$. Note that Φ_M^* represents the projection operator into the basis $\{\phi_k^{R_{0,0}},\ k = 1, \dots, M\}$, and Φ_M its adjoint.

In practice, model fitting is performed by solving the functional linear equation system (A3.16) in terms of the empirical covariance operators

$$\begin{aligned} \hat{R}_{0,0} &= \frac{1}{KN} \sum_{i=1}^{K} \sum_{j=1}^{N} Z_{i,j} \otimes Z_{i,j} \\ \hat{R}_{1,0} &= \frac{1}{(K-1)N} \sum_{i=1}^{K-1} \sum_{j=1}^{N} Z_{i+1,j} \otimes Z_{i,j} \\ \hat{R}_{0,1} &= \frac{1}{K(N-1)} \sum_{i=1}^{K} \sum_{j=1}^{N-1} Z_{i,j+1} \otimes Z_{i,j} \\ \hat{R}_{1,1} &= \frac{1}{(K-1)(N-1)} \sum_{i=1}^{K-1} \sum_{j=1}^{N-1} Z_{i+1,j+1} \otimes Z_{i,j}, \end{aligned} \tag{A3.17}$$

where $\{Z_{i,j},\ i = 1,\ldots,K,\ j = 1,\ldots,N\}$ denotes the detrended functional sample located at the spatial nodes of the regular grid considered. Specifically, the empirical eigenvectors $\{\widehat{\phi}_k^{\widehat{R}_{0,0}},\ k \geq 1\}$ and eigenvalues $\{\lambda_k(\widehat{R}_{0,0}),\ k \geq 1\}$ of $\widehat{R}_{0,0}$ are considered, providing its spectral decomposition

$$\widehat{R}_{0,0} = [\widehat{\Phi}^{\widehat{R}_{0,0}}]^* \mathbf{\Lambda}(\widehat{R}_{0,0}) \widehat{\Phi}^{\widehat{R}_{0,0}}.$$

Hence, for a truncation level M, equation (A3.16) is approximated by

$$\begin{bmatrix} \widehat{\Phi^*}_M \widehat{L}_1 \widehat{\Phi}_M \\ \widehat{\Phi}^*_M \widehat{L}_2 \widehat{\Phi}_M \\ \widehat{\Phi}^*_M \widehat{L}_3 \widehat{\Phi}_M \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}(\widehat{R}_{0,0}) & \mathbf{C}(\widehat{R}_{1,1}) & \mathbf{C}(\widehat{R}_{0,1}) \\ \mathbf{C}(\widehat{R}_{1,1}) & \mathbf{\Lambda}(\widehat{R}_{0,0}) & \mathbf{C}(\widehat{R}_{1,0}) \\ \mathbf{C}(\widehat{R}_{0,1}) & \mathbf{C}(\widehat{R}_{1,0}) & \mathbf{\Lambda}(\widehat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}(\widehat{R}_{1,0}) \\ \mathbf{C}(\widehat{R}_{0,1}) \\ \mathbf{C}(\widehat{R}_{1,1}) \end{bmatrix}, \tag{A3.18}$$

where $\mathbf{C}(\widehat{R}_{i,j})$ denotes the matrix of coordinates of $\widehat{R}_{i,j}$ with respect to the truncated basis $\{\phi_k^{\widehat{R}_{0,0}},\ k = 1,\ldots M\}$, that is, the matrix with entries $\widehat{R}_{i,j}(\phi_k^{\widehat{R}_{0,0}})(\phi_l^{\widehat{R}_{0,0}})$, for $k, l \in \{1,\ldots M\}$.

After projection, the SARH(1) plug-in predictor is then computed as

$$\begin{aligned} E\left[Z_{i,j} \mid Z_{k,l},\ k < i,\ \text{or } l < j,\ \text{or } k < i \text{ and } l < j\right] &= L_1 Z_{i-1,j} + L_2 Z_{i,j-1} + L_3 Z_{i-1,j-1} \\ &\simeq \widehat{L_1} Z_{i-1,j} + \widehat{L_2} Z_{i,j-1} + \widehat{L_3} Z_{i-1,j-1}, \end{aligned} \tag{A3.19}$$

where $\widehat{L}_k$, $k = 1, 2, 3$, denote the solution to equation system (A3.18).

A3.4 Functional regression with spatially varying regression operators

A large number of empirical studies supported the fact that the firm capital structure is heterogeneous through different sectors and geographical regions. This key aspect provides the main motivation of this section, where a spatial heterogeneous functional multiple regression model is introduced.

Definition 7. *Let us consider $Y \in \mathcal{H}$ to be the functional response of the following multiple regression model:*

$$Y(s) = \boldsymbol{\mathcal{L}}(s)\mathbf{X} + \varepsilon(s), \quad s \in \mathcal{S}, \tag{A3.20}$$

where, for every $s \in \mathcal{S}$, $Y(s), \varepsilon(s) \in \mathcal{H}$, and $\boldsymbol{\mathcal{L}}(s) = (\mathcal{L}_1(s), \ldots, \mathcal{L}_q(s))$ is a $1 \times q$ vector of unknown spatially dependent linear bounded operators on $\mathcal{H}$. Here, $\mathbf{X} = (X_1, \ldots, X_q)^T \in \mathcal{H}^q$ represents a fixed vector of functional regressors. The spatial functional error process ε is assumed to satisfy

$$\varepsilon_{s_1,s_2} = L_1\varepsilon_{s_1-1,s_2} + L_2\varepsilon_{s_1,s_2-1} + L_3\varepsilon_{s_1-1,s_2-1} + \epsilon,$$

where L_k, $k = 1, 2, 3$, belong to $\mathcal{L}(\mathcal{H})$, and ϵ is Hilbertian white noise in the strong sense, uncorrelated with the random initial values $\varepsilon_{1,0}$, $\varepsilon_{0,1}$ and $\varepsilon_{0,0}$.

An abstract spatial regular grid will be constructed in the next section, where rows are defined by industry sector categories and columns are given by Spanish communities. Model (A3.20) will then be formulated as

$$Z(i,j) = \mathcal{L}_1(i,j)X_1 + \ldots + \mathcal{L}_q(i,j)X_q + \varepsilon(i,j), \tag{A3.21}$$

for $i = 1, \ldots, K$, and $j = 1, \ldots, N$. Define now, for $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$,

$$\begin{aligned}
\widehat{R}_0(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} Z(i,j,h) \otimes Z(i,j,h) \\
\widehat{R}_l(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} Z(i,j,h) \otimes X_{l,h}, \quad l = 1, \ldots, q \\
\widehat{R}_{m,n} &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} X_{m,h} \otimes X_{n,h}, \quad m, n = 1, \ldots, q,
\end{aligned} \tag{A3.22}$$

the empirical covariance operators, where $N(i,j)$ denotes the functional sample size at node (i,j), for $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$. From equation (A3.21), the parameter vector $\boldsymbol{\mathcal{L}}(i,j) =$

$(\mathcal{L}_1(i,j),\ldots,\mathcal{L}_q(i,j))$ can be estimated by solving the following functional equation system

$$\begin{aligned}
\widehat{R}_1(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{11} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{12} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{1q},\\
\widehat{R}_2(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{21} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{22} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{2q},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\widehat{R}_q(i,j) &= \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{q1} + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{q2} + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{qq},
\end{aligned} \tag{A3.23}$$

for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$. Numerical projection methods are applied in the resolution of (A3.23). Specifically, for a given truncation level $M(i,j)$, for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, operators $L_k(i,j)$, $k = 1,\ldots,q$, $\widehat{R}_k(i,j)$, $k = 1,\ldots,q$, $\widehat{R}_0(i,j)$, and $\widehat{R}_{m,n}$, $m,n = 1,\ldots,q$, in (A3.23) are replaced by their projections into the basis

$$\left\{\widehat{\phi}_l^{\widehat{R}_0(i,j)} \otimes \widehat{\phi}_m^{\widehat{R}_0(i,j)},\ l,m = 1,\ldots M(i,j)\right\},$$

leading to the following finite-dimensional approximation of (A3.23):

$$\begin{bmatrix} \Phi_M^*(i,j)L_1(i,j)\Phi_M(i,j) \\ \ldots \\ \Phi_M^*(i,j)L_q(i,j)\Phi_M(i,j) \end{bmatrix} = \begin{bmatrix} \mathbf{C}\left(\hat{R}_{1,1}\right) & \mathbf{C}\left(\hat{R}_{1,2}\right) & \ldots & \mathbf{C}\left(\hat{R}_{1,q}\right) \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{C}\left(\hat{R}_{q,1}\right) & \mathbf{C}\left(\hat{R}_{q,2}\right) & \ldots & \mathbf{C}\left(\hat{R}_{q,q}\right) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}\left(\hat{R}_1(i,j)\right) \\ \ldots \\ \mathbf{C}\left(\hat{R}_q(i,j)\right) \end{bmatrix}.$$

where, as before, for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, $\mathbf{C}(\mathcal{K}(i,j))$ represents the coordinate matrix of operator $\mathcal{K}(i,j)$ with respect to the basis $\left\{\widehat{\phi}_l^{\widehat{R}_0(i,j)} \otimes \widehat{\phi}_m^{\widehat{R}_0(i,j)},\ l,m = 1,\ldots M(i,j)\right\}$.

Remark 8. *Note that the spatially dependent truncation level $M(i,j)$ can be selected by cross-validation at each spatial node (i,j), $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, of the regular grid considered, according to the sample size at such a node. However, for simplification, an optimal minimum homogeneous $\mathcal{M}$ truncation level, common to all the spatial nodes of the regular grid, can also be fitted, as showed in the validation results derived in the next section in relation to the real data example analyzed. Several model selection problems arise in this spatially heterogeneous projection context, especially in relation to the choice of a suitable orthogonal or biorthogonal*

basis, as well as in relation to the functional probabilistic characteristics of the spatial sample (see, for example, Ruiz-Medina and Espejo, 2012, 2013a).

Finally, following the methodology described in the previous section, SARH(1) model fitting is performed from the functional errors

$$\widehat{\varepsilon}(i,j) = Z(i,j) - \widehat{\mathcal{L}}_1(i,j)X_1 - \ldots - \widehat{\mathcal{L}}_q(i,j)X_q, \quad (i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}.$$

A3.5 Real data example

In this section, financing decisions of the firms during the period 1999-2007 are analyzed in a panel constituted by 638 Spanish companies, belonging to 9 different sectors, and located at 17 autonomous Spanish communities. This period corresponds to the economic crisis recuperation of Spain, since its beginning during the years 1993-1994, when the highest number of unemployed people, three and a half million, was reached in Spain. Data have been collected from the SABI (*Sistema de Análisis de Balances Ibéricos*) database. This database contains financial statements of more than a million non-financial Spanish firms.

A3.5.1 Spatial heterogeneous functional multiple regression

The firm factor determinants of leverage considered in the analysis of the financing decisions of the 638 Spanish companies studied are: *Firm size*, which is measured as the log of total assets; *Asset structure*, which consists of the net fixed assets divided by the total assets of the firm; *Profitability*, computed as the ratio between earnings before interest, taxes amortization and depreciation, and the total assets; *Growth*, for which as proxy, we consider the growth of the assets, calculated as the annual change of the total assets of the firm; *Firm risk* given by the *Business risk*, defined as the standard deviation of earnings before interest and taxes over book value of total assets during the sample period; and, finally we consider the *Age*, measured as the logarithm of the number of years that the firm has been operating.

These factors are assumed to be fixed over the companies belonging to the 9 sectors, and the 17 geographical regions analyzed. Leverage is estimated as the quotient between the total debt and the total assets.

An abstract economic 9×17 spatial regular grid is defined according to the 9 categories of industry sectors and 17 categories of geographical regions distinguished in the functional data analyzed. At each node of this spatial regular grid, the empirical covariance operators (A3.22) are computed, and equation system (A3.23) is numerically inverted after projection. A finite-dimensional approximation to the regression operators at such a node, and to the corresponding plug-in estimation of the mean debt curve is then obtained.

A 4-fold cross-validation (4-fold CV) procedure is applied by considering a validation sample constituted by 9 companies, and a learning or training sample consisting of 27 firms randomly selected from each node of the 9×17 spatial regular grid. The average of the L^∞-norms of the validation absolute error curves at such nodes of the spatial regular grid constructed is also displayed in Table A3.1. Figure A3.1 shows the 4-fold CV results for certain nodes, considering a homogeneous truncation order $\mathcal{M} = 4$. From the cross-validation performed, we can conclude that the truncation order $\mathcal{M} = 4$ provides the threshold to obtain acceptable projection estimation results.

To illustrate the increased accuracy when the truncation order is higher, in Table A3.2, the truncation order $\mathcal{M} = 9$ is displayed. Specifically, the mean L^∞-norms of the absolute error curves obtained after fitting the projected spatial functional regression model with this higher truncation order are showed in this table. It can be seen that the order of magnitude of these L^∞-norms ranges between 10^{-14} and 10^{-17}, except at the nodes $(3, 7)$, $(5, 7)$ and $(5, 11)$, which have been eliminated in the graphical representation at Figure A3.2, achieved in terms of the following spatial ordering

$$(i, j) \leq (k, l) \quad \begin{cases} \text{if} \quad i = k, \quad \text{then} \quad j \leq l \\ \text{otherwise} \quad i \leq k \end{cases}$$

Figure A3.1: After applying 4-fold CV, from top to bottom, four mean absolute error curves (two at the left and two at the right), computed at the validation firms randomly selected from the non-empty nodes $(3, 13); (4, 7); (5, 12); (7, 11)$ and $(8, 7)$, are showed

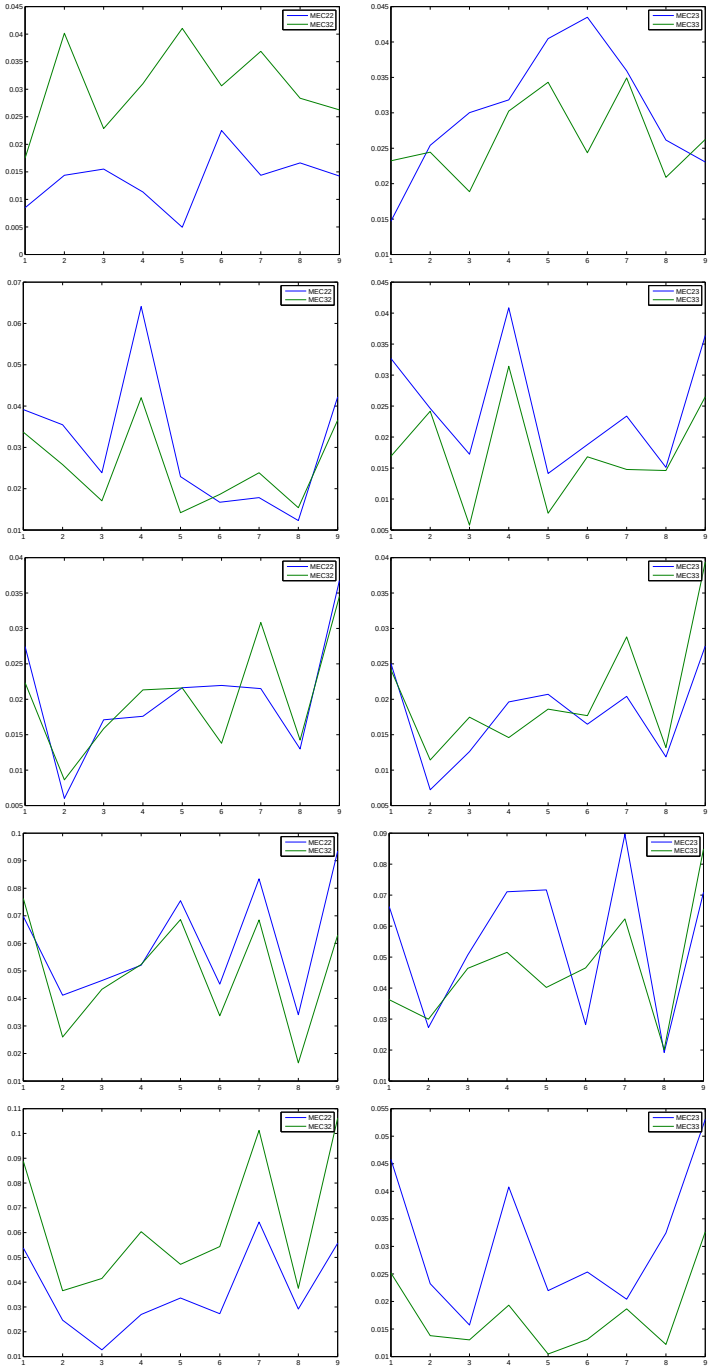
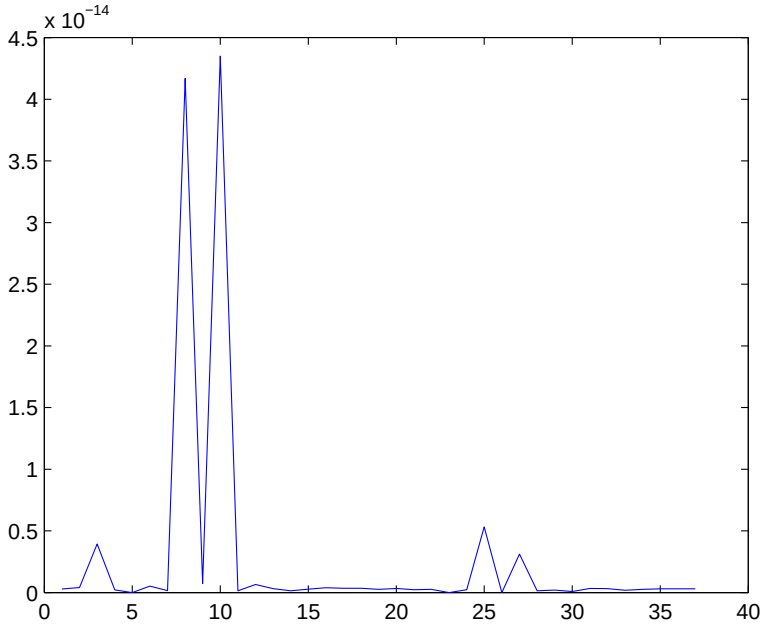


Table A3.1: *Averaged L^∞-norms of the 4-fold CV mean absolute error curves for homogeneous truncation order $\mathcal{M} = 4$*

Nodes	$(3,4)$	$(3,6)$	$(3,7)$	$(3,8)$	$(3,11)$
L^∞-norm	0.1034	0.0492	0.0315	0.0572	0.0805
Nodes	$(3,12)$	$(3,13)$	$(3,15)$	$(4,7)$	$(4,8)$
L^∞-norm	0.0290	0.0396	0.0579	0.0446	0.0242
Nodes	$(4,11)$	$(5,7)$	$(5,11)$	$(5,12)$	$(5,13)$
L^∞-norm	0.0435	0.0654	0.0225	0.0345	0.0609
Nodes	$(5,14)$	$(5,15)$	$(7,7)$	$(7,11)$	$(8,7)$
L^∞-norm	0.0400	0.0586	0.0657	0.0861	0.0443

Figure A3.2: Averaged L^∞-norms of the functional multiple regression absolute error curves, after eliminating $(3,7)$, $(5,7)$, $(5,11)$ nodes



Note that the considered spatial functional multiple regression model fitting assumes heterogeneous spatial behavior between companies clustered at different spatial nodes, and homogene-

Table A3.2: Averaged L^∞-*norms of the functional multiple regression absolute error curves for homogeneous truncation order* $\mathcal{M} = 9$

Nodes	$(3,1)$	$(3,2)$	$(3,4)$	$(3,6)$	$(3,7)$	$(3,8)$
L^∞-norm	$2.82 \cdot 10^{-16}$	$1.73 \cdot 10^{-16}$	$4.08 \cdot 10^{-15}$	$2.66 \cdot 10^{-16}$	$1.3 \cdot 10^{-2}$	$4.12 \cdot 10^{-16}$
Nodes	$(3,9)$	$(3,11)$	$(3,12)$	$(3,13)$	$(3,14)$	$(3,15)$
L^∞-norm	$1.84 \cdot 10^{-16}$	$3.33 \cdot 10^{-14}$	$6.39 \cdot 10^{-16}$	$4.33 \cdot 10^{-14}$	$3.13 \cdot 10^{-16}$	$5.59 \cdot 10^{-16}$
Nodes	$(3,16)$	$(4,1)$	$(4,7)$	$(4,8)$	$(4,11)$	$(4,13)$
L^∞-norm	$2.51 \cdot 10^{-16}$	$1.34 \cdot 10^{-16}$	$2.60 \cdot 10^{-16}$	$3.83 \cdot 10^{-16}$	$2.88 \cdot 10^{-16}$	$3.67 \cdot 10^{-16}$
Nodes	$(4,15)$	$(5,1)$	$(5,4)$	$(5,6)$	$(5,7)$	$(5,8)$
L^∞-norm	$2.34 \cdot 10^{-16}$	$1.62 \cdot 10^{-16}$	$2.02 \cdot 10^{-16}$	$2.30 \cdot 10^{-16}$	$3.14 \cdot 10^{-2}$	$1.93 \cdot 10^{-16}$
Nodes	$(5,10)$	$(5,11)$	$(5,12)$	$(5,13)$	$(5,14)$	$(5,15)$
L^∞-norm	$1.86 \cdot 10^{-16}$	$1.4 \cdot 10^{-2}$	$2.16 \cdot 10^{-16}$	$8.04 \cdot 10^{-15}$	$1.69 \cdot 10^{-16}$	$2.80 \cdot 10^{-15}$
Nodes	$(5,16)$	$(5,17)$	$(6,14)$	$(7,7)$	$(7,11)$	$(8,7)$
L^∞-norm	$1.25 \cdot 10^{-16}$	$2.11 \cdot 10^{-16}$	$9.82 \cdot 10^{-17}$	$3.37 \cdot 10^{-16}$	$3.14 \cdot 10^{-16}$	$3.13 \cdot 10^{-16}$
Nodes	$(8,11)$	$(9,7)$	$(9,11)$			
L^∞-norm	$1.48 \cdot 10^{-16}$	$2.26 \cdot 10^{-16}$	$2.94 \cdot 10^{-16}$			

ity inside the company cluster at every node. It seems that the spatial homogeneity assumption inside the company clusters corresponding to the nodes $(3,7)$, $(5,7)$, and $(5,11)$ does not hold. Thus, a more careful financial analysis of the capital structure of the companies clustered in sector 3 and geographical region 7, as well as in sector 5 and geographical regions 7 and 11 must be performed for assessment of the heterogeneity/homogeneity, according to the particular characteristics of the firms grouped at such nodes. In particular, we note that regions 7 and 11 provide a 24% and 17% of the total number of companies analyzed in the sample. These areas have the highest level of industrial development in Spain. Thus, the diversity of firms is greater within these geographical regions. Moreover, sector 5, related to *Wholesale and retail trade*, represents a 50% of the companies at the geographical region 11, and a 40% of the companies at the geographical region 7. Sector 3 corresponds to *Manufacturing*, and represents a 32% of the firms at region 7. An important number of companies are then clustered at nodes $(3,7), (5,7)$ and $(5,11)$, requiring to increase resolution in the analysis.

A3.5.2 SARH(1) model fitting

This section provides the SARH(1) model fitting, following the steps described in Section A3.2. The mean debt curves are computed at each node from the results derived in the previous section. The empirical covariance operators (A3.17) of the estimated functional error term ε are then obtained. The estimates of the functional parameter L_i, $i = 1, 2, 3$, is derived from projection into the truncated eigenvector system of the empirical autocovariance operator of ε, considering truncation order $\mathcal{M} = 9$. After performing numerical inversion, the finite dimensional approximation of $\widehat{L}_i$, $i = 1, 2, 3$, is obtained, as well as the corresponding SARH(1) plug-in spatial prediction (A3.19) of mean debt curves. The L^∞-norms of the associated absolute error curves are displayed in Tables A3.3 and A3.4, for all the nodes of the spatial regular grid. Figure A3.3 also shows the computed L^∞-norms of the absolute error curves at every node of the spatial regular grid.

The order of magnitude of the computed L^∞-norms of the absolute error curves supports the spatial autoregressive dynamics of order one in the propagation of the functional errors, associated with the mean debt curve functional heterogeneous regression estimation previously performed at each node of the economic abstract regular grid.

A3.6 Conclusion

The Hilbert-valued modeling framework adopted in this paper allows to represent heterogeneous and Markovian behaviors in space in a functional context. Inference on two-parameter diffusion processes, in the class introduced in Nualart and Sanz-Solé (1979), can be achieved from the SARH(1) framework (see Ruiz-Medina, 2011, Section 3.2). The new modeling introduced allows the approximation of the solution to stochastic partial differential equations with spatially varying drift.

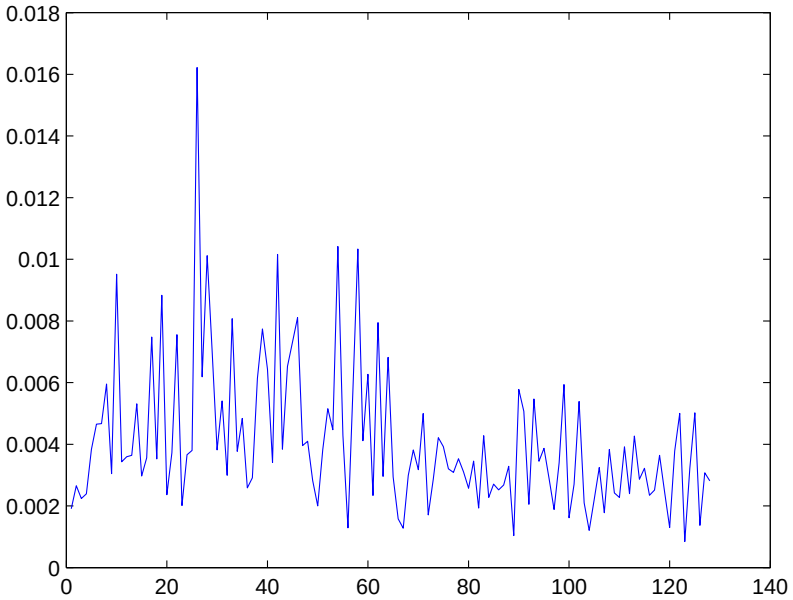
Table A3.3: *L^∞-norms of the SARH(1) absolute error curves at the nodes of the* 9×17 *regular grid*

Nodes	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$	$(2,7)$	$(2,8)$	$(2,9)$
L^∞-norm	0.0019	0.0027	0.0022	0.0024	0.0038	0.0047	0.0047	0.0060
Nodes	$(2,10)$	$(2,11)$	$(2,12)$	$(2,13)$	$(2,14)$	$(2,15)$	$(2,16)$	$(2,17)$
L^∞-norm	0.0030	0.0095	0.0034	0.0036	0.0036	0.0053	0.0030	0.0036
Nodes	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$	$(3,7)$	$(3,8)$	$(3,9)$
L^∞-norm	0.0075	0.0035	0.0088	0.0024	0.0037	0.0076	0.0020	0.0037
Nodes	$(3,10)$	$(3,11)$	$(3,12)$	$(3,13)$	$(3,14)$	$(3,15)$	$(3,16)$	$(3,17)$
L^∞-norm	0.0038	0.0162	0.0062	0.0101	0.0071	0.0038	0.0054	0.0030
Nodes	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$	$(4,7)$	$(4,8)$	$(4,9)$
L^∞-norm	0.0081	0.0038	0.0048	0.0026	0.0029	0.0061	0.0077	0.0064
Nodes	$(4,10)$	$(4,11)$	$(4,12)$	$(4,13)$	$(4,14)$	$(4,15)$	$(4,16)$	$(4,17)$
L^∞-norm	0.0034	0.0102	0.0038	0.0065	0.0073	0.0081	0.0040	0.0041
Nodes	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$	$(5,7)$	$(5,8)$	$(5,9)$
L^∞-norm	0.0028	0.0020	0.0038	0.0052	0.0045	0.0104	0.0043	0.0013
Nodes	$(5,10)$	$(5,11)$	$(5,12)$	$(5,13)$	$(5,14)$	$(5,15)$	$(5,16)$	$(5,17)$
L^∞-norm	0.0057	0.0103	0.0041	0.0063	0.0023	0.0079	0.0030	0.0068
Nodes	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$	$(6,7)$	$(6,8)$	$(6,9)$
L^∞-norm	0.0029	0.0016	0.0013	0.0030	0.0038	0.0032	0.0050	0.0017
Nodes	$(6,10)$	$(6,11)$	$(6,12)$	$(6,13)$	$(6,14)$	$(6,15)$	$(6,16)$	$(6,17)$
L^∞-norm	0.0029	0.0042	0.0039	0.0032	0.0031	0.0035	0.0031	0.0026

Table A3.4: *L^∞-norms of the SARH(1) absolute error curves at the nodes of the* 9×17 *regular grid*

Nodes	$(7,2)$	$(7,3)$	$(7,4)$	$(7,5)$	$(7,6)$	$(7,7)$	$(7,8)$	$(7,9)$
L^∞-norm	0.0035	0.0019	0.0043	0.0023	0.0027	0.0025	0.0027	0.0033
Nodes	$(7,10)$	$(7,11)$	$(7,12)$	$(7,13)$	$(7,14)$	$(7,15)$	$(7,16)$	$(7,17)$
L^∞-norm	0.0010	0.0058	0.0051	0.0021	0.0055	0.0034	0.0039	0.0029
Nodes	$(8,2)$	$(8,3)$	$(8,4)$	$(8,5)$	$(8,6)$	$(8,7)$	$(8,8)$	$(8,9)$
L^∞-norm	0.0019	0.0034	0.0059	0.0016	0.0027	0.0054	0.0021	0.0012
Nodes	$(8,10)$	$(8,11)$	$(8,12)$	$(8,13)$	$(8,14)$	$(8,15)$	$(8,16)$	$(8,17)$
L^∞-norm	0.0022	0.0032	0.0018	0.0038	0.0024	0.0023	0.0039	0.0024
Nodes	$(9,2)$	$(9,3)$	$(9,4)$	$(9,5)$	$(9,6)$	$(9,7)$	$(9,8)$	$(9,9)$
L^∞-norm	0.0043	0.0029	0.0032	0.0023	0.0025	0.0036	0.0025	0.0013
Nodes	$(9,10)$	$(9,11)$	$(9,12)$	$(9,13)$	$(9,14)$	$(9,15)$	$(9,16)$	$(9,17)$
L^∞-norm	0.0038	0.0050	0.0008	0.0032	0.0050	0.0014	0.0031	0.0028

Figure A3.3: L^∞-norms of the SARH(1) absolute error curves



In an applied statistical framework, the proposed estimation method opens new modeling and estimation research lines in the spatial econometric area. Further research must be developed in terms of the formulation of functional statistical hypothesis tests to contrast with the spatial interaction of order one between Hilbert-valued spatial random variables. In relation to the empirical study developed from financial panel data, the model fitting performed supports the non-stability of regression operators through different industry sectors and geographical regions, and the spatial interaction of order one between mean indebtedness curves.

Appendix A4

Spatial functional normal mixed effect approach for curve classification

Ruiz-Medina, M. D., Espejo, R. M. and Romano, E. (2014).

Spatial functional normal mixed effect approach for curve classification.

Advances in Data Analysis and Classification.

DOI : 10.1007/s11634-014-0174-6.

Abstract

A spatial functional formulation of the normal mixed effect model is proposed for the statistical classification of spatially dependent Gaussian curves, in a parametric and state space model frameworks. Their fixed effect parameters are represented in terms of a functional multiple regression model whose regression operators can change in space. Local spatial homogeneity of these operators is measured in terms of their Hilbert-Schmidt distances, leading to the classification of fixed effect curves in different groups. Assuming that the Gaussian random effect curves obey a spatial autoregressive dynamics of order one (SARH(1) dynamics), a second functional classification criterion is proposed in order to detect local spatially homogeneous patterns in the mean quadratic functional variation of Gaussian random effect curve increments. Finally, the two criteria are combined for detection of local spatially homogeneous patterns in the regression operators and in the functional mean quadratic

variation, under a state space approach. For illustration purposes, a real data example in the financial context is analyzed.

A4.1 Introduction

New criteria for classification arise in the context of Functional Statistics (see Ramsay and Silverman, 2005; Ferraty and Vieu, 2006, among others). In particular, these criteria allow the classification of random curves in the absence or in the presence of interactions between different individuals, as well as between different times (see Ferraty and Vieu, 2006; Aach and Church, 2001; Hall, Poskitt and Presnell, 2001; James and Hastie, 2001; Liu and Müller, 2003; Müller and Stadtmüller, 2005; among others).

Filtering methods are applied, for instance, in James and Hastie, (2001) who consider a variant of linear discriminant analysis, in terms of the curve projections assuming a Gaussian distribution with common covariance matrix for all classes. Their classification is derived from the minimization of the distance to the group mean. The likelihood-based approach presented in Hall, Poskitt and and Presnell, (2001) is inspired on quadratic discriminant analysis, since although they propose a fully nonparametric density estimation, in practice, multivariate Gaussian densities are considered. The model-based functional classification procedures proposed in Leng and Müller, (2006) and Rincón and Ruiz-Medina, (2012a) are derived in a generalized linear model framework. Specifically, Functional Principal Component Analysis (FPCA), and local wavelet-vaguelette decomposition are respectively considered for dimension reduction. In Biau, Bunea and Wegkamp, (2003), k-nearest neighbor method is applied to Fourier coefficients. Wavelet bases are selected for projection in Berlinet, Biau and Rouvière, (2008), while in James and Sugar, (2003) spline bases are used in a random effect model context, combining the best properties of filtering and regularization methods. However, the presented method is effective when the observations are sparse, irregularly spaced or occur at different time points for each

subject (see also Abraham *et al.*, 2003, where B-splines bases are previously chosen for projection in the application of k-means-based classification procedure).

Alternatively, the context of statistical learning methods based on kernels (see, for example, Schölkopf and Smola, 2002) has been widely used in the design of functional nonparametric statistical classification procedures. In this framework, the unknown function is estimated, considering its optimal approximation in a functional class given by a RKHS, under some prescribed criterion. Chaos game representation and multifractal analysis can also be considered in the classification of functional protein sequences displaying singular features (see, for example, Yang, *et al.*, 2009a, and Yang, Yu and Anh, 2009b).

Functional features in the data are considered in the design of the classification procedures presented, for example, in Ferraty and Vieu (2003), where kernel estimates of the group membership posterior probability are computed for the curves to be classified. Classification methodologies, based on the notion of depth for curves, are proposed in López-Pintado and Romo, (2006). Density-function-based classification is performed in Nerini and Ghattas, (2007), from functional regression trees (see also Baíllo and Cuevas, 2008, on functional k-nearest neighbor and Li and Yu, 2008, on the use of F-statistics for the application of linear discriminant-based approach to previously selected small intervals). We can also mention the papers by: Tarpey and Kinateder, (2003) who provide k-means algorithm over the probability distributions; Cuesta-Albertos and Fraiman, (2007), that combine impartial trimming with k-means; Chiou and Li, (2007), Chiou and Li, (2008a,b) where the information through the mean and the covariance/correlation functions is considered; Li and Chiou, (2011), where the cluster number selection problem for functional data is addressed. In general, continuous transformations can be considered for an efficient extraction of information, in particular, in relation to some features (see Ferraty and Vieu, 2009; Li and Yu, 2008; López-Pintado and Romo, 2006; Ramsay and Silverman, 2005; Rossi and Villa, 2006; Zhang and Müller, 2011; among others). Recently, the functional data

classification problem is transformed into a classical multivariate data discrimination problem in Alonso, Casado and Romo, (2012). Linear discriminant analysis is then applied to determine the combination of variables and coefficients. These variables are defined as the difference between the respective distances of the derivatives of an observed curve and the sample means of the corresponding derivatives computed from the training samples of two populations. Finally, we refer to the methods based on componentwise classification (see, for example, Delaigle, Hall and Bathia, 2012), functional logit regression (see Aguilera-Morillo *et al.*, 2013; Escabias, Aguilera and Valderrama, 2004, 2007; among others) and functional PLS (see Aguilera *et al.*, 2010; Delaigle and Hall, 2012a,b; Preda and Saporta, 2005a,b).

Spatial Functional Statistics emerges as a new branch of Functional Statistics to dealing with problems involving spatially dependent curves, or in general, functional data in space. The field of *Spatial Functional Statistics* still requires further development. Actually, literature in this framework covers, among others, statistical methodologies related to spatial functional regression (see Guillas and Lai, 2010), Bayesian inference (see Baladandayuthapani, *et al.*, 2008), nonparametric estimation methods (see Basse, Diop and Dabo-Niang, 2008), spatial functional prediction (see Nerini, Monestiez and Manté, 2010, Monestiez and Nerini, 2008, Giraldo, Delicado and Mateu, 2010; Delicado, *et al.*, 2010), spatial functional autoregressive time series models (see Ruiz-Medina, 2011, 2012a; Ruiz-Medina and Espejo, 2012; Ruiz-Medina, Anh, Espejo and Frías, 2013), outlier detection strategies (see Romano, Balzanella and Verde, 2013), non supervised clustering methods (see Romano and Verde, 2011; Romano, Balzanella and Verde, 2010, 2013).

In particular, model-based spatial functional statistical classification still remains as an unexplored area in most of its statistical subfields. In this paper, the spatial functional normal mixed effect approach is adopted for classification of spatially dependent curves. Fixed effect curves are represented in terms of a multiple regression model, where the regression operators can change

in space. The l^2 distance between the pure point spectra of the non-parametric estimates of these operators (see Ruiz-Medina and Espejo, 2012, for derivation of these estimators) allows to detect spatial heterogeneities in the functional linear relationship between the fixed effect curves and the functional regressors. Thus, two fixed-effect curves at different spatial locations are in the same group if the Hilbert-Schmidt distance between the associated vectors of regression operators is sufficiently small according to a prescribed partially data-driven threshold. On the other hand, assuming that the random effect curves display a spatial autoregressive dynamics of order one, the functional variance of their SARH(1) increments is computed for grouping curves with the same mean quadratic spatial local functional variation properties. Note that SARH(1) processes are introduced in Ruiz-Medina, (2011) under stationary assumption in space in order to ensure their invertibility. Summarizing, the proposed functional classification methodology is able to detect the local spatial homogeneity of the regression operators characterizing the fixed effect curves, as well as of the random effect curve increments in the mean-square sense, under a state-space approach.

The outline of the paper is as follows. Section A4.2 provides the spatial functional mixed effect model formulation. The elements involved in the non-parametric multiple functional regression approach adopted for classification of fixed effect curves are introduced in Section A4.3. SARH(1) variogram is approximated in Section A4.4 in the wavelet domain in terms of the empirical wavelet spectra of SARH(1) parameters and the observed random effect curves. Classification of random effect curves is then performed according to the local exponents reflected in this empirical variogram wavelet spectrum. In Section A4.5, the main steps involved in the implementation of the proposed functional classification algorithm in the real-data example analyzed in Section A4.6 are described. Section A4.6 provides the illustration of the classification methodology proposed with a real-data example. Final comments and discussion are conducted in Section A4.7.

A4.2 The spatial functional mixed effect model

Let us consider the following spatial functional normal mixed effect model:

$$\mathbf{Y}(\cdot,\mathbf{x}) = \mathbf{X}_{FE}\boldsymbol{\beta}(\cdot,\mathbf{x}) + \mathbf{X}_{RE}\vec{\nu}(\cdot,\mathbf{x}) + \sigma\vec{\epsilon}(\cdot,\mathbf{x}), \quad \mathbf{x} \in D \subset \mathbb{R}^n, \tag{A4.1}$$

where $\mathbf{X}_{FE}$ and $\mathbf{X}_{RE}$ represent the respective functional fixed and random effect design matrices. For each spatial location $\mathbf{x} \in D$, each component of the functional vectors $\mathbf{Y}(\cdot,\mathbf{x})$, $\boldsymbol{\beta}(\cdot,\mathbf{x})$, $\vec{\nu}(\cdot,\mathbf{x})$ and $\vec{\epsilon}(\cdot,\mathbf{x})$ lies in the real separable Hilbert space H. In a general setting, D could be an open bounded domain in $\mathbb{R}^n$. However, in the subsequent development, we assume that D is a finite set of points in $\mathbb{Z}^2$. Also, in the following, space H is constituted by functions with support contained in the real line, $\mathbf{Y}(\cdot,\mathbf{x})$, $\boldsymbol{\beta}(\cdot,\mathbf{x})$, $\vec{\nu}(\cdot,\mathbf{x})$ and $\vec{\epsilon}(\cdot,\mathbf{x})$ respectively denote the vectors of response, fixed effect, random effect and error curves observed at spatial location $\mathbf{x} \in D$.

In Section A4.6 the space H is constituted by the square integrable functions on the interval $[T_1, T_2]$, with $[T_1, T_2]$ being the time period where the firm panel analyzed is observed. The discrete observation of indebtedness and firm factor curves during the period 1999-2007 is smoothed by applying temporal local polynomial kernel smoothing, based on the Epanechnikov kernel. The firm panel is constituted by 638 Spanish companies, belonging to 4 different industry sectors, and located at 17 autonomous Spanish communities. (Data have been collected from the SABI, *Sistema de Análisis de Balances Ibéricos*, database). Since as commented in the analysis performed we consider D to be a finite set in $\mathbb{Z}^2$, high-dimensional spatial interpolation (see, for example, Stein, 2009) will be applied in Section A4.6 to obtaining the curve data samples (indebtedness and firm factor curves) on a 4×6 spatial regular grid. Here, the latitude and longitude defining the spatial location of the centroid associated with each Spanish community will be taken into account to define the suitable spatial regular grid. Indeed, we consider $\boldsymbol{\beta}$ to be modeled in terms of a multiple regression model with regression operators depending on the spatial location, and $\vec{\nu}$ is assumed to obey a spatial autoregressive equation of order one.

In the real-data example analyzed in Section A4.6, we consider the case where $\mathbf{X}_{FE}$ and $\mathbf{X}_{RE}$ in equation (A4.1) are functional identity matrices on H. This is the reason why in the subsequent development we concentrate on this case. In the general case where $\mathbf{X}_{FE}$ and $\mathbf{X}_{RE}$ are not functional identity matrices, the expressions defining the functional least-squares estimator of $\mathbf{Y}$ involve the matrices $\mathbf{X}_{FE}$ and $\mathbf{X}_{RE}$, since they appear in the corresponding least-squares estimators of $\vec{\beta}$ and $\vec{\nu}$. This functional mixed effect framework has been considered, for example, in Spitzner, Marron and Essick, (2003) for addressing the estimation problem related to the movement curves in human perception in Biomedical applications.

A4.3 Functional regression with spatially varying regression operators

Let us consider a spatial regular rectangular grid with nodes (i,j) in $\mathbb{Z}^2$. Assume that the fixed effect curve parameter vector $\boldsymbol{\beta}$ displays a linear functional relationship with a given H-valued vector $(X_1, \ldots, X_q)$ of random regressors, in terms of a $1 \times q$ vector of unknown spatially dependent linear bounded operators on H, $\vec{\mathcal{L}}(i,j) = (\mathcal{L}_1(i,j), \ldots, \mathcal{L}_q(i,j))$. In practice, a functional error term ε is included in the observation of the H-valued components of $\vec{\beta}(i,j)$ for estimation of $(\mathcal{L}_1(i,j), \ldots, \mathcal{L}_q(i,j))$ (see Ruiz-Medina, Anh, Espejo and Frías, 2013). Thus,

$$Y(i,j,h) = \beta_{OB}(i,j,h) = \mathcal{L}_1(i,j)X_{1,h} + \ldots + \mathcal{L}_q(i,j)X_{q,h} + \varepsilon(i,j,h), \tag{A4.2}$$

where $X_{k,h}, \varepsilon(i,j,h) \in H$ are centered Gaussian H-valued random variables, for $k = 1, \ldots q$, and $h = 1, \ldots, N(i,j)$, with $N(i,j)$ the number of individuals observed at location (i,j), $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$. Specifically, we are considering as $Y(i,j,h) = \beta_{OB}(i,j,h)$ the observed curve at the individual h, $h = 1, \ldots, N(i,j)$, located at (i,j), $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$. Hence, $\varepsilon(i,j,h) = \nu(i,j,h) + \sigma\epsilon(i,j,h)$. The random regressors X_i, $i = 1, \ldots, q$, are independent of ε. Also, the spatial H-valued Gaussian random effect ν is assumed to be independent of the spatial H-valued white noise ϵ, with ϵ being the error term introduced in the $H-$valued normal

mixed effect model (A4.1). Note that the functional regressors $X_1, \ldots, X_q$ are observable, with $X_{k,h}$ denoting the H-value at individual h, for $k = 1, \ldots, q$, with $h = 1, \ldots, N(i,j)$ for each node $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$.

Define now, for $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$,

$$\begin{aligned}
\widehat{R}_0(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} (Y(i,j,h) - \overline{Y}(i,j)) \otimes (Y(i,j,h) - \overline{Y}(i,j)) \\
\widehat{R}_l(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} (Y(i,j,h) - \overline{Y}(i,j)) \otimes (X_{l,h} - \overline{X}_l^{(i,j)}) \quad l = 1, \ldots, q \\
\widehat{R}_{m,n}(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} (X_{m,h} - \overline{X}_l^{(i,j)}) \otimes (X_{n,h} - \overline{X}_n^{(i,j)}), \quad m,n = 1, \ldots, q,
\end{aligned} \tag{A4.3}$$

the empirical covariance operators, where, as before, $N(i,j)$ denotes the functional sample size at node (i,j). Here,

$$\begin{aligned}
\overline{Y}(i,j) &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} Y(i,j,h), \quad (i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\} \\
\overline{X}_l^{(i,j)} &= \frac{1}{N(i,j)} \sum_{h=1}^{N(i,j)} X_{l,h}, \quad (i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\},
\end{aligned} \tag{A4.4}$$

for $l = 1, \ldots, q$. From equation (A4.2), we consider the nonparametric estimator $\vec{\widehat{\mathcal{L}}}(i,j) = (\widehat{\mathcal{L}}_1(i,j), \ldots, \widehat{\mathcal{L}}_q(i,j))$ of $\vec{\mathcal{L}}(i,j) = (\mathcal{L}_1(i,j), \ldots, \mathcal{L}_q(i,j))$ as the solution to the following functional equation system

$$\begin{aligned}
&\widehat{R}_1(i,j) = \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{11}(i,j) + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{12}(i,j) + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{1q}(i,j), \\
&\widehat{R}_2(i,j) = \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{21}(i,j) + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{22}(i,j) + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{2q}(i,j), \\
&\qquad\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\widehat{R}_q(i,j) = \widehat{\mathcal{L}}_1(i,j)\widehat{R}_{q1}(i,j) + \widehat{\mathcal{L}}_2(i,j)\widehat{R}_{q2}(i,j) + \ldots + \widehat{\mathcal{L}}_q(i,j)\widehat{R}_{qq}(i,j),
\end{aligned} \tag{A4.5}$$

for each $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$. Numerical projection methods are applied to solving equation system (A4.5). Specifically, we consider an unconditional orthonormal basis of H, $\{\widehat{\phi}_n,\ n \in \mathbb{N}\}$ (see, for example, Donoho, 1993). Note that, in practice, a finite-dimensional approximation of (A4.5) is obtained by applying the projection operator $\widehat{\mathbf{\Phi}}^*_{M_{i,j}}$ into the finite set $\{\widehat{\phi}_n,\ n = 1,\ldots,M_{ij}\}$, with $M_{i,j}$ denoting the truncation order at location (i,j), $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$. For simplicity, let us establish a common truncation order M at all the nodes of the spatial regular grid, i.e., $M = \max_{(i,j)\in\{1,\ldots,K\}\times\{1,\ldots,N\}} M_{ij}$. Hence, for every $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, $\widehat{\mathbf{\Phi}}^*_{M_{i,j}} = \widehat{\mathbf{\Phi}}^*_M$, and $[\widehat{\mathbf{\Phi}}_M]^*\widehat{\mathbf{\Phi}}_M = I$, by the orthogonality of the unconditional basis considered, with I denoting the identity operator on H, and A^* being the adjoint operator of A. Projection operator $[\widehat{\mathbf{\Phi}}_M]^*$ is applied to the left-hand side of equation (A4.5), and its adjoint $\widehat{\mathbf{\Phi}}_M$ to the right-hand-side of such an equation, obtaining

$$\widehat{\vec{\Lambda}}\left(\vec{\mathcal{L}}(i,j)\right) = \vec{\Lambda}(\widehat{\mathbf{R}_{YX}}(i,j))[\vec{\Lambda}(\widehat{\mathbf{R}_X})(i,j)]^{-1}, \tag{A4.6}$$

where

$$\begin{aligned}
\widehat{\vec{\Lambda}}\left(\vec{\mathcal{L}}\right) &= \begin{bmatrix} \left[\widehat{\mathbf{\Phi}}_M\right]^* \mathcal{L}_1(i,j)\widehat{\mathbf{\Phi}}_M \\ \ldots \\ \left[\widehat{\mathbf{\Phi}}_M\right]^* \mathcal{L}_q(i,j)\widehat{\mathbf{\Phi}}_M \end{bmatrix} = \begin{bmatrix} \widehat{\Lambda}\left(\mathcal{L}_1(i,j)\right) \\ \ldots \\ \widehat{\Lambda}\left(\mathcal{L}_q(i,j)\right) \end{bmatrix} \\
[\vec{\Lambda}(\widehat{\mathbf{R}_X})(i,j)]^{-1} &= \begin{bmatrix} \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_{1,1}(i,j)\widehat{\mathbf{\Phi}}_M & \ldots & \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_{1,q}(i,j)\widehat{\mathbf{\Phi}}_M \\ \ldots & \ldots & \ldots \\ \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_{q,1}(i,j)\widehat{\mathbf{\Phi}}_M & \ldots & \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_{q,q}(i,j)\widehat{\mathbf{\Phi}}_M \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \Lambda\left(\hat{R}_{1,1}(i,j)\right) & \Lambda\left(\hat{R}_{1,2}(i,j)\right) & \ldots & \Lambda\left(\hat{R}_{1,q}(i,j)\right) \\ \ldots & \ldots & \ldots & \ldots \\ \Lambda\left(\hat{R}_{q,1}(i,j)\right) & \Lambda\left(\hat{R}_{q,2}(i,j)\right) & \ldots & \Lambda\left(\hat{R}_{q,q}(i,j)\right) \end{bmatrix}^{-1} \\
\vec{\Lambda}(\widehat{\mathbf{R}_{YX}}) &= \begin{bmatrix} \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_1(i,j)\widehat{\mathbf{\Phi}}_M \\ \ldots \\ \left[\widehat{\mathbf{\Phi}}_M\right]^* \hat{R}_q(i,j)\widehat{\mathbf{\Phi}}_M \end{bmatrix} = \begin{bmatrix} \Lambda\left(\hat{R}_1(i,j)\right) \\ \ldots \\ \Lambda\left(\hat{R}_q(i,j)\right) \end{bmatrix}.
\end{aligned} \tag{A4.7}$$

The following assumption is made in relation to the homogeneity in space of the resolution of the identity of operators $\{\mathcal{L}_m(i,j),\ (i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, for $m = 1,\ldots,q$, that is, such operators have a common eigenvector system given by an unconditional basis of H (see, for example, Dautray and Lions, 1985; Donoho, 1993).

Assumption A0. For every $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, and $m = 1,\ldots,q$,

$$\mathcal{L}_m(i,j)\phi_k = \lambda_{km}^{\mathcal{L}_m}(i,j)\phi_k, \quad k \geq 1.$$

Equivalently, we are assuming that the possible spatial heterogeneity in the linear relationship between the regressors and the fixed effect curve is induced by the spatial heterogeneity of the point spectra of the compact and self-adjoint regression operators $\mathcal{L}_m$, $m = 1,\ldots,q$.

Remark 9. *Note that, given a self-adjoint and compact operator $\mathcal{A}$, any continuous function f of such an operator $f(\mathcal{A})$ has the same eigenvector system as operator $\mathcal{A}$ (see, for example, Dautray and Lions, 1985). In our case, under **Assumption A0**, the spatial heterogeneous decay velocity of the eigenvalues of the regression operators means that the stability of the functional linear relationship between the fixed effect curve and the regressors depends on space.*

Additionally, **Assumption A0** means that we look for a solution $\widehat{\boldsymbol{\mathcal{L}}} = (\widehat{\mathcal{L}}_1(i,j),\ldots,\widehat{\mathcal{L}}_q(i,j))$ to equation system (A4.5) in the subspace of $[\mathcal{L}(H)]^q$ generated by the basis $(\widehat{\phi}_k \otimes \widehat{\phi}_k, \underset{q}{\ldots}, \widehat{\phi}_k \otimes \widehat{\phi}_k)$, $k \geq 1$. That is, we look for a solution of the form

$$\widehat{\boldsymbol{\mathcal{L}}}(\varphi) = \left(\left\langle \sum_{k=1}^{\infty} \widehat{\lambda}_{k1}^{\mathcal{L}_1}(i,j)\widehat{\phi}_k \otimes \widehat{\phi}_k, \varphi \right\rangle_H, \ldots, \left\langle \sum_{k=1}^{\infty} \widehat{\lambda}_{kq}^{\mathcal{L}_q}(i,j)\widehat{\phi}_k \otimes \widehat{\phi}_k, \varphi \right\rangle_H \right) \tag{A4.8}$$

for $\varphi \in H$, and for $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, with $\mathcal{L}(H)$ denoting the space of bounded linear operators on H. In particular, for a given truncation order M, the finite-dimensional approximation of the solution to equation system (A4.5) is then obtained from equations (A4.6)-(A4.7), where, in this case, $\widehat{\boldsymbol{\Phi}}_M^*$ represents the projection operator into the common eigenvector

system $\{\widehat{\phi}_1, \ldots, \widehat{\phi}_M\}$ of operators $\widehat{\mathcal{L}}_m(i,j)$, for $m = 1, \ldots, q$, and $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$ (see Ruiz-Medina, Anh, Espejo and Frías, 2013).

In practice, we perform the non-parametric wavelet-kernel-based estimation of $R_0(i,j)$ from the empirical auto-covariance operator $\widehat{R}_0(i,j)$, for $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$. Hence, the empirical eigenvector system $\{\widehat{\phi}_k,\ k \geq 1\}$ of the obtained auto-covariance operator estimate is a compactly supported orthogonal wavelet basis (see, for example, Angelini, De Canditiis and Leblanc, 2003), which provides an unconditional basis for spaces in the Hölder, Sobolev, Besov, and Triebel scales (see, for example Donoho, 1993). This will be the unconditional basis of space H selected for projection in Section A4.6, where a real-data example is considered for illustration purposes. This is the reason why we have denoted with a *hat* $\widehat{\phi}$ the elements of the unconditional basis considered through this section for projection, except in **Assumption A0**, since we wants to reflecting its empirical nature in practice.

A4.3.1 Fixed-effect-curve-based discrimination

In the proposed spatial functional statistical classification procedure, two fixed effect curves located at different nodes will belong to the same group when the Hilbert-Schmidt distance between the corresponding regression operators is close to zero. Note that, under **Assumption A0**, the Hilbert-Schmidt distance can be computed as

$$\begin{aligned} \|\boldsymbol{\mathcal{L}}(i,j) - \boldsymbol{\mathcal{L}}(k,l)\|_{[\mathcal{S}(H)]^q} &= \|\vec{\lambda}^{\boldsymbol{\mathcal{L}}(i,j)} - \vec{\lambda}^{\boldsymbol{\mathcal{L}}(k,l)}\|_{[l^2]^q} \\ &= \left[\sum_{m=1}^{q} \|\vec{\lambda}^{\mathcal{L}_m}(i,j) - \vec{\lambda}^{\mathcal{L}_m}(k,l)\|_{l^2}^2\right]^{1/2} \\ &= \left[\sum_{m=1}^{q}\sum_{p=1}^{\infty} [\lambda_{pm}^{\mathcal{L}_m}(i,j) - \lambda_{pm}^{\mathcal{L}_m}(k,l)]^2\right]^{1/2}, \end{aligned} \tag{A4.9}$$

where $\mathcal{S}(H)$ denotes the Hilbert space of Hilbert-Schmidt operators on H. Here, $\{\lambda_{km}^{\mathcal{L}_m}(i,j),\ k \geq 1\}$, $m = 1, \ldots, q$, denote, as before, the pure point spectra associated with the regression operators $\mathcal{L}_m(i,j)$ at spatial location $(i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$, for $m = 1, \ldots, q$, respectively.

In real-data applications, these spectra are replaced by their empirical versions given by the spectra of the nonparametric estimates $\widehat{\mathcal{L}}_m(i,j)$, $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$ of the respective operators $\mathcal{L}_m(i,j)$, $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, for $m = 1,\ldots,q$. A fixed truncation level M is also considered for approximation of (A4.9) in practice. This truncation level can be chosen based on the percent of explained variance, cross-validation and/or sensitivity analysis. Information criteria can also be applied (see, for example, Fujikoshi and Satoh, 1997). Hence, we compute

$$\|\widehat{\vec{\lambda}}^{\mathcal{L}}(i,j) - \widehat{\vec{\lambda}}^{\mathcal{L}}(k,l)\|_{[l^2_M]^q} = \left[\sum_{m=1}^{q}\sum_{p=1}^{M}[\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(i,j) - \widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(k,l)]^2\right]^{1/2}. \tag{A4.10}$$

Remark 10. *Under* ***Assumption A0****, the spectral distance (A4.9) is equivalent to the Hilbert-Schmidt distance between vector operators* $(\mathcal{L}_1(i,j),\ldots,\mathcal{L}_q(i,j))$ *and* $(\mathcal{L}_1(k,l),\ldots,\mathcal{L}_q(k,l))$*. Small values of this distance means that the corresponding operators are* close *in the* L^2 *sense. In particular, when this distance is null, the involved operators coincide. The stability of the fixed effect curves against small perturbations of the regressors increases when the Hilbert-Schmidt norm of the regression operators decreases.*

In the case considered where eigenvectors are given by a common compactly supported orthogonal wavelet basis, the truncation order M *is interpreted as the number of resolution levels taken into consideration in the finite-dimensional approximation of curves from a previously defined coarsest scale. In this setting, when the selected wavelet system is* $[s]+1$*-regular for* $s > 1/2$*, Theorem 5.1 in Angelini, De Canditiis and Leblanc, (2003) provides the order of magnitude* $\mathcal{O}\left(n^{\frac{-2s}{2s+1}}\right)$*, with* n *denoting the sample size, of the functional mean-square error (also referred as mean integrated square error (MISE) in the cited paper) associated with the derived wavelet-kernel-based estimators of* $\mathcal{L}_i(\cdot,\cdot)$*,* $i = 1\ldots,q$*.*

Summarizing, the main steps involved in the fixed-effect classification procedure proposed

are the following:

Step FE1 For $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, compute the empirical auto-covariance operator $\widehat{R}_0(i,j)$, and the cross-covariance operators $\widehat{R}_l(i,j)$, $l = 1,\ldots,q$, $\widehat{R}_{m,n},(i,j)$, $m,n = 1,\ldots,q$.

Step FE2 For $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, the nonparametric wavelet-kernel-based estimation of the autocovariance operator $R_0(i,j)$ from the empirical auto-covariance operator $\widehat{R}_0(i,j)$ computed in Step FE1 is achieved for a given truncation order M.

Step FE3 For $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, numerical inversion of the equation system (A4.5) by projection into the compactly supported wavelet basis considered in Step FE2 is performed, obtaining $\widehat{\mathcal{L}}_k(i,j)$, $k = 1,\ldots,q$, from equations (A4.6)–(A4.8).

Step FE4 For $(i,j),(k,l) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, compute the distance (A4.10) in terms of the spectra of operators $\widehat{\mathcal{L}}_m(i,j)$, and $\widehat{\mathcal{L}}_m(k,l)$, $m = 1,\ldots,q$, obtained in Step FE3.

Step FE5 When $\|\widehat{\vec{\lambda}}^{\mathcal{L}}(i,j) - \widehat{\vec{\lambda}}^{\mathcal{L}}(k,l)\|_{[l^2_M]^q} \simeq 0$, the fixed effect curves located at (i,j) and (k,l) are classified in the same group. Otherwise, they will belong to different clusters.

Remark 11. *Under the conditions in Remark 10, for* $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$, *the functional mean-square errors of the wavelet-kernel-based estimators* $\widehat{\mathcal{L}}_m(i,j)$, $m = 1,\ldots,q$, *in the space* $\mathcal{S}(H)$ *of Hilbert-Schmidt operators on* H *display a hyperbolic decay with respect to the sample size. Hence, by Chebyshev's inequality, when the sample size goes to infinity, the limit superior of the corresponding sequence of wavelet-kernel-based estimators converges to the true regression operator in probability. Thus, the almost surely convergence of these estimators to the respective regression operators holds. Hence, in the case where* $(\mathcal{L}_1(i,j),\ldots,\mathcal{L}_q(i,j)) = (\mathcal{L}_1(k,l),\ldots,\mathcal{L}_q(k,l)) = (\mathcal{A}_1,\ldots,\mathcal{A}_q)$, *for certain fixed vector operator* $\boldsymbol{A} = (\mathcal{A}_1,\ldots,\mathcal{A}_q)$, *the*

following upper bound can be derived for equation (A4.10):

$$\begin{aligned}
\|\widehat{\vec{\lambda}}^{\mathcal{L}}(i,j) - \widehat{\vec{\lambda}}^{\mathcal{L}}(k,l)\|_{[l^2_M]^q} &= \left[\sum_{m=1}^{q}\sum_{p=1}^{M}[\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(i,j) - \widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(k,l)]^2\right]^{1/2} \\
&= \left[\sum_{m=1}^{q}\sum_{p=1}^{M}[\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(i,j) - \lambda^{\mathcal{A}_m}_{pm} - (\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(k,l) - \lambda^{\mathcal{A}_m}_{pm})]^2\right]^{1/2} \\
&\leq \left[\sum_{m=1}^{q}\sum_{p=1}^{M}[\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(i,j) - \lambda^{\mathcal{A}_m}_{pm}]^2\right]^{1/2} \\
&+ \left[\sum_{m=1}^{q}\sum_{p=1}^{M}[\widehat{\lambda}^{\widehat{\mathcal{L}}_m}_{pm}(k,l) - \lambda^{\mathcal{A}_m}_{pm}]^2\right]^{1/2} \longrightarrow 0, \quad a.s,
\end{aligned} \tag{A4.11}$$

as $N(i,j)\wedge N(k,l) \to \infty$. Here, as usual, a.s. means almost surely convergence, and $\lambda^{\mathcal{A}_m}_{pm}$ denotes the pth eigenvalue of $\mathcal{A}_m$, for $m = 1, \ldots, q$.

In Step FE5, the symbol $\simeq 0$ means that

$$\|\widehat{\vec{\lambda}}^{\mathcal{L}}(i,j) - \widehat{\vec{\lambda}}^{\mathcal{L}}(k,l)\|_{[l^2_M]^q} \leq \mathcal{O}(f(N(i,j)\wedge N(k,l))) \quad a.s,$$

with $f(x) \longrightarrow 0$, as $x \to \infty$, a.s.

A4.4 Spatial heterogeneity of the random effect curves

Let Z be defined by the following equation

$$Z(i,j) = \bar{Y}(i,j) - (\widehat{\mathcal{L}}_1(i,j)\bar{X}_1^{(i,j)} + \cdots + \widehat{\mathcal{L}}_q(i,j)\bar{X}_q^{(i,j)}, \tag{A4.12}$$

for $(i,j) \in \{1,\ldots,K\}\times\{1,\ldots,N\}$, where $\widehat{\mathcal{L}}_k(i,j)$, $k = 1,\ldots,q$, are the nonparametric estimators derived in the previous section, and, as before

$$\bar{Y}(i,j) = \frac{1}{N(i,j)}\sum_{h=1}^{N(i,j)} Y(i,j,h); \quad \bar{X}_k^{(i,j)} = \frac{1}{N(i,j)}\sum_{h=1}^{N(i,j)} X_{k,h},$$

for $k = 1, \ldots, q$, are the respective averaged values of the original functional data, and of the regressors over the individuals located at each $(i, j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$. Thus,

$$Z(i,j) = \widehat{\overline{\nu}(i,j)} + \widehat{\sigma\overline{\epsilon}(i,j)}, \quad (i,j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}. \tag{A4.13}$$

For implementation of the estimation procedure proposed, we apply wavelet shrinkage denoising to $Z(i, j)$, obtaining $\widehat{\overline{\nu}}(i, j)$, for each $(i, j) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$.

Given, as before, a $K \times N$ spatial regular grid, the following symmetric spatial autoregressive dynamics is assumed to be satisfied by $\overline{\nu}$:

$$\begin{aligned} \overline{\nu}(i,j) &= L_1\overline{\nu}(i-1,j) + L_2\overline{\nu}(i,j-1) + L_3\overline{\nu}(i-1,j-1) + \xi(i,j) \\ &= L_1\overline{\nu}(i+1,j) + L_2\overline{\nu}(i,j+1) + L_3\overline{\nu}(i+1,j+1) + \xi(i,j), \end{aligned} \tag{A4.14}$$

where, for $k = 1, 2, 3$, $L_k \in \mathcal{L}(H)$, with $\mathcal{L}(H)$ denoting as before the space of bounded linear operators on H. Note that operator L_1 acts on the vertical direction, operator L_2 acts on the horizontal direction, and operator L_3 acts on the diagonal direction. The H-valued process ξ is white noise in the strong sense, with $E\|\xi(i,j)\|_H^2 = \sigma_\xi^2$, $E[\xi(i,j) \otimes \xi(k,l)] = \delta_{(i,j)-(k,l)} E[\xi(0,0) \otimes \xi(0,0)]$, with $\delta_{(i,j)}$ denoting the Kronecker delta. In particular, ξ has finite functional variance σ_ξ^2, and is weak-sense stationary in space.

The functional second-order structure of $\overline{\nu}$ is defined in terms of the following covariance operators:

$$\begin{aligned} R_{0,0} &= E\left[\overline{\nu}(i,j) \otimes \overline{\nu}(i,j)\right] \\ R_{1,0} &= E\left[\overline{\nu}(i+1,j) \otimes \overline{\nu}(i,j)\right] \\ R_{0,1} &= E\left[\overline{\nu}(i,j+1) \otimes \overline{\nu}(i,j)\right] \\ R_{1,1} &= E\left[\overline{\nu}(i+1,j+1) \otimes \overline{\nu}(i,j)\right]. \end{aligned} \tag{A4.15}$$

Replacing in equation (A4.15), $\overline{\nu}(i+1,j)$, $\overline{\nu}(i,j+1)$, and $\overline{\nu}(i+1,j+1)$ by their expressions, according to the functional state equation (A4.14), we obtain the following equation system defining operators L_k, $k=1,2,3$, respectively (see Ruiz-Medina, 2012a):

$$\begin{aligned} R_{1,0} &= L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1} \\ R_{0,1} &= L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0} \\ R_{1,1} &= L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0}. \end{aligned} \tag{A4.16}$$

For solving equation system (A4.16), we consider the following assumption:

Assumption A1 For a given unconditional basis $\{\psi_l,\ l \geq 1\}$ of H, for $k=1,2,3$,

$$L_k \psi_l = \lambda_l(L_k)\psi_l, \quad l \geq 1.$$

Under **Assumption A1**, we look for a vector integral operator (L_1, L_2, L_3), solution of equation system (A4.16), whose components are of the form

$$L_k(\varphi) = \left\langle \sum_{l=1}^{\infty} \lambda_l(L_k)\psi_l \otimes \psi_l, \varphi \right\rangle_H, \quad \forall \varphi \in H, \quad k=1,2,3.$$

In practice, we follow a similar numerical projection scheme as in the previous section. Specifically, from the empirical covariance operator $\hat{R}_{0,0}$, we compute the wavelet-kernel-based estimation of $R_{0,0}$. The involved compactly supported orthogonal wavelet basis is such that it provides an unconditional basis of H, and the eigenvectors of the derived wavelet-kernel-based estimator of $R_{0,0}$. Specifically, let us denote by $\widehat{\mathbf{\Psi}}_M^*$ the projection operator into $\{\widehat{\phi}_{0,k},\ k \in \Gamma_0\} \cup \{\widehat{\psi}_{l,k},\ k \in \Gamma_l,\ l=0,\ldots,M\}$, where parameter M indicates the number of resolution levels considered in the finite-dimensional approximation we will compute for numerical inversion of (A4.16), and $\{\widehat{\phi}_{0,k},\ k \in \Gamma_0\}$ denotes the system of scaling functions defining the space V_0 where the *draft* of each function is generated. Here, $\{\widehat{\psi}_{l,k},\ k \in \Gamma_l,\ l=0,\ldots,M\}$ represents the wavelet systems generating subspaces W_j, $j=0,\ldots,M$, where functional details are displayed. Keeping

in mind that $\widehat{\mathbf{\Psi}}_M^* \widehat{\mathbf{\Psi}}_M = I$, the composition with $\widehat{\mathbf{\Psi}}_M^*$ at the left-hand-side of equation (A4.16), and with $\widehat{\mathbf{\Psi}}_M$ at the right-hand-side of such an equation leads to the following finite-dimensional approximation $\widehat{L}_k$ of the nonparametric estimator of L_k, for $k = 1, 2, 3$,

$$\begin{bmatrix} \widehat{\mathbf{\Lambda}(L_1)} \\ \widehat{\mathbf{\Lambda}(L_2)} \\ \widehat{\mathbf{\Lambda}(L_3)} \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda}(\widehat{R}_{0,0}) & \mathbf{\Lambda}(\widehat{R}_{1,1}) & \mathbf{\Lambda}(\widehat{R}_{0,1}) \\ \mathbf{\Lambda}(\widehat{R}_{1,1}) & \mathbf{\Lambda}(\widehat{R}_{0,0}) & \mathbf{\Lambda}(\widehat{R}_{1,0}) \\ \mathbf{\Lambda}(\widehat{R}_{0,1}) & \mathbf{\Lambda}(\widehat{R}_{1,0}) & \mathbf{\Lambda}(\widehat{R}_{0,0}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Lambda}(\widehat{R}_{1,0}) \\ \mathbf{\Lambda}(\widehat{R}_{0,1}) \\ \mathbf{\Lambda}(\widehat{R}_{1,1}) \end{bmatrix}, \tag{A4.17}$$

where, in this case, $\mathbf{\Lambda}(\mathcal{K}) = \widehat{\mathbf{\Psi}}_M^* \mathcal{K} \widehat{\mathbf{\Psi}}_M$ for a bounded linear operator $\mathcal{K}$. Hence,

$$\widehat{L}_k = \widehat{\mathbf{\Psi}}_M \widehat{\mathbf{\Lambda}(L_k)} \widehat{\mathbf{\Psi}}_M^*, \quad k = 1, 2, 3. \tag{A4.18}$$

In the computation of the empirical SARH(1) variogram we also consider the wavelet-kernel-based estimation of $R_{1,0}R_{1,0}^*$, $R_{0,1}R_{0,1}^*$ and $R_{1,1}R_{1,1}^*$ from the respective empirical covariance operators $\widehat{R}_{1,0}$, $\widehat{R}_{0,1}$, and $\widehat{R}_{1,1}$, as well as from the compactly supported orthogonal wavelet basis $\{\widehat{\phi}_{0,k},\ k \in \Gamma_0\} \cup \{\widehat{\psi}_{l,k},\ k \in \Gamma_l,\ l = 0, \ldots, M\}$, previously considered for estimation of $R_{0,0}$, and L_k, $k = 1, 2, 3$.

Let us now consider the curve spatial increments $\Delta_{i,j} = \overline{\nu}_{i,j} - \overline{\nu}_{i-1,j} - \overline{\nu}_{i,j-1} + \overline{\nu}_{i-1,j-1}$. The following empirical version of SARH(1) variogram can be computed: For $(i, j) \in \{1, \ldots, K\} \times$

$\{1,\ldots,N\}$,

$$\begin{aligned}
\widehat{\gamma}(i,j) &= E\widehat{\|\Delta_{i,j}\|}_H^2 = E\|\widehat{\overline{\nu}}_{i,j} - \widehat{\overline{\nu}}_{i-1,j} - \widehat{\overline{\nu}}_{i,j-1} + \widehat{\overline{\nu}}_{i-1,j-1}\|_H^2 \\
&= E\|[\widehat{L}_1 - I]\widehat{\overline{\nu}}_{i-1,j} + [\widehat{L}_2 - I]\widehat{\overline{\nu}}_{i,j-1} + [\widehat{L}_3 + I]\widehat{\overline{\nu}}_{i-1,j-1}\|_H^2 \\
&= E\|[\widehat{L}_1 - I]\widehat{\overline{\nu}}_{i-1,j}\|_H^2 + E\left[\left\langle [\widehat{L}_1 - I]\widehat{\overline{\nu}}_{i-1,j}, \widehat{\overline{\nu}}_{i,j-1}[\widehat{L}_2 - I]^*\right\rangle_H\right] \\
&+ E\left[\left\langle [\widehat{L}_1 - I]\widehat{\overline{\nu}}_{i-1,j}, \widehat{\overline{\nu}}_{i-1,j-1}[\widehat{L}_3 + I]^*\right\rangle_H\right] \\
&+ E\left[\left\langle [\widehat{L}_2 - I]\widehat{\overline{\nu}}_{i,j-1}, \widehat{\overline{\nu}}_{i-1,j}[\widehat{L}_1 - I]^*\right\rangle_H\right] \\
&+ E\|[\widehat{L}_2 - I]\widehat{\overline{\nu}}_{i,j-1}\|_H^2 + E\left[\left\langle [\widehat{L}_2 - I]\widehat{\overline{\nu}}_{i,j-1}, \widehat{\overline{\nu}}_{i-1,j-1}[\widehat{L}_3 + I]^*\right\rangle_H\right] \\
&+ E\left[\left\langle [\widehat{L}_3 + I]\widehat{\overline{\nu}}_{i-1,j-1}, \widehat{\overline{\nu}}_{i-1,j}[\widehat{L}_1 - I]^*\right\rangle_H\right] \\
&+ E\left[\left\langle [\widehat{L}_3 + I]\widehat{\overline{\nu}}_{i-1,j-1}, \widehat{\overline{\nu}}_{i,j-1}[\widehat{L}_2 - I]^*\right\rangle_H\right] \\
&+ E\|[\widehat{L}_3 + I]\widehat{\overline{\nu}}_{i-1,j-1}\|_H^2. \qquad \text{(A4.19)}
\end{aligned}$$

Under **A.1**, considering, as before, the projections into the empirical wavelet eigenvector basis $\{\widehat{\phi}_{0,k},\ k \in \Gamma_0\} \cup \{\widehat{\psi}_{l,k},\ k \in \Gamma_l,\ l \in \mathbb{Z}\}$, equation (A4.19) can be rewritten as

$$\begin{aligned}
\widehat{\gamma}(i,j) &= \sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l^2(L_1 - I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i-1,j)] \\
&+ 2\sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l(L_1 - I)\widehat{\lambda}_l(L_2 - I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}(i-1,j)\widehat{\overline{\nu}}_{l,k}(i,j-1)] \\
&+ 2\sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l(L_1 - I)\widehat{\lambda}_l(L_3 + I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}(i-1,j)\widehat{\overline{\nu}}_{l,k}(i-1,j-1)] \\
&+ \sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l^2(L_2 - I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i,j-1)] \\
&+ 2\sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l(L_2 - I)\widehat{\lambda}_l(L_3 + I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}(i,j-1)\widehat{\overline{\nu}}_{l,k}(i-1,j-1)] \\
&+ \sum_{l=-\infty}^{\infty}\sum_{k\in\Gamma_l} \widehat{\lambda}_l^2(L_3 + I)\widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i-1,j-1)], \qquad \text{(A4.20)}
\end{aligned}$$

where $\{\widehat{\lambda}_l(A)\}$ denotes the empirical eigenvalue system of the estimator of operator A on H,

and $\widehat{\widehat{\nu}}_{l,k}(i,j) = \left\langle \widehat{\widehat{\nu}}(i,j), \psi_{l,k} \right\rangle_H$, for $k \in \Gamma_l$, $l \in \mathbb{Z}$, and $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$. Note that the expectation estimates appearing at the right-hand side of the sums in equation (A4.20) are computed in terms of the empirical wavelet covariogram of process $\widehat{\widehat{\nu}}$.

Without loss of generality, we now consider for simplicity the space $H = L^2([0,1])$. Hence, for a suitable compactly supported orthogonal wavelet basis providing a multiresolution analysis of $L^2([0,1])$ which is $([s]+1)$-regular, we obtain, for $k = 0,\ldots,2^l-1$, $l \geq l_0$, the following asymptotic approximations for l sufficiently large (see, for example, Proposition 2.1 in Angelini, De Canditiis and Leblanc, 2003, for the case $d = 1$):

$$\begin{aligned}
\widehat{E}[\widehat{\widehat{\nu}}^2_{l,k}(i-1,j-1)] &\simeq 2^{-2ls(i-1,j-1)/d} \\
\widehat{E}[\widehat{\widehat{\nu}}_{l,k}(i-1,j)\widehat{\widehat{\nu}}_{l,k}(i,j-1)] &\simeq 2^{-ls(i-1,j)/d}2^{-ls(i,j-1)/d} \\
\widehat{E}[\widehat{\widehat{\nu}}_{l,k}(i-1,j)\widehat{\widehat{\nu}}_{l,k}(i-1,j-1)] &\simeq 2^{-ls(i-1,j)/d}2^{-ls(i-1,j-1)/d} \\
\widehat{E}[\widehat{\widehat{\nu}}^2_{l,k}(i-1,j)] &\simeq 2^{-2ls(i-1,j)/d} \\
\widehat{E}[\widehat{\widehat{\nu}}^2_{l,k}(i,j-1)] &\simeq 2^{-2ls(i,j-1)/d} \\
\widehat{E}[\widehat{\widehat{\nu}}_{l,k}(i,j-1)\widehat{\widehat{\nu}}_{l,k}(i-1,j-1)] &\simeq 2^{-ls(i,j-1)/d}2^{-ls(i-1,j-1)/d}.
\end{aligned} \tag{A4.21}$$

For a given resolution level $l \geq l_0$, with l_0 sufficiently large, from equations (A4.20) and (A4.21), we obtain

$$\begin{aligned}
\widehat{\gamma}_l(i,j) \simeq\ & 2^{-ls(i-1,j)} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2(L_1 - I) + 2^{-ls(i,j-1)} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2(L_2 - I) + 2^{-ls(i-1,j-1)} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2(L_3 + I) \\
+\ & 2^{-ls(i-1,j)/2}2^{-ls(i,j-1)/2} \sum_{k=0}^{2^l-1} 2\widehat{\lambda}_l(L_1 - I)\widehat{\lambda}_l(L_2 - I) \\
+\ & 2^{-ls(i-1,j)/2}2^{-ls(i-1,j-1)/2} \sum_{k=0}^{2^l-1} 2\widehat{\lambda}_l(L_1 - I)\widehat{\lambda}_l(L_3 + I) \\
+\ & 2^{-ls(i,j-1)/2}2^{-ls(i-1,j-1)/2} \sum_{k=0}^{2^l-1} 2\widehat{\lambda}_l(L_2 - I)\widehat{\lambda}_l(L_3 + I),
\end{aligned} \tag{A4.22}$$

where $\widehat{\gamma}_l(i,j)$ represents the wavelet approximation of the local variation of the empirical SARH(1) variogram at resolution level l. For our classification purposes, we can choose level l_0 for generating the draft of the functions in $H = L^2([0,1])$, in terms of the translations and dilations of the scaling function, and a number M of higher resolution levels for the construction of the details, in terms of the corresponding wavelet functions. Specifically, we consider the projection of the random effect curve observations $\widehat{\overline{\nu}}$ into the spaces W_l, $l = l_0, \ldots, l_0 + M$, generated by the translations and dilations of the wavelet functions at resolution levels $l = l_0, \ldots, l_0 + M$. From equation (A4.22), we then have

$$\widehat{\gamma}_{l_0}^{l_0+M}(i,j) = \sum_{l=l_0}^{l_0+M} \widehat{\gamma}_l(i,j) \simeq \sum_{n=1}^{6} A_{n,l_0,M} f_n(i,j,l_0,M), \tag{A4.23}$$

for l_0 sufficiently large. In the next section we describe a functional statistical classification procedure of random effect curves based on function $\widehat{\gamma}_{l_0}^{l_0+M}(i,j)$ in equation (A4.23). Note that, given two nodes (i,j) and (k,l), in the spatial regular grid considered, the spatial homogeneity of SARH(1) increments located at these nodes means that functions $f_n(i,j,l_0,M)$, and $f_n(k,l,l_0,M)$, $n = 1, \ldots, 6$, in equation (A4.23) coincide. Specifically, ergodic theorems (see, for example, Friedman, 1970) ensure the almost surely convergence of the wavelet covariogram, the wavelet transform of the empirical covariance operator of $\widehat{\overline{\nu}}$, to the variance and covariance of the corresponding wavelet coefficients when the sample size goes to infinity. As $l \to \infty$, which is equivalent to consider that the sample size goes to infinity, equation (A4.21) then provides the almost surely order of magnitude of the sample second-order moments of wavelet coefficients of $\widehat{\overline{\nu}}$. Thus, in the case where $\widehat{\overline{\nu}}$ displays homogeneous spatial local variation at a region containing the nodes (i,j) and (k,l), the random variables $f_n(i,j,l_0,M)$, and $f_n(k,l,l_0,M)$ converge almost surely to the same limit $\sum_{l=l_0}^{l_0+M} \gamma_l$, as $M \to \infty$. Equivalently,

$$\left|\widehat{\gamma}_{l_0}^{l_0+M}(i,j)-\widehat{\gamma}_{l_0}^{l_0+M}(k,l)\right| = \left|\widehat{\gamma}_{l_0}^{l_0+M}(i,j)-\sum_{l=l_0}^{l_0+M}\gamma_l-\left(\widehat{\gamma}_{l_0}^{l_0+M}(k,l)-\sum_{l=l_0}^{l_0+M}\gamma_l\right)\right|$$
$$\leq \left|\widehat{\gamma}_{l_0}^{l_0+M}(i,j)-\sum_{l=l_0}^{l_0+M}\gamma_l\right| + \left|\widehat{\gamma}_{l_0}^{l_0+M}(k,l)-\sum_{l=l_0}^{l_0+M}\gamma_l\right| \longrightarrow 0, \quad M\to\infty, \quad \text{a.s.} \tag{A4.24}$$

Therefore,

$$\left|\widehat{\gamma}_{l_0}^{l_0+M}(i,j)-\widehat{\gamma}_{l_0}^{l_0+M}(k,l)\right| \leq \mathcal{O}(g(M)), \quad \text{a.s.}, \tag{A4.25}$$

where $g(M)\longrightarrow 0$, as $M\to\infty$, a.s. Note that, given the sparsity of wavelet transform, g goes to zero relatively fast (see Daubechies, 1992; Harten, 1993; Kumar and Mehra, 2005; among others).

A4.4.1 Random-effect-curve-based discrimination

Summarizing, the main steps involved, in practice, in the random-effect-curve–based classification procedure proposed are

Step RE1 Compute the observed averaged random effect curves $\widehat{\overline{\nu}}(i,j)$, $(i,j)\in\{1,\ldots,K\}\times\{1,\ldots,N\}$, from equations (A4.12) and (A4.13) applying wavelet shrinkage denoising.

Step RE2 Compute the empirical covariance operators $\widehat{R}_{0,0}, \widehat{R}_{1,0}, \widehat{R}_{0,1}$, and $\widehat{R}_{1,1}$.

Step RE3 Compute the covariance operator wavelet-kernel-based estimators from $\widehat{R}_{0,0}$, $\widehat{R}_{1,0}\widehat{R}_{1,0}^*$, $\widehat{R}_{0,1}\widehat{R}_{0,1}^*$, and $\widehat{R}_{1,1}\widehat{R}_{1,1}^*$.

Step RE4 Compute the non-parametric wavelet-based estimators $\widehat{L}_k$, $k=1,2,3$, in equation (A4.18).

Step RE5 Compute the wavelet covariogram of $\widehat{\overline{\nu}}$.

Step RE6 Compute $\widehat{\gamma}_{l_0}^{l_0+M}(i,j)$ for $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$ from a truncated version of equation (A4.20) as described in equation (A4.23), for a suitable l_0 sufficiently large. That is, compute, for $(i,j) \in \{1,\ldots,K\} \times \{1,\ldots,N\}$,

$$\begin{aligned}
\widehat{\gamma}_{l_0}^{l_0+M}(i,j) &= \widehat{\gamma}(i,j,l_0,M) = \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2 (L_1 - I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i-1,j)] \\
&+ 2 \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l (L_1 - I) \widehat{\lambda}_l (L_2 - I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}(i-1,j) \widehat{\overline{\nu}}_{l,k}(i,j-1)] \\
&+ 2 \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l (L_1 - I) \widehat{\lambda}_l (L_3 + I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}(i-1,j) \widehat{\overline{\nu}}_{l,k}(i-1,j-1)] \\
&+ \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2 (L_2 - I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i,j-1)] \\
&+ 2 \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l (L_2 - I) \widehat{\lambda}_l (L_3 + I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}(i,j-1) \widehat{\overline{\nu}}_{l,k}(i-1,j-1)] \\
&+ \sum_{l=l_0}^{l_0+M} \sum_{k=0}^{2^l-1} \widehat{\lambda}_l^2 (L_3 + I) \widehat{E}[\widehat{\overline{\nu}}_{l,k}^2(i-1,j-1)]. \qquad \text{(A4.26)}
\end{aligned}$$

Step RE7 Given two nodes (i,j) and (k,l), if $\widehat{\gamma}_{l_0}^{l_0+M}(i,j) - \widehat{\gamma}_{l_0}^{l_0+M}(k,l) \simeq 0$, in the sense of equation (A4.25), the random effect curves located at such nodes will be in the same group. Otherwise, they will belong to different clusters.

A4.5 Functional statistical classification algorithm

Let us now summarize the main steps involved in the implementation of the functional statistical classification methodology proposed in this paper to analyzing the firm panel data example in the next section.

Step 1 *Functional fixed effect model validation*

Step 1.1 A centroid with coordinates (x_{1i}, x_{2i}) is assigned to the Spanish community i, for $i = 1, \ldots, C$, with C denoting the number of Spanish communities analyzed in the firm panel data.

Step 1.2 For each industry sector k, $k = 1, \ldots, s$, and for $i = 1, \ldots, C$, let us consider the indebtedness curve sample of size $N(i,k)$ at community i, where $N(i,k)$ denotes the number of firms located at region i within the industry sector k. For $k = 1, \ldots, s$, and $i = 1, \ldots C$, compute the empirical covariance operators

$$\widehat{R}_0(x_{1i}, x_{2i}, k), \widehat{R}_l(x_{1i}, x_{2i}, k), \widehat{R}_{m,n}(x_{1i}, x_{2i}, k), \quad l, m, n = 1, \ldots, q.$$

Step 1.3 For each industry sector, Leave-One-Out Cross-Validation (LOOCV) is applied at each one of the Spanish communities studied. Specifically, for each industry sector $k = 1, \ldots, s$, and within each Spanish community i, $i = 1, \ldots, C$, the corresponding $N(i,k)$ iterations of the LOOCV are running, consisting of removing the firm h_0 used for validation, and implementing Steps FE1-FE3, in terms of the corresponding training firm sample, to obtaining the estimates $\widehat{\mathcal{L}}_1(i,k), \ldots, \widehat{\mathcal{L}}_q(i,k)$. Then, $\widehat{\beta}(x_{1i}, x_{2i}, h_0, k) = \sum_{m=1}^{q} \widehat{\mathcal{L}}_m(i,k) X_{m,h_0,k}$, and

$$\text{LOOCVE}(x_{1i}, x_{2i}, h_0, k) = \left\| \sum_{m=1}^{q} \widehat{\mathcal{L}}_m(i,k) X_{m,h_0,k} - Y(x_{1i}, x_{2i}, h_0, k) \right\|_{L^\infty},$$

where $Y(x_{1i}, x_{2i}, h_0, k)$ denotes the observed indebtedness curve at the firm h_0 in the Spanish community i and industry sector k, for $h_0 = 1, \ldots, N(i,k)$, and $i = 1, \ldots, C$, $k = 1, \ldots, s$.

Step 1.4 The mean of LOOCVE$(\cdot, \cdot, \cdot, \cdot)$ error over the firms of each community within a particular industry sector is calculated for assessment of the suitability of the functional multiple regression model to be considered in the fixed-effect classification procedure.

Step 1.5 If mean LOOCV error is small, spatial interpolation methods are applied to allocating data on a spatial regular grid (see, for example, Stein, 1999, 2009). The nodes of the

grid are spatially located close to the spatial coordinates of the centroid assigned to each one of the Spanish communities analyzed. This fact is also taken into account in the determination of the number of nodes defining the spatial regular grid selected.

Step 2 *Fixed effect curve classification.* Steps FE1-FE5 are implemented in terms of the spatial interpolated indebtedness curve samples at the nodes of the spatial regular grid. The members of each one of the clusters distinguished in Step FE5 are indebtedness curves whose fixed effect parameters have regression operators spatially homogeneous.

Step 3 *SARH(1) model validation*

Step 3.1 For each industry sector, LOOCV is applied to testing suitability of SARH(1) model for representing spatial interaction between indebtedness curves located at the nodes of the spatial regular grid considered in Step 1.5. After removing the validation node (i_0, j_0), Steps RE1-RE4 are computed from the training functional sample, considering a suitable scaling and translation of the interval $[0, 1]$ to the temporal period, years, analyzed. Hence, for $k = 1, \ldots, s$,

$$\begin{aligned} \text{LOOCVE}(i_0, j_0, k) &= \|\widehat{\widehat{\overline{\nu}}}(i_0, j_0, k) - \widehat{\overline{\nu}}(i_0, j_0, k)\|_{L^\infty} \\ &= \|\widehat{L}_1\widehat{\overline{\nu}}(i_0 - 1, j_0, k) + \widehat{L}_2\widehat{\overline{\nu}}(i_0, j_0 - 1, k) \\ &\qquad + \widehat{L}_3\widehat{\overline{\nu}}(i_0 - 1, j_0 - 1, k) - \widehat{\overline{\nu}}(i_0, j_0, k)\|_{L^\infty}, \end{aligned} \tag{A4.27}$$

for each $(i_0, j_0) \in \{1, \ldots, K\} \times \{1, \ldots, N\}$.

Step 3.2 The mean of LOOCVE$(\cdot, \cdot, \cdot)$ is computed over the nodes of the spatial regular grid considered, for each industry sector.

Step 4 *Random effect curve classification.* When the LOOCVE mean is small, Steps RE1-RE7 are implemented for obtaining the clustering of the indebtedness curves at the nodes of the spatial regular grid, according to the homogeneous spatial patterns displayed by the SARH(1) increments of the random effect parameters, in the mean quadratic sense.

Step 5 *Fixed and random effect curve classification.* The intersection between the clusters distinguished in Step 2 and in Step 4 is considered to detecting groups of indebtedness curves with spatially homogeneous regression operators and mean-quadratic functional local variation.

A4.6 Real data example

In this section, the financing decisions of firms during the period 1999-2007 are analyzed in a panel constituted by 638 Spanish companies, belonging to 4 different industry sectors, and located at 17 autonomous Spanish communities. Data have been collected from the SABI (*Sistema de Análisis de Balances Ibéricos*) database (see Ruiz-Medina and Espejo, 2012). The functional classification methodologies proposed are illustrated with the analysis of this firm panel data with the final objective of establishing:

- The groups of Spanish communities whose firms within a sector have indebtedness curves with a spatially homogeneous firm factor relationship.

- The groups of Spanish communities whose firms have indebtedness curves with spatially homogeneous mean-quadratic local variation.

- The groups of Spanish communities whose firms within a sector display indebtedness curves with spatially homogeneous regression operators and mean-quadratic functional local variation.

The firm factor determinants of leverage considered in the analysis of the financing decisions of the 638 Spanish companies studied are: *Firm size*, which is measured as the log of total assets; *Asset structure*, which consists of the net fixed assets divided by the total assets of the firm; *Profitability*, computed as the ratio between earnings before interest, taxes amortization and depreciation, and the total assets; *Growth*, for which as proxy, consider the growth of the assets, calculated as the annual change of the total assets of the firm; *Firm risk* given by the *Business risk*, defined as the standard deviation of earnings before interest and taxes over book value of total assets during the sample period; and, finally we consider the *Age*, measured as the logarithm of the number of years that the firm has been operating. These factors are assumed to be fixed over the companies belonging to the 17 geographical regions analyzed, and the 4 industry sectors studied (*Factories*, *Building*, *Commerce* and *Several*). Leverage is estimated as the quotient between the total debt and the total assets. There exist a large number of empirical studies undertaken to demonstrate the influence of the industry sector, and of the geographical region in the statistical analysis of firm panel data. They are usually incorporated as dummy variables (see Booth, *et al.*, 2001; and Degryse, De Goeij and Kappert, 2012).

The subsequent implementation has been developed with MatLab Language. Specifically, *Kernel Smoothing*, *Spatial Statistics* and *Wavelet* constitute the main toolboxes used. Indebtedness and firm factor curves are approximated by applying temporal local polynomial kernel smoothing, based on the Epanechnikov kernel (see, for example Rincón and Ruiz-Medina, 2012a, Appendix A.1). Functional fixed effect model validation is first implemented. Specifically, after computing Steps 1.1-1.4, the L^{∞}-norms of the mean LOOCV errors at the 17 Spanish communities are displayed in Table A4.1, for each one of the four industry sectors. Note that —— means that a region has not firms at some specific industry sector.

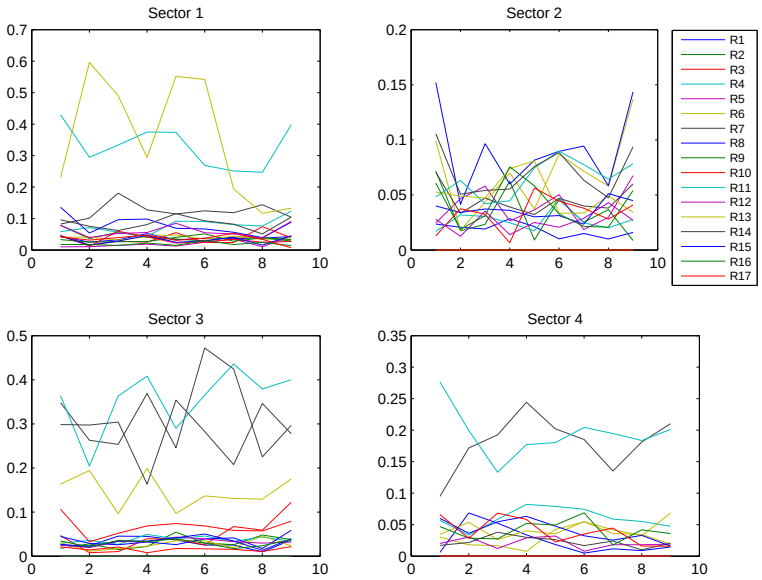
The mean LOOCV error curves at the 17 communities for each one of the four industry sectors appear in Figure A4.6.

Table A4.1: the L^{∞}-norms of the mean LOOCV errors at the 17 Spanish communities

Mean LOOCV error	SECTOR 1	SECTOR 2	SECTOR 3	SECTOR 4
GALICIA	0.0519	0.0284	0.0452	——
ASTURIAS	0.0781	0.0715	0.0544	——
CANTABRIA	0.0743	——	0.1222	——
PAÍS VASCO	0.1234	0.0470	0.0494	0.0821
NAVARRA	0.0432	0.0675	——	——
ARAGÓN	0.0476	0.1371	0.0460	0.0546
CATALUÑA	0.1805	0.0705	0.4723	0.2443
CAST. LEÓN	0.1361	0.1522	0.0586	0.0685
LA RIOJA	0.0511	——	——	——
EXTREMADURA	0.0597	——	0.0216	——
MADRID	0.4299	0.0897	0.4358	0.2768
CAST. MANCHA	0.0911	0.0430	0.0431	0.0320
C. VALENCIANA	0.5960	0.0809	0.1989	0.0683
ANDALUCÍA	0.1153	0.1055	0.3539	0.0379
MURCIA	0.0461	0.0511	0.0503	0.0632
BALEARES	0.0298	0.0754	0.0390	0.0688
CANARIAS	0.0534	0.0562	0.1063	0.0684

Since the above displayed validation results support the functional multiple regression modeling of H-valued fixed effect parameters at each community within each industry sector, Steps 1.5 and 2 are now implemented. In particular, in the implementation of Step 1.5, high-dimensional spatial interpolation (see, for example, Stein, 2009) is applied to obtaining the curve data samples (indebtedness and firm factor curves) on a 4×6 spatial regular grid. Here, the latitude and longitude defining the spatial location of the centroid associated with each Spanish community have been taken into account to define such a spatial regular grid. Specifically, Figure A4.6 shows the distribution of the 15 Spanish communities over the considered 4×6 spatial regular grid located at the square interval $[38°, 43°] \times [-6°, -1°]$ for covering the Iberian Peninsula. Note that Baleares and Canarias islands have not been included since they are considered from

Figure A4.1: Mean LOOCV error curves at the 17 communities, using the following abbreviations: R1 for Galicia, R2 for P. Asturias, R3 for Cantabria, R4 for Pais Vasco, R5 for Navarra, R6 for La Rioja, R7 for Aragon, R8 for Cataluña, R9 for Castilla Leon, R10 for Madrid, R11 for Castilla La Mancha, R12 for C. Valenciana, R13 for Canarias, R14 for Extremadura, R15 for Andalucia, R16 for Murcia, R17 for Baleares



the beginning in separated groups due to their latitude and longitude.

Figure A4.2: Distribution of the 15 Spanish communities centroids, located at the Iberian Peninsula, on the spatial regular grid considered

(43°, -6°)	*	*	*	*	*		(43°, -1°)
			*	*	*	*	
		*	*		*	*	
(36°, -6°)		*				*	(36°, -1°)

A4.6.1 Fixed-effect curve classification results

According to the local variability displayed by the approximated fixed-effect curves, $M = 9$ resolution levels have been considered, since detail wavelet empirical coefficients are null for higher resolution levels. For each industry sector, in the determination of the clusters from Step F5, we evaluate equation (A4.10) for the possible pairs of communities. Two communities located at nodes (i, j) and (k, l) of the 4×6 spatial regular grid considered are in the same cluster if

(i) $[\widehat{\lambda}_{pm}^{\widehat{\mathcal{L}}_m}(i,j) - \widehat{\lambda}_{pm}^{\widehat{\mathcal{L}}_m}(k,l)]^2 = 0$, for at least $p \geq 4$, and for $m = 1, \ldots, 6$,

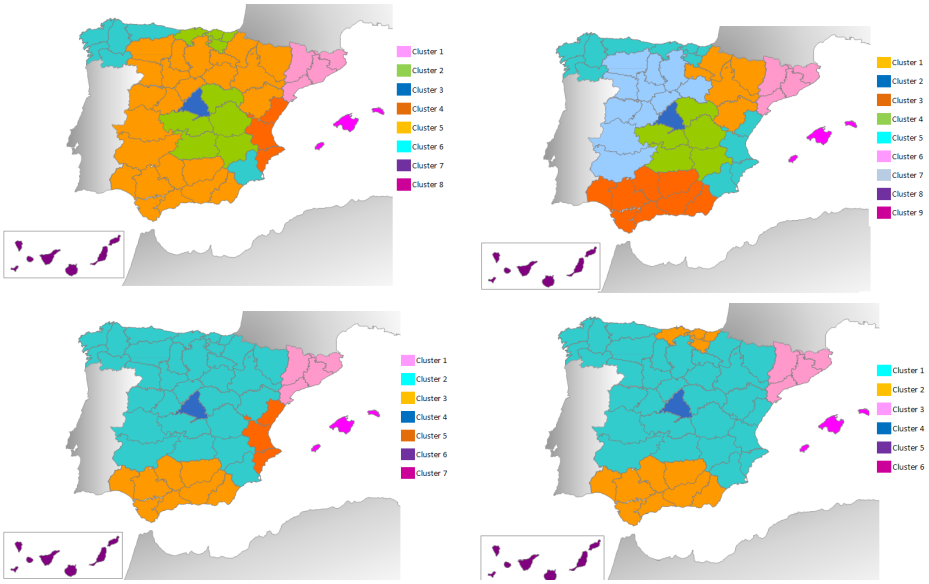
(ii) $\left[\sum_{m=1}^{6}\sum_{p=1}^{3}[\widehat{\lambda}_{pm}^{\widehat{\mathcal{L}}_m}(i,j) - \widehat{\lambda}_{pm}^{\widehat{\mathcal{L}}_m}(k,l)]^2\right]^{1/2} \leq T_s$, with the thresholds T_s, $s = 1, 2, 3, 4$, being selected according to Remarks 10 and 11 (see equation (A4.11)), and taking into account the empirical values of (A4.10) within the different candidate groups determined under restriction (i).

The threshold values used for each one on the industry sectors analyzed, T_s, $s = 1, 2, 3, 4$, are displayed in the following table:

THRESHOLDS	T_1	T_2	T_3	T_4
	0.208	0.161	0.215	0.305

When one community located at node (k, l) does not satisfy restrictions (i)-(ii) in the evaluation of (A4.10), in terms of its combination with the rest of 14 communities, it is allocated in a separated group, i.e., it constitutes a cluster. The resulting clusters for each one of the four industry sectors are now displayed in Figure A4.3 (see also tables below).

Figure A4.3: Classification results after implementation of Step 2 for the 4 industry sectors. *Factories* (top-left); *Building* (top-right); *Commerce* (bottom-left); *Several* (bottom- right)



SECTOR 1. *Factories*

Cluster 1	Cataluña
Cluster 2	Cantabria Pais Vasco Castilla La Mancha
Cluster 3	Madrid
Cluster 4	C. Valenciana
Cluster 5	Navarra La Rioja Aragón Castilla León Extremadura Andalucía
Cluster 6	Galicia P. Asturias Murcia
Cluster 7	Canarias
Cluster 8	Baleares

SECTOR 2. *Building*

Cluster 1	Navarra La Rioja Aragón
Cluster 2	Madrid
Cluster 3	Andalucía
Cluster 4	Castilla La Mancha
Cluster 5	Galicia Asturias Cantabria Pais Vasco C. Valenciana Murcia
Cluster 6	Cataluña
Cluster 7	Castilla León Extremadura
Cluster 8	Canarias
Cluster 9	Baleares

SECTOR 3. *Commerce*

Cluster 1	Cataluña
Cluster 2	Galicia P. Asturias Cantabria Pais Vasco Navarra La Rioja Aragón Castilla León Extremadura Castilla La Mancha Murcia
Cluster 3	Andalucía
Cluster 4	Madrid
Cluster 5	C. Valenciana
Cluster 6	Canarias
Cluster 7	Baleares

SECTOR 4. *Several*

Cluster 1	Galicia P. Asturias Navarra La Rioja Aragón Castilla León Extremadura Castilla La Mancha C. Valenciana Murcia
Cluster 2	Cantabria Pais Vasco Andalucía
Cluster 3	Cataluña
Cluster 4	Madrid
Cluster 5	Canarias
Cluster 6	Baleares

It can be appreciated that the industry Sector 2 on *building* displays a higher order of spatial heterogeneity in the functional relationship between indebtedness and firm factor curves, followed by Sector 1 in relation to *Factories*. Similar spatial heterogeneous patterns are displayed by Sectors 3 and 4 on *Commerce* and *Several*, respectively. Although a larger number of clusters

is distinguished in *Commerce*, with the most significant differences between these two last sectors localized in the Spanish communities of C. Valenciana, País Vasco and Cantabria.

A4.6.2 SARH(1)-based variogram classification

After implementation of Steps 3.1 and 3.2 for each one of the four industry sectors, the following table displays the L^∞-norm of the mean, over the nodes of the 4×6 spatial regular grid, of the LOOCV error computed in the SARH(1) model validation.

SARH(1) validation	SECTOR 1	SECTOR 2	SECTOR 3	SECTOR 4
	0.2318	0.1700	0.0916	0.8972

In the computation of Step 4, consisting of Steps RE1-RE6, we consider $l_0 = 5$ and $M = 9$. Step RE7 is then implemented in the following way: Equation (A4.24) is evaluated in the possible pairs of communities to be constructed. Two communities, located at nodes (i, j) and (k, n), are in the same cluster if

(i) $|\widehat{\gamma}_l(i,j) - \widehat{\gamma}_l(k,n)| = 0$, for $l \geq 4$

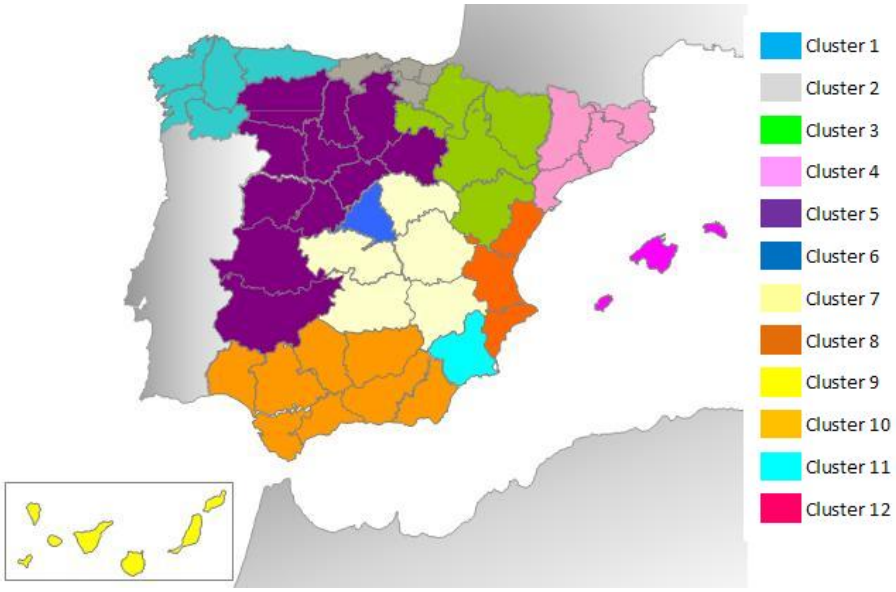
(ii) $\left|\sum_{l=5}^{5+3} \widehat{\gamma}_l(i,j) - \sum_{l=5}^{5+3} \widehat{\gamma}_l(k,n)\right| \leq S_s$, with the thresholds S_s, $s = 1, 2, 3, 4$, being established from the empirical values of (A4.24) within a candidate group determined under restriction (i), according to equation (A4.25).

The thresholds S_s, $s = 1, 2, 3, 4$, used are given in the following table:

THRESHOLDS	S_1	S_2	S_3	S_4
	0.015	0.035	0.01	0.025

As commented in the previous section, communities which do not satisfy (i)-(ii) in its evaluation of (A4.24) with the rest of 14 communities are allocated in a separated cluster. After implementation of (i)-(ii), Spanish communities are grouped into 12 common clusters for the four industry sectors considered, represented with different colors in the map of Spain given in the figure below.

Figure A4.4: Clusters defined by evaluation of the empirical wavelet approximation (A4.24) of the variogram differences



These clustering results are also collected in the following table:

Cluster	Spanish Communities
Cluster 1	Galicia P. Asturias
Cluster 2	Cantabria País Vasco
Cluster 3	Navarra La Rioja Aragón
Cluster 4	Cataluña
Cluster 5	Castilla León Extremadura
Cluster 6	Madrid
Cluster 7	Castilla La Mancha
Cluster 8	C. Valenciana
Cluster 9	Canarias
Cluster 10	Andalucía
Cluster 11	Murcia
Cluster 12	Baleares

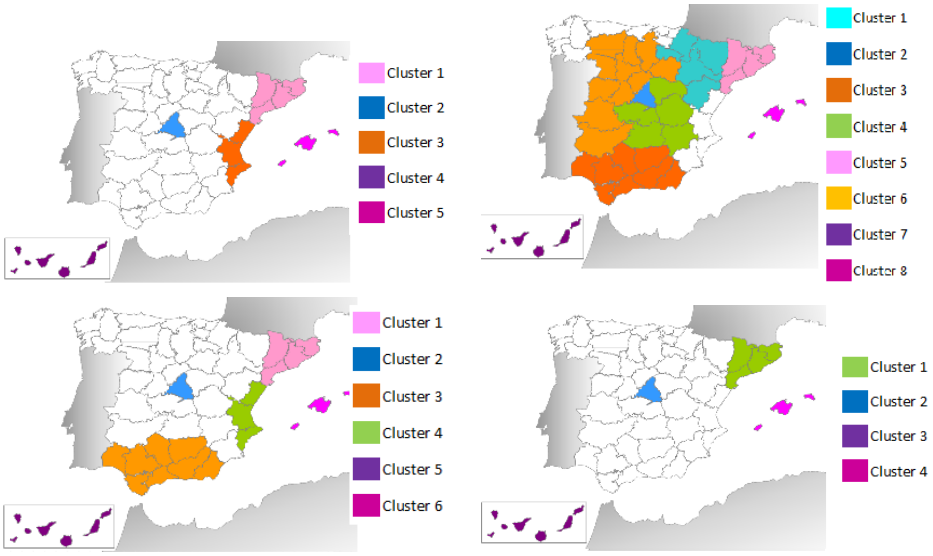
Since the same clusters are detected in relation to the mean-quadratic spatial local functional variation of indebtedness curves for the four industry sectors analyzed, real differences between industry sectors in the distribution of local spatially homogeneous patterns are only appreciated

in the functional relationship between indebtedness curves and firm factors. On the other hand, a higher level of spatial heterogeneity is displayed by the mean quadratic local functional variation properties. The conclusion of the study is that regional factors affect both, the firm structure (relationship between indebtedness curves and firm factors) and the mean quadratic local functional variation properties of indebtedness curves. However, it seems that the industry sector only affects the firm structure.

A4.6.3 Fixed and random effect curve classification

Groups of Spanish communities whose firms within a sector display indebtedness curves with spatially homogeneous relationships with firm factors, and spatially homogeneous mean-quadratic local functional variation are obtained as intersection of the clusters distinguished in Steps 2 and 4 of the functional classification algorithm described in Section A4.5. Specifically, the common groups distinguished by both methods are marked in different colors in the following maps corresponding to each one of the four industry sectors. Note that the communities which are not located at common clusters by both methods appear white on the map.

Figure A4.5: Common clusters for Step 2 and 4 of the functional classification algorithm proposed. Sector 1 (top-left), Sector 2 (top-right), Sector 3 (bottom-left) and Sector 4 (bottom-right)



The results displayed in Figure A4.5 are also reflected in the following table.

SECTOR 1. *Factories*

Cluster 1	Cataluña
Cluster 2	Madrid
Cluster 3	C. Valenciana
Cluster 4	Canarias
Cluster 5	Baleares

SECTOR 2. *Building*

Cluster 1	Navarra La Rioja Aragón
Cluster 2	Madrid
Cluster 3	Andalucía
Cluster 4	Castilla La Mancha
Cluster 5	Cataluña
Cluster 6	Castilla León Extremadura
Cluster 7	Canarias
Cluster 8	Baleares

SECTOR 3. *Commerce*

Cluster 1	Cataluña
Cluster 2	Andalucía
Cluster 3	Madrid
Cluster 4	C. Valenciana
Cluster 5	Canarias
Cluster 6	Baleares

SECTOR 4. *Several*

Cluster 1	Cataluña
Cluster 2	Madrid
Cluster 3	Canarias
Cluster 4	Baleares

Note that Canarias and Baleares have been considered since the beginning as two different groups due to their latitude and longitude outside of the Iberian Peninsula, where the spatial regular grid displayed in Figure A4.6 is located. We have also to note that Cataluña and Madrid are Spanish communities conforming separated groups common to the four industry sectors in the fixed and random effect based classification. The spatial distribution of groups constituted by the rest of Spanish communities changes with the industry sector when the fixed effect classification is considered, and, hence, when both spatial homogeneity criteria are simultaneously taken into consideration.

A4.7 Final comments

Functional statistical classification of curves incorporating spatial interactions, or interaction between individuals, constitutes an open research problem. In this paper, we present a first approach to the Gaussian case, when a spatial autoregressive dynamics of order one is assumed.

Specifically, the spatial functional mixed-effect approach proposed allows to detect heterogeneous trends and non-homogeneous local variability properties in space, in the mean-quadratic sense. Heterogeneity analysis in relation to trends is achieved in terms of the l^2 distance between the empirical spectra of the regression operators, while a functional version of the empirical variogram is formulated, based on SARH(1) dynamics, for random effect curve classification according to the functional temporal local variability in space displayed by such curves. The wavelet covariogram is considered in the last case for computing mean quadratic local functional variation exponents.

Numerical projection into the auto-covariance eigenvector system leads to an approximation of the functional variance in an optimal way. The wavelet-kernel-based estimation of the auto-covariance operator allows us to consider orthogonal wavelet functions providing a regular multiresolution analysis as empirical eigenvectors (see Angelini, De Canditiis and Leblanc, 2003). The finite-dimensional nonparametric estimates of the regression operators, and of the SARH(1) parameters are respectively computed in terms of the solution to the projected systems of equations derived from the spatial functional regression model, in the fixed-effect curve classification, as well as from the SARH(1) state equation, in the random-effect curve classification. Note that the presented approach for clustering spatial functional data adopts filtering methods in the implementation of Steps 3-4 to defining the observed random effect curves. Hence, the problem of selecting a suitable basis for filtering is crucial. Some recent approaches for clustering functional data choose spline bases for projection (see, for example, James and Sugar, 2003). Alternatively, wavelet bases can be considered for implementation of filtering methods in the classification of functional data (see, for example, Giacofci, *et al.*, 2013, in relation to wavelet-based classification in the context of functional mixed-effect models). In our approach, the wavelet spectra of the regression operators are used for fixed effect curve classification. Similarly, the wavelet spectra of the SARH(1) parameters and covariance operators lead to the designing of a parallel

methodology for random effect curve classification, based on the wavelet approximation of the variogram.

The presented approach can be viewed as a functional modified version of classical Maximum Likelihood Classifier, where instead of estimation of the trend, the regression operators are estimated, and where instead of considering the estimation of the covariance matrix, the SARH(1) variogram is approximated from the nonparametric projection estimation of SARH(1) parameters, as well as from the wavelet covariogram spectra of the observed random effect curves. Summarizing, the spatial functional statistical classification methodology proposed can be applied to detect local spatial homogeneity of regression operators defining fixed effect parameters, as well as local homogeneity patterns in the mean quadratic spatial local functional variation of Gaussian random effect curves, under a state space approach.

Appendix A5

Gegenbauer random field

Espejo, R. M., Leonenko, N and Ruiz-Medina, M. D. (2014).

Gegenbauer random fields.

Random Operators and Stochastic Equations, 22, DOI: 10.1515/rose-2014-0001.

Abstract

This paper introduces spatial long-range dependence time series models, based on the consideration of fractional difference operators associated with Gegenbauer polynomials. Their structural properties are analyzed. The spatial autoregressive Gegenbauer case is also studied, including the case of k factors with multiple singularities. An extension to the Hilbert-valued context is finally formulated leading to the introduction of a new class of spatial functional time series models.

A5.1 Introduction

In the last three decades, a big number of contributions have been developed in the context of spatial time series theory (see, Andel, 1986; Boissy *et al.*, 2005; Brockwell and Davis, 1991; Hosking, 1981; Martin, 1979; among others). Namely, since the initial work on spatial statistical models developed in Basu and Reinsel (1993); Bhattacharyya, Khalil and Richardson (1996); Guyon (1995); Jain (1981); Martin (1979, 1990, 1996); and Tjostheim (1978, 1981, 1983); the

spatial series modeling framework has been widely considered in several applied fields such as geology, geophysics, biology, agriculture, spatial econometrics, image processing, etc. This framework is useful when data are collected on a regular grid. Although irregularly spaced data can sometimes be replaced by equally spaced data, using interpolation techniques as the ones given, for example, in Delfiner and Delhomme (1975) and Papari and Petkov (2009).

The initial work by Martin (1979) is one of the pioneers in this field and was taken as a reference in the majority of later articles in this research area. Doubly geometric processes are introduced in this paper as a first attempt for modeling spatial dynamics (see also Martin, 1996). One of the most famous spatial time series models is the spatial autoregressive and moving average (SARMA) model. For example, recently, in Bosq (2010), the tensorial product of autoregressive Hilbertian processes is investigated, as extension to the functional context of the doubly geometric processes introduced by Martin (1979). Also, the Hilbert-valued formulation of the spatial unilateral dynamics displayed by the spatial autoregressive models of order one studied in Martin, (1979) is provided in Ruiz-Medina (2011). The structural properties of this class of spatial Hilbert-valued processes are also investigated in this paper. A functional spatial extrapolator is obtained in Ruiz-Medina (2012a), based on the method of moments, by projection into the biorthogonal eigenfunction system that provides the spectral decomposition of the infinite-dimensional parameters involved in the state equation. Alternative projection methodologies are investigated in Ruiz-Medina and Espejo (2012, 2013a), respectively based on the autocovariance eigenfunction system and the discrete interval wavelet transform. In the spatial time series context, weak dependence is usually assumed, but this condition is not always satisfied. This fact constitutes the motivation of the present paper. Specifically, this work will focus on Spatial Autoregressive and Moving Average (SARMA) models with long range dependence (LRD). Long-range dependence time series models have been essentially investigated in the one-dimensional case (see, for instance, Andel, 1986 or Brockwell, and Davis, 1991). In

the paper by Brockwell and Davis (1991), among other cases studied, the processes whose autocorrelation function asymptotically displays a power law behavior are investigated. We also cited the papers by Hosking (1981) and Granger and Joyeux (1980), as pioneering works regarding autoregressive and moving-average process (ARMA) with long memory. Both papers study the case where the degree of differentiation of the time series take fractional values. This type of fractional models are referred to as fractional autoregressive and moving average models FARMA models. In the context of FARMA models, Gray, Zhang and Woodward (1989) applied the generating function methodology in terms of the Gegenbauer polynomials. They refer to the introduced class of processes as Gegenbauer autoregressive moving average models, GARMA models. Later, Woodward, Cheng and Gray (1998) analyzed this model from a perspective closer to the long range dependence modeling framework. There exist different generalizations of GARMA processes discussed below. For example, Chung (1996a) introduced the so-called FARMA models, in terms of Gegenbauer autoregressive and moving average models displaying long memory, which, in particular, include the fractionally integrated processes.

The above referred long range dependence (LRD) models are introduced in the time domain. Recently, Boissy *et al.* (2005) have studied spatial strong-dependence models. In this paper, the structural properties of spatial Gegenbauer autoregressive processes displaying long range dependence are investigated.

A5.2 Preliminary concepts

For the development and better understanding of this chapter we initially introduce three spatial time series models. Specifically, the description of two-dimensional autoregressive models is performed in this section, as starting point in our work.

Model 1: (see Boissy *et. al.*, 2005) A process Y_{t_1,t_2} is called doubly-geometric process (Martin, 1979), if its correlation function is the product of two geometric terms, obeying the spatial

autoregressive model:

$$Y_{t_1,t_2} = \alpha Y_{t_1-1,t_2} + \beta Y_{t_1,t_2-1} - \alpha\beta Y_{t_1-1,t_2-1} + \varepsilon_{t_1,t_2}, \quad (t_1,t_2) \in \mathbb{Z}^2, \tag{A5.1}$$

where $-1 < \alpha, \beta < 1$, and $\varepsilon_{t_1,t_2} \in \mathbb{Z}^2$, is a white noise random field with mean zero $E(\varepsilon_{t_1,t_2}) = 0$, common variance σ_ε^2 and $E(\varepsilon_{t_1,t_2}, \varepsilon_{t_1',t_2'}) = \sigma_\varepsilon^2 \delta_{t_1}^{s_1'} \delta_{t_2}^{t_2'}$. Martin (1979) pointed out the desirability that the process displays reflection symmetric, that is, $\rho_{k,l} = \rho_{-k,l} = \rho_{k,-l} = \rho_{-k,-l}$, where $\rho_{k,l}$ is the autocorrelation between the lags k and l, with $\rho_{k,l} = corr(Y_{t_1,t_2}, Y_{t_1+k,t_2+l})$. From equation (A5.1), these requirements lead to the simple form of the correlation function $\rho_{k,l} = \alpha^{|k|}\beta^{|l|}$, $(k,l) \in \mathbb{Z}^2$, having reflection symmetric, and the resulting random field then has the second order moments.

Model 2: (see Basu and Reinsel, 1993) Let Y_{t_1,t_2} be a simple spatial autoregressive bidimensional process, expressed as follows:

$$Y_{t_1,t_2} = \alpha Y_{t_1-1,t_2} + \beta Y_{t_1,t_2-1} + \varepsilon_{t_1,t_2} \quad (t_1,t_2) \in \mathbb{Z}^2, \tag{A5.2}$$

which is stationary in the case $|\alpha| + |\beta| < 1$, where ε_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$, is again white noise random field. In this case there is a representation in L_2-sense:

$$Y_{t_1,t_2} = \sum_{(i,j)\in\mathcal{V}_{t_1,t_2}} \left(\begin{array}{c} t_1+t_2-i-j \\ t_1-i \end{array} \right) \alpha^{t_1-i}\beta^{t_2-j}\varepsilon_{i,j} = \sum_{i=0}^{\infty}\sum_{j=0}^{i} \left(\begin{array}{c} i \\ j \end{array} \right) \alpha^i\beta^{i-j}\varepsilon_{t_1-i,t_2-i+j},$$

where $\mathcal{V}_{t_1,t_2} = \{(i,j) \in \mathbb{Z}^2, i \le t_1,\, j \le t_2\}$.

Model 3. We consider a polynomial in a complex plain

$$\phi(Z_1,Z_2) = \sum_{k\in\mathcal{S}}\sum_{l\in\mathcal{S}} a_{k,l} Z_1^k Z_2^l, \quad Z_1, Z_2 \in \mathbb{C}',$$

where $\mathcal{S}$ is a finite set at $\mathbb{Z}^2$, and $a_{k,l}$, $(k,l) \in \mathcal{S}$, are real coefficients such that, for all $(t_1,t_2) \in \mathbb{Z}^2$,

$$\phi(B_1,B_2)Y_{t_1,t_2} = \varepsilon_{t_1,t_2}, \quad (t_1,t_2) \in \mathbb{Z}^2, \tag{A5.3}$$

where the backward-shift operators can be defined as $B_1 Y_{t_1,t_2} = Y_{t_1-1,t_2}$, and $B_2 Y_{t_1,t_2} = Y_{t_1,t_2-1}$. It is known (Guyon, 1995; Rosenblatt, 1985) that equation (A5.3) admits a stationary solution if and only if $\phi(e^{i\lambda_i}, e^{i\lambda_2}) \neq 0$, for all $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [-\pi, \pi]^2$. In this case, the homogeneous fields $\{Y_{t_1,t_2},\ (t_1, t_2) \in \mathbb{Z}^2\}$ may be written as

$$Y_{t_1,t_2} = \sum_{k \in \mathbb{Z}^2} \sum_{l \in \mathbb{Z}^2} C_{k,l} \varepsilon_{t_1-k,t_2-l}, \tag{A5.4}$$

where $C_{k,l}$ are the coefficients involved in the Laurent expansion of $\phi^{-1}(Z_1, Z_2)$ in the neighborhood of $\{|Z_1| = 1, |Z_2| = 1\}$, and satisfy $\sum_k \sum_l |c_{k,l}|^2 < \infty$. Note that (A5.4) converges almost surely and in L_2 (the proof is similar to Brockwell and Davis, 1991, Proposition 3.1.1). The spectral density of stationary field (A5.4) is of the form

$$f_x(\lambda_1, \lambda_2) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} |\phi(e^{i\lambda_1}, e^{i\lambda_2})|^{-2}, \quad (\lambda_1, \lambda_2) \in [-\pi, \pi]^2.$$

Finally, in Ruiz-Medina (2011), the following spatial autoregressive Hilbertian model is introduced:

$$Y_{t_1,t_2} = R + A(Y_{t_1-1,t_2}) + B(Y_{t_1,t_2-1}) + C(Y_{t_1-1,t_2-1}) + \varepsilon_{t_1,t_2}, \tag{A5.5}$$

where $R \in \mathcal{H}$, with $\mathcal{H}$ being a separable Hilbert space, A, B and C are linear bounded operators on $\mathcal{H}$, i.e., A, B, $C \in \mathcal{L}(\mathcal{H})$, and ε is a spatial Hilbert-valued martingale difference process. In this chapter, similar formulations to those given in equations (A5.1)-(A5.3) and (A5.5) are considered, in the introduction of long-range dependence spatial time series models, having singular spectra, in terms of spatial Gegenbauer innovation processes.

A5.3 Long-range dependence

The spatial autoregressive processes $\{Y_{t_1,t_2}\}$ which are usually studied typically have short memory, as mentioned earlier in the Introduction. This means that the covariance function associated with these processes satisfies that $\sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |\gamma_{k,l}| < \infty$. From now on, we will focus our

study in the spatial autoregressive models (SAR), displaying long range dependence (LRD), i.e., in the case where the autocovariance function satisfies that:

$$\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}|\gamma_{k,l}|=\infty.$$

This section discusses and provides the main structural properties of SAR processes with strong spatial dependence.

Definition 8. *Let Y_{t_1,t_2} be a spatial autoregressive model in two-dimensions. From equation (A5.1), we can define the spatial autoregressive process as:*

$$\phi(B_1,B_2)Y_{t_1,t_2}=\varepsilon_{t_1,t_2},\quad (t_1,t_2)\in\mathbb{Z}^2, \tag{A5.6}$$

where ε_{t_1,t_2} is assumed to be a white noise process, with zero-mean and common variance σ_ε^2, and $\phi(B_1,B_2)$ is the autoregressive operator, such that, $\phi(Z_1,Z_2)=(1-\alpha Z_1)(1-\beta Z_2)=1-\alpha Z_1-\beta Z_2+\alpha\beta Z_1Z_2$, for all Z_i, with $i=1,2$, within the closed unit polydisc $|Z_i|\leq 1$, $i=1,2$.

The spatial autoregressive process Y_{t_1,t_2} is said to be casual if it can be expressed as $Y_{t_1,t_2}=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\psi_{k,l}\varepsilon_{t_1-k,t_2-l}$, or equivalently,

$$Y_{t_1,t_2}=(1-\alpha B_1-\beta B_2-\alpha\beta B_1B_2)^{-1}\varepsilon_{t_1,t_2}$$

with $(1-\alpha B_1-\beta B_2-\alpha\beta B_1B_2)\neq 0$. We introduce in the previous spatial model, equation (A5.6), LRD in terms of the binomial difference operator ∇^d, for $d\in(-1/2,1/2)$, given by

$$\nabla^d=(1-B)^d=\sum_{j=0}^{\infty}\pi_jB^j=1+dB+\frac{d(d-1)}{2}B^2+\ldots \tag{A5.7}$$

with $\pi_j=\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}=\prod_{0\leq k\leq j}\frac{k-1-d}{k}$, $\quad j=0,1,2,\ldots,$ and $d\in\left(-\frac{1}{2},\frac{1}{2}\right)$.

The inverse of this backward shift operator is defined as

$$\nabla^{-d}=(1-B)^{-d}=\sum_{j=0}^{\infty}\Psi_jB^j,$$

with $\Psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{0<k\leq j} \frac{k-1+d}{k}$, $j = 0, 1, 2, \ldots$, and $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, see Granger and Joyeux (1980); Hosking(1981); and Brockwell and Davis (1991). Spatial long range dependence requires the introduction of the two-dimensional operator

$$\nabla^{\boldsymbol{d}} = \nabla^{d_1} \circ \nabla^{d_2} = (1 - B_1)^{d_1} \circ (1 - B_2)^{d_2},$$

where now $\boldsymbol{d} = (d_1, d_2) \in (-1/2, 1/2)^2$. The backward-shift operators for each spatial coordinate are given by $B_1 Y_{t_1,t_2} = Y_{t_1-1,t_2}$, and $B_2 Y_{t_1,t_2} = Y_{t_1,t_2-1}$. Similarly, ε_{t_1,t_2} denotes the two-dimensional white noise process, and $B_1 \varepsilon_{t_1,t_2} = \varepsilon_{t_1-1,t_2}$, $B_2 \varepsilon_{t_1,t_2} = \varepsilon_{t_1,t_2-1}$.

Consider now equation (A5.6) (see Boissy *et al.*, 2005; Guo, Lim and Meerschaert, 2009; Beran, Ghosh and Schell, 2009), we can introduce the two-dimensional LRD operator, for $\boldsymbol{d} \in (-1/2, 1/2)^2$, as

$$\phi(B_1, B_2) \nabla^{\boldsymbol{d}} Y_{t_1,t_2} = \varepsilon_{t_1,t_2},$$

for $(t_1, t_2) \in \mathbb{Z}^2$, with $\boldsymbol{d} = (d_1, d_2)$ and the operator $\nabla^{\boldsymbol{d}}$ is defined by its corresponding series representation in (Z_1, Z_2). The coefficients of $(1 - Z_1)^{d_1} \circ (1 - Z_2)^{d_2}$ are in the unit polydisc, with $\{Z_i \in \mathbb{Z};\ |Z_i| < 1,\ i = 1, 2\}$, and denoting, as before, by $\mathbb{Z}$ the set of all integers.

Remark 12. *Let Y_{t_1,t_2} be a spatial long range dependence process, with $d_i \in (-1/2, 1/2)$, $i = 1, 2$, then, there exists a unique stationary solution Y_{t_1,t_2} such that:*

$$Y_{t_1,t_2} = \nabla^{-\boldsymbol{d}} \varepsilon_{t_1,t_2}.$$

The autocorrelation function of spatial autoregressive process Y_{t_1,t_2} is then given by (see Brockwell and Davis, 1991, and Andel, 1986):

$$\rho(k, l) = \frac{\Gamma(1 - d_1)\Gamma(k + d_1)}{\Gamma(d_1)\Gamma(k + 1 - d_1)} \frac{\Gamma(1 - d_2)\Gamma(l + d_2)}{\Gamma(d_2)\Gamma(l + 1 - d_2)}.$$

In addition, if we consider the autoregressive operator in model (A5.6), we obtain:

$$f(\boldsymbol{\lambda}) = f(\lambda_1, \lambda_2) = \frac{\sigma_\varepsilon^2}{4\pi^2} \frac{1}{|\phi(e^{-i\lambda_1}, e^{-i\lambda_2}, \alpha, \beta)|^2} \frac{1}{|1 - e^{-i\lambda_1}|^{2d_1}} \frac{1}{|1 - e^{-i\lambda_2}|^{2d_2}}$$
$$= \frac{\sigma_\varepsilon^2}{4\pi^2} \frac{1}{|2sin(\lambda_1/2)|^{2d_1}} \frac{1}{|2sin(\lambda_2/2)|^{2d_2}} \frac{1}{|\phi(cos(\lambda_1), cos(\lambda_2), \alpha, \beta)|^2},$$

where $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, $\boldsymbol{d} \in (-1/2, 1/2)$ and $|\alpha| < 1$ and $|\beta| < 1$. Whereas it has been taken into account the following relationship:

$$|1 - e^{-i\lambda}| = 2\,(1 - \cos(\lambda)) = 4\,(\sin(\lambda/2))^2 = 4\left(\sin^2(\lambda/2)\right).$$

Remark 13. *The spectral density $f(\lambda_1, \lambda_2)$ has singular properties for $(\lambda_1, \lambda_2) \to (0, 0)$. Obviously, we have $f(\lambda_1, \lambda_2) \longrightarrow \infty$, for $(\lambda_1, \lambda_2) \longrightarrow (0, 0)$, if only if, $|d_i| < 1/2$, $i = 1, 2$.*

From the above expression of the spectral density, the autocovariance function of the spatial autoregressive long-range dependence process is given by:

$$\begin{aligned}
\gamma(k, l) &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ik\lambda_1} e^{il\lambda_2} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \\
&= \frac{\sigma_\varepsilon^2}{4\pi^2} \int_0^{\pi} \int_{-\pi}^{\pi} e^{i(k\lambda_1 + l\lambda_2)} \frac{1}{|1-e^{-i\lambda_1}|^{2d_1} |1-e^{-i\lambda_2}|^{2d_2}} \frac{1}{|\phi(e^{-i\lambda_1}, e^{-i\lambda_2})|^2}\, d\lambda_1 d\lambda_2 \\
&= \sigma_\varepsilon^2 \frac{(-1)^{k+l} \Gamma(1-2d_1) \Gamma(1-2d_2)}{\Gamma(k-d_1+1) \Gamma(1-k-d_1) \Gamma(l-d_2+1) \Gamma(1-l-d_2)},
\end{aligned}$$

or equivalently, taking into account the relationship $|1 - e^{-i\lambda}| = 2(1 - \cos\lambda) = (2\sin(\lambda/2))^2$ showed above, we can re-write the autocovariance function as

$$\gamma(k, l) = \frac{\sigma_\varepsilon^2}{4\pi^2} \int_0^{\pi} \int_{-\pi}^{\pi} \frac{1}{\left|2\sin\left(\frac{\lambda_1}{2}\right)\right|^{2d_1} \left|2\sin\left(\frac{\lambda_2}{2}\right)\right|^{2d_2}} \frac{1}{|\phi(\cos(k\lambda_1), \cos(l\lambda_2))|^2}\, d\lambda_1 d\lambda_2,$$

where $k = 0, 1, 2, \ldots$, $l = 0, 1, 2, \ldots$, and $d_i \in (-1/2, 1/2)$, $i = 1, 2$.

Remark 14. *We say that spatial process Y_{t_1,t_2} is a long-memory spatial autoregressive process for $d_i \in (-1/2,\, 1/2), i = 1, 2,$ if $\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\rho(k, l)| = \infty$.*

Remark 15. *The second-order process $Y_{t_1,t_2} : t_1, t_2 \in \mathbb{Z}\}$ is said to be stationary when $E(Y_{t_1,t_2}) = \mu$, for all $(t_1, t_2) \in \mathbb{Z}$, and, for each $(k, l) \in \mathbb{Z}$, $Cov(Y_{t_1,t_2}, Y_{t_1+k,t_2+l})$ is independent of $(t_1, t_2) \in \mathbb{Z}$.*

Remark 16. *If $d_i \in (-1/2, 1/2)$, $i = 1, 2$, then, the solution $\{Y_{t_1,t_2} : t_1, t_2 \in \mathbb{Z}\}$ is casual if and only if $\phi(Z_1, Z_2) \neq 0$ for $|Z_i| \leq 1$, $i = 1, 2$.*

Applying Sterling's formula $\Gamma(\lambda_i) = \sqrt{2\pi}e^{-\lambda_i+1}(\lambda_i - 1)^{\lambda_i - \frac{1}{2}}$, $i = 1, 2$, with $\lambda_i \longrightarrow \infty$ $i = 1, 2$, then we can re-write Π_j and Ψ_j as:

$$\Pi_j \simeq \frac{j^{-d-1}}{\Gamma(-d)}, \quad \text{and } \Psi_j \simeq \frac{j^{d-1}}{\Gamma(d)}, \qquad j \to \infty,$$

and therefore, the autocorrelation function is defined as:

$$\rho(\lambda_1, \lambda_2) = \lambda_1^{2d_1-1}\lambda_2^{2d_2-1}\frac{\Gamma(d_1 - 1)}{\Gamma(d_1)}\frac{\Gamma(d_2 - 1)}{\Gamma(d_2)}, \quad \lambda_i \to \infty,\ d_i \neq 0, \quad i = 1, 2.$$

If we assume that $\sin(\lambda) \sim \lambda$ with $\lambda_i \longrightarrow 0$ $i = 1, 2$, then we can write now the spectral density as

$$f(\lambda_1, \lambda_2) \simeq \frac{\sigma_\varepsilon^2}{4\pi^2}\frac{1}{\lambda_1^{2d_1}\lambda_2^{2d_2}}, \quad \lambda_i \longrightarrow 0, \quad i = 1, 2, \tag{A5.8}$$

which, for $\lambda_i = 0$, $i = 1, 2$, is finite if and only if $d_1 \leq 0$ and $d_2 \leq 0$.

A5.4 Gegenbauer white noise

In this section, Gegenbauer polynomials in two dimensions are considered to introduce long range dependence in the spatial case (see, for example, Chung, 1996a; Gray, Zhang and Woodward, 1989; and Woodward, Cheng and Gray, 1998). Note that Gegenbauer polynomials constitute a class of orthogonal polynomials, widely considered in applied mathematics, due to their properties of recursion and orthogonality.

Definition 9. *Let $d \neq 0$ and $|Z| < 1$; then, for $|u| \leq 1$, we define the Gegenbauer polynomials $C_n^{(d)}(u)$ by*

$$\left(1 - 2uZ + Z^2\right)^{-d} = \sum_{n=0}^{\infty} C_n^{(d)}(u)Z^n, \tag{A5.9}$$

where $C_n^{(d)}(u)$ *can be expressed as*

$$C_n^{(d)}(u) = \sum_{k=0}^{[n/2]} \frac{(-1)^k (d)_{n-k} (2u)^{n-2k}}{k!(n-2k)!}, \tag{A5.10}$$

with parameter $(d)_n$ *being defined as* $(d)_n = \frac{\Gamma(d+n)}{\Gamma(d)}$.

Following Chung (1996a) and Gray, Zhang and Woodward (1989), to develop the Gegenbauer model a more general fractional operator is needed. This operator is defined from equations (A5.9) and (A5.10)). The easiest way to compute $C_n^{(d)}(u)$ comes from the following recursion formula:

$$C_n^{(d)}(u) = 2u\left(\frac{d-1}{n}+1\right) C_{n-1}^{(d)}(u) - \left(\frac{2(d-1)}{n}+1\right) C_{n-2}^{(2)}(u),$$

where $C_0^{(d)}(u) = 1$, $C_1^{(d)}(u) = 2du$, $C_2^{(d)}(u) = 2d(d+1)u^2$. Also it is known that

$$C_n^{(d)}(u) \sim \frac{\cos[(n+d)\nu - (\frac{d}{2}\pi)]}{\Gamma(d) sin^d(\nu)} \left(\frac{2}{n}\right)^{1-d},$$

as $n \longrightarrow \infty$, and $\nu = cos^{-1}(u) = arcos(u)$. Let us consider Y_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$, be a process defined on discrete space. Let us assume that its autocovariance function satisfies $\sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |\gamma_{k,l}| = \infty$. We consider the backward-shift operator defined as

$$\nabla_u^d = (I - 2uB + B^2)^d = (1 - 2\cos\nu B + B^2)^d = [(1 - e^{i\nu}B)(1 - e^{-i\nu}B)]^d, \tag{A5.11}$$

for $u = \cos\nu$, or $\nu = \cos^{-1}(u)$, and $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. The following state equation is assumed to be satisfied by Y, which is referred as spatial Gegenbauer white noise:

$$\nabla_{u_1}^{d_1} \circ \nabla_{u_2}^{d_2} Y_{t_1,t_2} = \left(I - 2u_1 B_1 + B_1^2\right)^{d_1} \circ \left(I - 2u_2 B_2 + B_2^2\right)^{d_2} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}, \quad (t_1,t_2) \in \mathbb{Z}^2, \tag{A5.12}$$

where $\nabla_{u_i}^{d_i}$, $i = 1,2$, is given in equation (A5.11), with B_i, for $i = 1,2$, denoting the backward-shift operators for each spatial coordinate, that is, $B_1 Y_{t_1,t_2} = Y_{t_1-1,t_2}$, and $B_2 Y_{t_1,t_2} = Y_{t_1,t_2-1}$. Here, ε_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$ is assumed to be white noise process with zero-mean $E[\varepsilon_{t_1,t_2}] = 0$, and

common variance $E[\varepsilon_{t_1,t_2}, \varepsilon_{s_1,s_2}] = \sigma_\varepsilon^2$. From equation (A5.12), the spatial Gegenbauer process can be defined in terms of the inverse of operator $\nabla_{u_1}^{d_1} \circ \nabla_{u_2}^{d_2}$ expanded in a Gegenbauer polynomial series as follows:

$$\begin{aligned} Y_{t_1,t_2} &= \nabla_{u_2}^{-d_2} \circ \nabla_{u_1}^{-d_1} \varepsilon_{t_1,t_2} = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)\, C_{n_2}^{(d_2)}(u_2) B_1^{n_1} B_2^{n_2} \varepsilon_{t_1,t_2} \\ &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)\, C_{n_2}^{(d_2)}(u_2) \varepsilon_{t_1-n_1,t_2-n_2}. \end{aligned}$$

The spectral density of a Gegenbauer white noise (see Chung, 1996a, and Hsu and Tsai, 2009, for the one-parameter case) is given by

$$\begin{aligned} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \tfrac{\sigma_\varepsilon^2}{(2\pi)^2} \left|1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}\right|^{-2d_1} \left|1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}\right|^{-2d_2} \\ &= \tfrac{\sigma_\varepsilon^2}{(2\pi)^2} \left\{|2\cos\lambda_1 - 2u_1|\right\}^{-2d_1} \left\{|2\cos\lambda_2 - 2u_2|\right\}^{-2d_2}, \end{aligned}$$

where $u_i = \cos\nu_i$, and $-\pi \le \lambda_i \le \pi$, $i = 1, 2$. From the spectral density function (A5.13), we can compute the auto-covariance function values as follows:

$$\begin{aligned} \gamma(j_1, j_2) &= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{ij_1\lambda_1} e^{ij_2\lambda_2} f(\lambda_1, \lambda_2) d\lambda_1 d\lambda_2 \\ &= \frac{\sigma_\varepsilon^2}{(2\sqrt{\pi})^2} \prod_{i=1}^{2} \Gamma(1 - 2d_i) [2\sin(\nu_i)]^{\frac{1}{2} - 2d_1} \left[P_{j_i - \frac{1}{2}}^{2d_i - \frac{1}{2}}(u_i) + (-1)^{j_i} P_{j_i - \frac{1}{2}}^{2d_i - \frac{1}{2}}(-u_i) \right] \end{aligned} \quad \text{(A5.13)}$$

with $P_a^b(z)$ being the Legendre functions with initial values

$$P_{-\frac{1}{2}}^{2d_i - \frac{1}{2}}(u_i) = \left(\frac{1+u_i}{1-u_i}\right)^{d_i - \frac{1}{4}} \frac{1}{\Gamma(\frac{3}{2} - 2d_i)} \sum_{n=1}^{\infty} \frac{\Gamma\left(\frac{3}{2} - 2d_i\right) \Gamma\left(\frac{1}{2} + n\right) \Gamma\left(\frac{1}{2} + n\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{3}{2} - 2d_i + n\right) \Gamma(n+1)} \left(\frac{1-u_i}{2}\right)^n. \quad \text{(A5.14)}$$

$$P_{-\frac{1}{2}}^{2d_i - \frac{1}{2}}(-u_i) = \left(\frac{1+u_i}{1-u_i}\right)^{d_i - \frac{1}{4}} \frac{1}{\Gamma(\frac{3}{2} - 2d_i)} \sum_{n=1}^{\infty} \frac{\Gamma\left(\frac{3}{2} - 2d_i\right) \Gamma\left(\frac{-1}{2} + n\right) \Gamma\left(\frac{3}{2} + n\right)}{\Gamma\left(\frac{-1}{2}\right) \Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{3}{2} - 2d_i + n\right) \Gamma(n+1)} \left(\frac{1-u_i}{2}\right)^n. \quad \text{(A5.15)}$$

From Chung (1996a), Gray, Zhang and Woodward (1989) and Gradshteyn and Ryzhik (1980), the following asymptotic approximation of the autocovariance function is obtained:

$$\gamma(j_1, j_2) = \prod_{i=1}^{2} \frac{2^{1-d_i} \sigma_\epsilon^2}{\pi} \sin^{-2d_i}(\nu_i) \sin(d_i\pi) \Gamma(1 - 2d_i) \cos(j_i\nu_i) \frac{\Gamma(j_i + 2d_i)}{\Gamma(j_i + 1)} [1 + \mathcal{O}(j_i^{-1})].$$

Theorem 3. *The following assertions hold:*

1. *A Gegenbauer white noise process is stationary*

   1.1 *if* $|u_i| < 1$ *and* $d_i < 1/2, \quad i = 1, 2,$

   1.2 *or if* $|u_i| = 1$, *and* $d_i < 1/4, \quad i = 1, 2.$

2. *A Gegenbauer white noise process is invertible*

   2.1 *if* $|u_i| < 1$, *and* $d_i > -1/2, \quad i = 1, 2$

   2.2 *or if* $|u_i| = 1$ *and* $d_i > -1/4, \quad i = 1, 2.$

Proof

1 We started this show with stationarity of Gegenbauer process.

1.1 In this point, we consider that $|u_i| < 1$. Then,

$$C_{n_i}^{d_i}(u_i) = \frac{\Gamma(d_i + \frac{1}{2})\Gamma(2d_i + n_i)}{n_i!\Gamma(2d_i)} \left(\frac{1 - u_i^2}{4}\right)^{\frac{1}{4} - \frac{d_i}{2}} P_{n_i + d_i - \frac{1}{2}}^{(\frac{1}{2} - d_i)}(u_i), \quad i = 1, 2,$$

where $P_\alpha^\beta(u)$ is a Legendre given by the equation (A5.14) and (A5.15). Now, for fixed u_i and $\beta_i = 1/2 - d_i$, and $\alpha_i = n_i + d_i - 1/2$, the following asymptotic result holds, when $|u_i| < 1$, and $i = 1, 2$,

$$P_{\alpha_i}^{\beta_i}(u_i) = \frac{2}{\pi^{1/2}} \frac{\Gamma(\alpha_i + \beta_i + 1)}{\Gamma(\alpha_i + \frac{3}{2})} \frac{\cos\left\{(\alpha_i + \frac{1}{2})\phi - \frac{\pi}{4} + \beta_i \frac{\pi}{2}\right\}}{\sqrt{2 \sin \phi}} \left\{1 + \mathcal{O}\left(\frac{1}{\alpha_i}\right)\right\},$$

where $\phi = \cos^{-1}(u_i)$. Then, if we substitute α_i and β_i, P can be expressed as

$$\begin{aligned} P_{n_i + d_i - \frac{1}{2}}^{(\frac{1}{2} - d_i)}(u_i) &= \frac{2}{\sqrt{\pi}} \frac{\Gamma\big((n_i + d_i - \frac{1}{2}) + (\frac{1}{2} - d_i) + 1\big)}{\Gamma\big((n_i + d_i - \frac{1}{2}) + \frac{3}{2}\big)} \frac{\cos\{\big((n_i + d_i - \frac{1}{2}) + \frac{1}{2}\big)\phi - \frac{\pi}{4} + (\frac{1}{2} - d_i)\frac{\pi}{2}\}}{\sqrt{2 \sin \phi}} \left\{1 + \frac{\mathcal{O}}{(n_i + d_i - \frac{1}{2})}\right\} \\ &= \frac{2}{\sqrt{\pi}} \frac{\Gamma(n_i + 1)}{\Gamma(n_i + d_i + 1)} \frac{\cos\{(n_i + d_i)\phi - d_i \frac{\pi}{2}\}}{\sqrt{2 \sin \phi}} \left\{1 + \frac{\mathcal{O}}{(n_i + d_i - \frac{1}{2})}\right\}, \quad i = 1, 2. \end{aligned}$$

Therefore with $i = 1, 2$,

$$C_{n_i}^{(d_i)}(u_i) = \frac{2}{\sqrt{\pi}} \frac{\Gamma(d_i + \frac{1}{2})}{\Gamma(n_i + d_i + 1)} \frac{\Gamma(2d_i + n_i)}{\Gamma(2d_i)} \left(\frac{1 - u_i^2}{4}\right)^{\frac{1}{4} - \frac{d_i}{2}} \frac{\cos\{(n_i + d_i)\phi - d_i \frac{\pi}{2}\}}{\sqrt{2 \sin \phi}} \left\{1 + \frac{\mathcal{O}}{(n_i + d_i - \frac{1}{2})}\right\}.$$

Then by Sterling's formula

$$C_{n_i}^{(d_i)}(u) \simeq \frac{2}{\pi^{1/2}} \frac{\Gamma(n_i+1)\Gamma(d_i+\frac{1}{2})}{\Gamma(2d_i)} \left(\frac{1-u_i^2}{4}\right)^{1/4-d_i/2} \cos\left((n_i+d_i)\phi - d_i\pi/2\right) n_i^{d_i-1}, \tag{A5.16}$$

as $n_i \longrightarrow \infty$. Now

$$\gamma(0) = \sum_{k=0}^{\infty} \{C_k^d(u)\}^2 \sigma_\varepsilon^2. \tag{A5.17}$$

However, from (A5.16), when $|u_i| < 1$, $i = 1, 2$, the series in (A5.17) converges if $d_i < \frac{1}{2}$, $i = 1, 2$ which completes the proof for the case $|u_i| < 1$, $i = 1, 2$.

1.2 The case $u_i = 1$ was shown by Hosking (1981), and $u_i = -1$ follows similarly with $i = 1, 2$.

2 Now, we prove the condition of invertibility for this type of processes.

2.1 From equation (A5.12), it can be seen that Y_{t_1,t_2} can be written in the form

$$Y_{t_1,t_2} = \varepsilon_{t_1,t_2} - \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} C_{k,l}^{(-d_i)}(u_i) Y_{t_1-k,t_2-l}$$

if $|u_i| < 1$ and $d_i > -\frac{1}{2}$ or if $d_i > -\frac{1}{4}$ when $u_i = \pm 1$, and thus the result follows. □

Remark 17. *A stationary Gegenbauer white noise process is long memory*

- *if* $0 < d_i < \frac{1}{2}$ *and* $|u_i| < 1$, $i = 1, 2$,
- *or if* $0 < d_i < \frac{1}{4}$ *and* $|u_i| = 1$, $i = 1, 2$.

Theorem 4. *Let* Y_{t_1,t_2} *be a stationary long-memory Gegenbauer process given as in equation (A5.12), i.e.* $\left(I - 2u_1B_1 + B_1^2\right)^{d_1} \circ \left(I - 2u_2B_2 + B_2^2\right)^{d_2} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}$, $(t_1, t_2) \in \mathbb{Z}^2$.

- *When* $u_i = 1$ *and* $0 < d_i < \frac{1}{4}$, $i = 1, 2$, *the autocorrelation function of* Y_{t_1,t_2} *is*

$$\rho(k,l) = \frac{\Gamma(1-2d_1)\Gamma(k+2d_1)}{\Gamma(2d_1)\Gamma(k-2d_1+1)} \frac{\Gamma(1-2d_2)\Gamma(l+2d_2)}{\Gamma(2d_2)\Gamma(l-2d_2+1)},$$

As $k \to \infty$ *and* $l \to \infty$, *then we re-write* $\rho(k,l) \simeq k^{4d_1-1} l^{4d_2-1}$.

- *When $u_i = -1$ and $0 < d_i < \frac{1}{4}$, $i = 1, 2$, the autocorrelation function of Y_{t_1,t_2} is given by*

$$\rho(k,l) = (-1)^{k+l} \frac{\Gamma(1-2d_1)\Gamma(k+2d_1)}{\Gamma(2d_1)\Gamma(k-2d_1+1)} \frac{\Gamma(1-2d_2)\Gamma(l+2d_2)}{\Gamma(2d_2)\Gamma(l-2d_2+1)},$$

As $k \to \infty$ and $l \to \infty$, then we re-write $\rho(k,l) \simeq (-1)^{k+l} k^{4d_1-1} l^{4d_2-1}$.

- *When $|u_i| < 1$ and $0 < d_i < \frac{1}{2}$, $i = 1, 2$, the autocorrelation function of Y_{t_1,t_2} is*

$$\rho(k,l) \simeq k^{2d_1-1} l^{2d_2-1} \sin(\pi d_1 - k u_1) \sin(\pi d_2 - l u_2),$$

as $k \to \infty$, $l \to \infty$, with u_1 and u_2 being the Gegenbauer spatial frequencies.

Proof

(i) Since the case $u_i = 1$, with $i = 1, 2$, corresponds to the fractional process, the result has been shown by Hosking (1981).

(ii) When $u_i = -1$, with $i = 1, 2$, the Gegenbauer process becomes $(1+B_1)^{2d_1} \circ (1+B_2)^{2d_2} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}$, with $(t_1, t_2) \in \mathbb{Z}$, i.e., $Y_{t_1,t_2} = \left((1+B_1)^{-2d_1} \circ (1+B_2)^{-2d_2}\right) \varepsilon_{t_1,t_2}$, with $(t_1, t_2) \in \mathbb{Z}$. The ψ weights in the general linear process form

$$Y_{t_1,t_2} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \psi(k,l) \varepsilon_{t_1-k,t_2-l}$$

$\psi(k,l)$ can be written as

$$\psi(k,l) = (-1)^{k+l} \frac{\Gamma(k+2d_1)}{k!\Gamma(2d_1)} \frac{\Gamma(l+2d_2)}{l!\Gamma(2d_2)} = (-1)^{k+l} \eta(k)\eta(l),$$

where $\eta(k)$ and $\eta(l)$ are the ψ weights found in Hosking (1981), in the case $u_i = 1$, with $i = 1, 2$, and the result follows from the proof provided in this paper about this identity.

(iii) The following lemma is needed for this proof (see Gray, Zhang and Woodward, 1989).

Lemma 1. *Let $R(\tau) = \int_0^{\pi} P(\omega) \cos(\tau\omega) d\omega$, with τ an integer. Let $\omega_0 \in (0, \pi)$, and suppose that $P(\omega)$ can be expressed as $P(\omega) = b(\omega)|\omega - \omega_0|^{-\beta}$, with $0 < \beta < 1/2$. Further, suppose*

that $b(\omega)$ is nonnegative and of bounded variation in $(0, \omega_0 - \varepsilon) \cup (\omega_0 + \varepsilon, \pi)$, for $\varepsilon > 0$. Suppose also that $b(\omega)$ is slowly varying at ω_0. Then, when $\tau \longrightarrow \infty$,

$$R(\tau) \simeq \tau^{\beta-1} \sin\left(\frac{\pi\beta}{2} - \tau\omega_0\right) \left\{ b_1\left(\frac{1}{\tau}\right) + b_2\left(\frac{1}{\tau}\right) \right\},$$

where $b_1(x) = b(x + \omega_0)$ and $b_2(x) = b_1(-x)$.

For the spatial Gegenbauer process with $|u_i| < 1$, $i = 1, 2$, the spectrum is given by

$$f(\lambda_1, \lambda_2) = \sigma_\varepsilon^2 \left\{4(\cos(\lambda_1) - u_1)^2\right\}^{-d_1} \left\{4(\cos(\lambda_2) - u_2)^2\right\}^{-d_2}.$$

Considering $\nu_i = \cos^{-1}(u_i)$, $i = 1, 2$, we get

$$P(\lambda_1, \lambda_2) \quad = \sigma_\varepsilon^2 \left\{2\sin^2\left(\tfrac{\lambda_1-\nu_1}{2}\right)\sin^2\left(\tfrac{\lambda_1+\nu_1}{2}\right)\right\}^{-d_1} \left\{2\sin^2\left(\tfrac{\lambda_2-\nu_2}{2}\right)\sin^2\left(\tfrac{\lambda_2+\nu_2}{2}\right)\right\}^{-d_2}.$$

Then, we can write $P(\lambda_1, \lambda_2) = \prod_{i=1}^2 |\lambda_i - \nu_i|^{-2d_i} b(\lambda_i)$, where, for $i = 1, 2$,

$$b(\lambda_i) = \sigma_\varepsilon^2 2^{-d_i} \left\{\sin^2\left(\frac{\lambda_i + \nu_i}{2}\right)\right\}^{-d_i} \left(\left[\frac{\sin((\lambda_i - \nu_i)/2)}{\lambda_i - \nu_i}\right]^2\right)^{-d_i}.$$

The rest of the proof follows from the fact that function b satisfies the conditions required in Lemma 1. □

A5.5 Autoregressive Gegenbauer random fields

In the definition given in the previous section about spatial Gegenbauer processes, we now incorporate the autoregressive part. That is, we consider the following model (see Chung, 1996a): For $(t_1, t_2) \in \mathbb{Z}^2$,

$$\phi(B_1, B_2)(I - 2u_1B_1 + B_1^2)^{d_1}(I - 2u_2B_2 + B_2^2)^{d_2}(Y_{t_1,t_2} - \mu) = \varepsilon_{t_1,t_2}, \tag{A5.18}$$

where ε_{t_1,t_2} is white noise with zero-mean and variance σ^2, and the operators B_1 and B_2 are the spatial lag operators. As usual, the parameter μ denotes the mean of the process. As before,

$|u_i| \leq 1$, $|d_i| < 1/2$, $i = 1, 2$. We pay special attention to the following particular cases of ϕ polynomial:

Model 1: $\phi(B_1, B_2) = 1 - \alpha B_1 - \alpha B_2 + \alpha\beta B_1 B_2, \quad -1 \leq \alpha, \beta \leq 1.$

Model 2: $\phi(B_1, B_2) = 1 - \alpha B_1 - \beta B_2, \quad |\alpha| + |\beta| < 1.$

In the computation of the autocovariance function, we will focus on Model 2:

$$\phi(B_1, B_2)(Y_{t_1,t_2} - \mu) = \varepsilon_{t_1,t_2},$$

where, as before, $\phi(B_1, B_2) = 1 - \alpha B_1 - \beta B_2$.

We then can define our Gegenbauer autoregressive model with long memory in two-dimensions, as the previous model (A5.18), considering the following two steps: First, the inverse of the autoregressive operator leads to the moving average representation of infinite order given by:

$$\begin{aligned} Y_{t_1,t_2} &= \mu + \sum_{i=0}^{\infty} \sum_{j=0}^{i} \binom{i}{j} \alpha^i \beta^{i-j} \varepsilon_{t_1-j,t_2-i+j} \\ &= \mu + \sum_{(i,j)\in\nu_{t_1,t_2}} \binom{t_1+t_2-i-j}{t_1-i} \alpha^{t_1-i} \beta^{t_2-j} \varepsilon_{i,j}, \end{aligned} \tag{A5.19}$$

where $\nu_{t_1,t_2} = \{(i,j) \in \mathbb{Z}^2, \quad i \leq t_1, \ j \leq t_2\}$. After incorporation of the fractional order operators, the inversion of the autoregressive part is first performed:

$$\phi(B_1, B_2) \nabla_{u_1}^{d_1} \nabla_{u_2}^{d_2} (Y_{t_1,t_2} - \mu) = \varepsilon_{t_1,t_2}$$

and

$$\begin{aligned} \nabla_{u_1}^{d_1} \nabla_{u_2}^{d_2} (Y_{t_1,t_2} - \mu) &= \phi^{-1}(B_1, B_2) \varepsilon_{t_1,t_2} \\ &= \sum_{(i,j)\in\nu_{t_1,t_2}} \binom{t_1+t_2-i-j}{t_1-i} \alpha^{t_1-i} \beta^{t_2-j} \varepsilon_{i,j}. \end{aligned} \tag{A5.20}$$

The inversion of the fractional-order filter, in terms of Gegenbauer polynomials, then leads

to the following expression:

$$\begin{aligned}
Y_{t_1,t_2} &= \mu + \nabla_{u_1}^{-d_1}\nabla_{u_2}^{-d_2} \sum_{(i,j)\in\nu_{t_1,t_2}} \left(\begin{array}{c} t_1+t_2-i-j \\ t_1-i \end{array} \right) \alpha^{t_1-i}\beta^{t_2-j}\varepsilon_{i,j} \\
&= \mu + \sum_{(i,j)\in\nu_{t_1,t_2}} \left(\begin{array}{c} t_1+t_2-i-j \\ t_1-i \end{array} \right) \alpha^{t_1-i}\beta^{t_2-j}\nabla_{u_1}^{-d_1}\nabla_{u_2}^{-d_2}\varepsilon_{i,j} \\
&= \mu + \sum_{(i,j)\in\nu_{t_1,t_2}} \left(\begin{array}{c} t_1+t_2-i-j \\ t_1-i \end{array} \right) \alpha^{t_1-i}\beta^{t_2-j} \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)C_{n_2}^{(d_2)}(u_2) \\
&\qquad \times B_1^{n_1}B_2^{n_2}\varepsilon_{t_1,t_2} \\
&= \mu + \sum_{(i,j)\in\nu_{t_1,t_2}} \left(\begin{array}{c} t_1+t_2-i-j \\ t_1-i \end{array} \right) \alpha^{t_1-i}\beta^{t_2-j} \sum_{n_1=0}^{\infty}\sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)C_{n_2}^{(d_2)}(u_2) \\
&\qquad \times \varepsilon_{t_1-n_1,t_2-n_2} \qquad \text{(A5.21)}
\end{aligned}$$

If we assume $\mu = 0$ in both models, the spectral density of the stationary solution is then given by

$$f(\lambda_1,\lambda_2) = \frac{\sigma_\varepsilon^2}{(2\pi)^2}\frac{1}{|\phi(e^{i\lambda_1},e^{i\lambda_2})|^2} f_G(\lambda_1,\lambda_2),$$

where

$$f_G(\lambda_1,\lambda_2) = \frac{1}{(2\pi)^2}\,\frac{1}{|2\cos(\lambda_1-u_1)|^{2d_1}\,|2\cos(\lambda_2-u_2)|^{2d_2}}.$$

The autocovariance function is then obtained from the identity

$$\begin{aligned}
\gamma(k,l) &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{ik\lambda_1}e^{il\lambda_2} f(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2 \\
&= 4\int_0^{\pi}\int_0^{\pi} \frac{\sigma_\varepsilon^2}{4\pi^2}\frac{\cos(k\lambda_1)\cos(l\lambda_2)}{|\phi(\cos(\lambda_1),\cos(\lambda_2))|^2}\frac{1}{|2\cos(\lambda_1-u_1)|^{2d_1}|2\cos(\lambda_2-u_2)|^{2d_2}}d\lambda_1 d\lambda_2.
\end{aligned}$$

Note that the autocovariance function can be approximated by

$$\gamma(k,l) \simeq A\cos(k\nu_1)k^{2d_1-1}\cos(l\nu_2)l^{2d_2-1},$$

when $k\to\infty$ and $l\to\infty$, with $\nu_i = \cos^{-1}(u_i)\in[0,\pi]$, $i=1,2$, being the Gegenbauer frequency, and A being a nonzero constant independent of k and l. In the range $0<d_i<1/2$, $i=1,2$, we

obtain the definition of a long memory process. The spectral density of the Gegenbauer process takes its values around the Gegenbauer frequency ν_i, $i = 1, 2$,

$$f(\lambda_1, \lambda_2) \simeq |\lambda_1 - \nu_1|^{-2d_1} |\lambda_2 - \nu_2|^{-2d_2},$$

as $\lambda_1 \to \infty$, and $\lambda_2 \to \infty$. For $0 < d_i < 1/2$, and $\nu_i = cos^{-1}(u_i) \in [0, \pi]$, $i = 1, 2$, the spectral density is unbounded, as $\lambda_1 \to \nu_1$ and $\lambda_2 \to \nu_2$. The spectral density of the autoregressive Gegenbauer process also admits the following expression in terms of the Gegenbauer frequency, $\nu_i = \cos^{-1}(u_i)$, $i = 1, 2$, given as above:

$$f(\lambda_1, \lambda_2) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} \frac{\left|1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}\right|^{-2d_1}}{\left|\phi(e^{-i\lambda_1})\right|^2} \frac{\left|1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}\right|^{-2d_2}}{\left|\phi(e^{-i\lambda_2})\right|^2}$$

with $0 \leq \lambda_1 \leq \pi$ and $0 \leq \lambda_2 \leq \pi$.

A5.6 Gegenbauer fields with multiple singularities

Inspired by the work of Gray, Zhang and Woodward (1989) and Woodward, Cheng and Gray (1998), we establish the formulation of spatial Gegenbauer autoregressive models with multiple singularities (GARMA models), in terms of $k - 1$ factors, as follows:

$$\phi(B_1, B_2) \prod_{j=1}^{k-1} \left(I - 2B_1 \cos(\mu_1^j) + B_1^2\right)^{d_{j1}} \prod_{j=1}^{k-1} \left(I - 2B_2 \cos(\mu_2^j) + B_2^2\right)^{d_{j2}} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}. \quad \text{(A5.22)}$$

First, we consider the spatial $k - 1$ factor Gegenbauer process

$$\begin{aligned} Y_{t_1,t_2} &= \prod_{j=1}^{k-1} \left(I - 2B_1 \cos(\mu_1^j) + B_1^2\right)^{-d_{j1}} \prod_{j=1}^{k-1} \left(I - 2B_2 \cos(\mu_2^j) + B_2^2\right)^{-d_{j2}} \varepsilon_{t_1,t_2} \\ &= \prod_{j=1}^{k-1} \left(\sum_{i=0}^{\infty} C_{1i}^{d_{j1}}(u_{1j}) B_1^i\right) \left(\sum_{i=0}^{\infty} C_{2i}^{d_{j2}}(u_{2j}) B_2^i\right) \varepsilon_{t_1,t_2} \\ &= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \psi_{m_1} B_1^{m_1} \psi_{m_2} B_2^{m_2} \varepsilon_{t_1,t_2}. \end{aligned} \quad \text{(A5.23)}$$

Model (A5.22) then has spectral density given by

$$
\begin{aligned}
f_x(\lambda_1,\lambda_2) &= \sigma_\varepsilon^2 \left|\phi(e^{i2\pi\lambda_1}, e^{i2\pi\lambda_2})\right|^{-2} \left|\prod_{j=1}^{k-1}\{1-2u_{j1}e^{-i2\pi\lambda_1}+e^{i4\pi\lambda_1}\}^{-d_{j1}}\right|^2 \\
&\quad \times \left|\prod_{j=1}^{k-1}\{1-2u_{j2}e^{-i2\pi\lambda_2}+e^{i4\pi\lambda_2}\}^{-d_{j2}}\right|^2 \\
&= \sigma_\varepsilon^2 \left|\phi(e^{i2\pi\lambda_1}, e^{i2\pi\lambda_2})\right|^{-2} \prod_{j=1}^{k-1}\left[4\{\cos(2\pi\lambda_1)-u_{j1}\}^2\right]^{-d_{j1}} \\
&\quad \times \prod_{j=1}^{k-1}\left[4\{\cos(2\pi\lambda_2)-u_{j2}\}^2\right]^{-d_{j2}}, \qquad \text{(A5.24)}
\end{aligned}
$$

where $\nu_{ij} = \frac{\cos^{-1}(u_{ij})}{2\pi}$, with $i = 1, 2$, and $j = 1, 2, \ldots, k-1$.

Theorem 5. *The spatial $k-1$ factor Gegenbauer process is stationary if u_{j1} and u_{j2} do not coincide, for $j = 1, \ldots, k-1$:*

- $d_{ji} < \frac{1}{2}$, *for* $|u_{ji}| \neq 1$, $i = 1, 2$, *and* $j = 1, 2, \ldots, k-1$.
- $d_{ji} < \frac{1}{4}$, *for* $|u_{ji}| = 1$, $i = 1, 2$, *and* $j = 1, 2, \ldots, k-1$.

Proof (see Woodward, Cheng and Gray, 1998, p. 502) In order to prove stationarity, we must show that $\gamma_{0,0} < \infty$, with

$$
\gamma_{0,0} = 2\sigma_\varepsilon^2 \int_0^{1/2}\int_0^{1/2} \prod_{j=1}^{k-1}\left[4\{cos(2\pi\lambda_1)-u_{j1}\}^2\right]^{-d_{j1}} \left[4\{cos(2\pi\lambda_2)-u_{j2}\}^2\right]^{-d_{j2}} d\lambda_1 d\lambda_2. \qquad \text{(A5.25)}
$$

Note that the singularities in the integrand of equation (A5.25) occur at $f_{ji} = (cos^{-1}u_{ji})/2\pi$, for $j = 1, \ldots, k-1$, $i = 1, 2$.

Case 1: $k = 1$. The stationary results for $k = 1$ are known, but we include the proof here since the methodology of the proof can be easily generalized to the case $k > 1$. Define

$$
Q(\lambda_1,\lambda_2) = 2^2\{\cos(2\pi\lambda_1)-u_1\}\{\cos(2\pi\lambda_2)-u_2\},
$$

i.e.,

$$P(\lambda_1, \lambda_2) = cQ_1^{-2d_1}(\lambda_1)Q_2^{-2d_2}(\lambda_2) = Q^{-2(d_1,d_2)}(\lambda_1, \lambda_2),$$

where c is a positive finite constant. We examine the singularity at $f_i = (\cos^{-1} u_i)/2\pi$, $i = 1, 2$, by examining the limit $lim_{\varepsilon\to 0}Q(f_1 + \varepsilon, f_2 + \varepsilon)$. Now,

$$\begin{aligned} Q(f_1 + \varepsilon, f_2 + \varepsilon) &= 2^2 \left[\cos\left\{2\pi(f_1 + \varepsilon)\right\} - u_1\right] \left[\cos\left\{2\pi(f_2 + \varepsilon)\right\} - u_2\right] \\ &= 2^2 \prod_{i=1}^{2} \left[u_i \left\{\cos(2\pi\varepsilon) - 1\right\} - \sin(2\pi f_i)\sin(2\pi\varepsilon)\right] \end{aligned}$$

- For $|u_i| < 1$, $i = 1, 2$, using the l'Hopital rule, it can be computed the following limit:

$$lim_{\varepsilon\to 0}\frac{Q_i(f_i + \varepsilon)}{\varepsilon} \longrightarrow C_i, \quad i = 1, 2,$$

  where C_i, $i = 1, 2$, are finite nonzero constants. Thus, it follows that

$$lim_{\varepsilon\to 0}\left|\frac{P(f_1 + \varepsilon, f_2 + \varepsilon)}{\varepsilon^{-2d_1-2d_2}}\right| \longrightarrow C,$$

  where $0 < C < \infty$, i.e., $|P(\lambda_1, \lambda_2)| \sim |\lambda_1 - f_1|^{-2d_1} |\lambda_2 - f_2|^{-2d_2}$, as $|\lambda_i - f_i| \to 0$, $i = 1, 2$. Therefore, $P(\lambda_1, \lambda_2)$ is well behaved at its singularity, i.e., as $|\lambda_i - f_i| \to 0$, $i = 1, 2$, it behaves like $\prod_{i=1}^{2} |\lambda_i - f_i|^{-2d_i}$, which has a finite integral over the interval $[0, 0.5] \times [0, 0.5]$, whenever $0 < d_i < 0.5$, $i = 1, 2$.

- Considering $|u_i| = 1$, $i = 1, 2$, it follows that

$$lim_{\varepsilon\to 0}\left|\frac{P(f_1 + \varepsilon, f_2 + \varepsilon)}{\varepsilon^{-4d_1-4d_2}}\right| \longrightarrow K,$$

  where $0 < K < \infty$, i.e., $|P(\lambda_1, \lambda_2)| \sim \prod_{i=1}^{2} |\lambda_i - f_i|^{-4d_i}$, as $|\lambda_i - f_i| \to 0$, $i = 1, 2$. Thus, the integral in (A5.25) is finite as long as $0 < d_i < 0.25$.

Case 2: For $k > 1$, consider the singularity at f_{ji}, $1 \le j \le k-1$, $i = 1, 2$. Define, for $i = 1, 2$, and $1 \le j \le k - 1$, $Q_i(f_{ji}) = 2\{\cos(2\pi f_{ji}) - u_{ji}\}$. Now $lim_{\varepsilon\to 0}Q_i(f_{ji} + \varepsilon) = 2\{\cos(2\pi f_{ji} + \varepsilon) -$

$u_{ji}\} \neq 0$, and

$$lim_{\varepsilon \to 0} \prod_{j=1}^{k-1} \left\{ \prod_{i=1}^{2} Q_i^{-2d_{ji}} (f_{ji} + \varepsilon) \right\} < \infty.$$

Thus, it follows that

$$|P(\lambda_1, \lambda_2)| \sim \prod_{j=1}^{k-1} \prod_{i=1}^{2} |\lambda_i - f_{ji}|^{-2d_{ji}}, \quad \text{or } |P(\lambda_1, \lambda_2)| \sim \prod_{j=1}^{k-1} \prod_{i=1}^{2} |\lambda_i - f_{ji}|^{-4d_{ji}},$$

as $|\lambda_i - f_{ji}| \to 0$, depending upon whether $|u_{ji}| < 1$, or $|u_{ji}| = 1$, for $i = 1, 2$, and $1 \leq j \leq k-1$, respectively. The integrand in (A5.25) is then well behaved at each one of its singularities, and the stationarity follows for Case 2, as in Case 1. □

Theorem 6. *The $k-1$ factor Gegenbauer process is long memory if the conditions of Theorem 5 are satisfied in terms of the ranges of d_{ji} considered, for $i = 1, 2$, $j = 1, \ldots, k-1$.*

Proof The result follows directly from the definition of long memory and from the spectral density given in the equation (A5.24). □

Similar to the case of fractional difference processes given by the Gegenbauer processes, the $k-1$ factor Gegenbauer processes in (A5.23) can also include autoregressive factors. The following $k-1$ factor Gegenbauer autoregressive process is then introduced:

$$\phi(B_1, B_2) \prod_{j=1}^{k-1} (1 - 2u_{j1}B_1 + B_1^2)^{d_{j1}} \prod_{j=1}^{k-1} (1 - 2u_{j2}B_2 + B_2^2)^{d_{j2}} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}. \tag{A5.26}$$

Theorem 7. *Let $\{Y_{t_1,t_2}\}$ be a $k-1$ factor Gegenbauer autoregressive process. Then, the following assertions hold:*

- *Y_{t_1,t_2} is stationary if all the roots of the equation of $\phi(z_1, z_2) = 0$ lie outside the unit circle, and d_{ji} and u_{ji}, $j = 1, 2, \ldots, k-1$, $i = 1, 2$, satisfy the conditions in Theorem 5.*
- *Y_{t_1,t_2} is stationary and long memory if all the roots of the equation $\phi(z_1, z_1) = 0$ lie outside the unit circle, and d_{ji} and u_{ji}, $i = 1, 2$, $j = 1, 2, \ldots, k-1$, satisfy the conditions in Theorem 6.*

Proof

The proof is similar to the one for Hosking (1981) (Theorem 2). Equation (A5.22) can be rewritten as $Y_{t_1,t_2} = \Psi(B_1, B_2)\varepsilon_{t_1,t_2}$, where

$$\Psi(z_1, z_2) = \phi^{-1}(z_1, z_2) \prod_{j=1}^{k-1} (1 - 2u_{j1}z_1 + z_1^2)^{-d_{j1}} \prod_{j=1}^{k-1} (1 - 2u_{j2}z_2 + z_2^2)^{-d_{j2}} \theta(z_1, z_2).$$

Thus, $\Psi(z_1, z_2)$ is convergent for all $|z_i| \leq 1$, $i = 1, 2$, under the conditions assumed in this theorem, since the conditions in Theorem 5 guarantee the convergence of the power expansion of $\prod_{j=1}^{k-1} \prod_{i=1}^{2} (1 - 2u_{ji}z_i + z_i^2)^{d_{ji}}$, for all $|z_i| \leq 1$, $i = 1, 2$, and $\phi^{-1}(z_1, z_2)$ converges, for all $|z_i| \leq 1$, $i = 1, 2$, when all the roots of equation $\phi(z_1, z_2) = 0$ lie outside the unit circle.

The spectral density of the $k-1$ factor GARMA is given by

$$P(\lambda_1, \lambda_2) = \sigma_\varepsilon^2 \left| \phi(e^{i2\pi\lambda_1}, e^{i2\pi\lambda_2}) \right|^{-2} \prod_{j=1}^{k-1} \prod_{i=1}^{2} [4\{\cos(2\pi\lambda_i) - u_{ji}\}]^{-2d_{ji}}.$$

The proof of this part then follows directly from the above expression of the spectral density. □

Remark 18. *Cheng, (1993) studied the $k-1$ factor model in which $k = 3$. Also, Hassler (1994) considered the "flexible" seasonal model*

$$\left(\prod_{j=1}^{k-1} A_j^{d_j} \right) Y_t = a_t,$$

where the A_j, $j = 1, 2, \ldots, k-1$, are the irreducible factors of $1 - B^s$, and proves stationary results for this case. However, he does not prove these results for arbitrarily selected Gegenbauer factors. Giraitis and Leipus, (1995) have used an approach different from ours to prove the stationary and long-memory results in the setting considered here.

When $|u_{ji}| < 1$ and $0 < d_{ji} < 1/2$, $i = 1, 2$, $j = 1, 2, \ldots, k-1$, then, the autocorrelation function satisfies

$$\rho(k, l) \simeq \prod_{j=1}^{k-1} k^{2d_{j1}-1} \cos(2\pi k u_{j1}) \prod_{j=1}^{k-1} k^{2d_{j2}-1} \cos(2\pi l u_{j2}),$$

as $k \to \infty$ and $l \to \infty$, where, as before, u_{ji}, $i =, 1, 2$, and $j = 1, 2, \ldots, k-1$ are the Gegenbauer frequencies.

The autocovariance function $\{\gamma_{k,l}\}$ of the $k-1$ factor Gegenbauer process is defined as the inverse Fourier transformation of its spectrum as usual

$$\gamma(k,l) = 2\int_0^\pi \int_0^\pi f(\lambda_1, \lambda_2)\cos(2\pi\lambda_1 k)\cos(2\pi\lambda_2 l)d\lambda_1\lambda_2.$$

A5.7 Hilbert-valued spatial autoregressive Gegenbauer random fields

This section introduces the class of Hilbert-valued autoregressive Gegenbauer random fields. Specifically, we formulate the Hilbert-valued version of equation (A5.18). The moving average Hilbertian representation of infinite order is also obtained under suitable conditions on the eigenvalues, that define the spectral decomposition of the infinite-dimensional parameters (operators) involved in the autoregressive part of the formulated functional state equation, as well as on the spectra of the operators defining the coefficients of the functional version of the fractional operators. Let us consider $\mathcal{H}$ to be a separable Hilbert space, and let Y_{t_1,t_2} to be the solution to the following functional state equation:

$$\phi(B_1, B_2)(I - 2M_1B_1 + B_1^2)^{d_1}(I - 2M_2B_2 + B_2^2)^{d_2}(Y_{i,j} - R) = \varepsilon(i,j), \tag{A5.27}$$

where M_i, $i = 1, 2$, are bounded linear operators on $\mathcal{H}$, $R \in \mathcal{H}$ denotes the functional mean, and for all $(i,j) \in \mathbb{Z}^2$, $Y_{i,j}, \varepsilon(i,j) \in \mathcal{H}$, and B_1 and B_2 are the lag operators. Here, $\varepsilon(i,j)$ is the functional innovation process which is assumed to be strong-Hilbertian white noise, i.e.,

$$E[\varepsilon_{i,j} \otimes \varepsilon_{k,l}] = \delta_{i,k}\delta_{j,l}R_\varepsilon,$$

with the autocovariance operator $R_\varepsilon \in \mathcal{S}(\mathcal{H})$, the space of Hilbert-Schmidt operators on $\mathcal{H}$. Also,

$$\phi(B_1, B_2) = 1 - L_1B_1 - L_2B_2 - L_3B_1B_2,$$

where L_i, $i = 1, 2, 3$, satisfy de conditions given in Proposition 3 of Ruiz-Medina (2011). Specifically, they satisfy the following conditions:

C1. For $i = 1, 2, 3$, operator $L_i \in \mathcal{L}(\mathcal{H})$ is assumed to admit a spectral decomposition in terms of the eigenvalue sequence $\{\lambda_{ki},\ k \in \mathbb{N}\}$, and the biorthonormal systems of left eigenvectors $\{\psi_k,\ k \in \mathbb{N}\}$ and right eigenvectors $\{\phi_k,\ k \in \mathbb{N}\}$, defining dual Riesz bases of $\mathcal{H}$ and $\mathcal{H}^*$, and satisfying the following equations:

$$L_i(\psi_k) = \lambda_{ki}\psi_k, \qquad L_i^*(\phi_k) = \lambda_{ki}\phi_k, \quad k \in \mathbb{N}, \tag{A5.28}$$

where L_i^* denotes the adjoint of L_i, for $i = 1, 2, 3$. Under (A5.28), for $i = 1, 2, 3$, L_i admits the following spectral representation

$$L_i(g)(f) = \sum_{k\in\mathbb{N}} \lambda_{ki}\psi_k(f)\phi_k(g), \quad \forall f, g \in \mathcal{H}. \tag{A5.29}$$

C2. $|\lambda_{ki}| < 1$, for $i = 1, 2, 3$, $k \in \mathbb{N}$.

C3. $\left(1 + \lambda_{k1}^2 - \lambda_{k2}^2 - \lambda_{k3}^2\right)^2 - 4(\lambda_{k1} + \lambda_{k2}\lambda_{k3})^2 > 0$

C4. $1 - \lambda_{k2}^2 > |\lambda_{k1} + \lambda_{k2}\lambda_{k3}|$.

Similarly, we assume that operators M_i, $i = 1, 2$, satisfy the following equations:

$$M_i(\chi_k) = \xi_{ki}\chi_k, \qquad M_i^*(\varphi_k) = \xi_{ki}\varphi_k, \quad k \in \mathbb{N}, \tag{A5.30}$$

where M_i^* denotes the adjoint of M_i, for $i = 1, 2$, and $\{\xi_{ki},\ k \in \mathbb{N}\}$ are the eigenvalues of M_i, $i = 1, 2$. The eigenvector systems $\{\chi_k,\ k \in \mathbb{N}\}$ and $\{\xi_{ki},\ k \in \mathbb{N}\}$ are bio-orthogonal. Under (A5.30), for $i = 1, 2$, M_i then admits the corresponding spectral kernel representation

$$M_i(g)(f) = \sum_{k\in\mathbb{N}} \xi_{ki}\chi_k(f)\varphi_k(g), \quad \forall f, g \in \mathcal{H}. \tag{A5.31}$$

Theorem 8. *Under conditions **C1-C4**, the following hold.*

(i) If $\sup_k |\xi_{ki}| < 1$, *and* $d_i < 1/2$, $i = 1, 2$, *or if* $\sup_k |\xi_{ki}| = 1$, *and* $d_i < 1/4$, $i = 1, 2$, *the Hilbert-valued process* Y_{t_1,t_2} *satisfying equation (A5.27) is stationary.*

(ii) If $\sup_k |\xi_{ki}| < 1$, *and* $d_i > -1/2$, $i = 1, 2$, *or if* $\sup_k |\xi_{ki}| = 1$, *and* $d_i > -1/4$, $i = 1, 2$, *equation (A5.27) is invertible.*

The proof of Theorem 8 follows from Proposition 4 of Ruiz-Medina (2011), and from Theorem 4.

Corollary 1. *Under the conditions assumed in Theorem 8, and under the restrictions imposed in this theorem to the spectra of* M_i, $i = 1, 2$, *as well as considering the ranges established in such a theorem for* d_i, $i = 1, 2$, *equation (A5.27) admits a unique stationary solution given by*

$$\begin{aligned} Y_{i,j} - R &= (I - 2M_1B_1 + B_1^2)^{-d_1}(I - 2M_2B_2 + B_2^2)^{-d_2} \\ &\quad \times \left[\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} L_1^k L_2^l L_3^r(\varepsilon_{i-k-r,j-l-r})\right]. \\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} L_1^k L_2^l L_3^r (I - 2M_1B_1 + B_1^2)^{-d_1} \\ &\quad \times (I - 2M_2B_2 + B_2^2)^{-d_2}(\varepsilon_{i-k-r,j-l-r}) \end{aligned}$$

The proof follows straightforward from Proposition 4 of Ruiz-Medina (2011), keeping in mind equation (A5.21), and Theorem 3.

A5.8 Discussion

Further research developments are required in relation to the parameter estimation problem, as well as to the investigation of asymptotical properties, since these problems still remain open in the spatial long-range dependence Gegenbauer case. In particular, the minimum contrast estimation approach can be adopted to address the parameter estimation problem (see Anh, Leonenko and Sakhno, 2004). In the Hilbert-valued context, the derivation of the autocorrelation

operator, the functional spectrum, as well as the functional projection parameter estimators will be also addressed in a subsequent paper, extending the functional estimation results obtained in Ruiz-Medina, (2012a) for the spatial autoregressive Hilbertian case, as well as some asymptotic results formulated in Bosq, (2000), for the temporal Hilbert-valued time series context.

Appendix A6

Functional time series analysis of spatio-temporal epidemiological data

Ruiz-Medina, M. D., Espejo, R. M., Ugarte, M. D. and Militino, A.F. (2013).

Functional time series analysis of spatio-temporal epidemiological data.

Stochastic Environmental Research and Risk Assessment,

DOI: 10.1007/s00477-013-0794-y.

Abstract

Spatio-temporal statistical models have been proposed for the analysis of the temporal evolution of the geographical pattern of mortality (or incidence) risks in disease mapping. However, as far as we know, functional approaches based on Hilbert-valued processes have not been used so far in this area. In this paper, the autoregressive Hilbertian process framework is adopted to estimate the functional temporal evolution of mortality relative risk maps. Specifically, the penalized functional estimation of log-relative risk maps is considered to smooth the classical standardized mortality ratio (SMR). The Reproducing Kernel Hilbert Space (RKHS) norm is selected for definition of the penalty term. This RKHS-based approach is combined with the Kalman filtering algorithm for the spatio-temporal estimation of risk. Functional confidence intervals are also derived for detecting high risk areas. The proposed methodology is illustrated analyzing breast cancer mortality data in the provinces of Spain during the period 1975-2005. A simulation study is performed to compare the ARH(1) based estimation with the

classical spatio-temporal Conditional Autoregressive approach.

A6.1 Introduction

Functional time series theory has become an active research area in the last two decades. A complete treatment of the autoregressive Hilbertian process estimation, based on the method of moments, considering projection into the eigenvector system of the autocovariance operator can be found for example in Antoniadis, Paparoditis and Sapatinas, (2009), Bosq, (1996, 2000), Bosq and Blanke, (2007), Damon and Guillas, (2005), and Soltani and Hashemi, (2011). Maximum-likelihood estimation in the Gaussian context has been addressed in Salmerón and Ruiz-Medina, (2009) and Ruiz-Medina and Salmerón, (2010). Wavelet-based projection for continuous time prediction has been considered by Antoniadis and Sapatinas, (2003), and Laukaitis, (2007), in the ARH(1) framework. Some contributions in the spatial functional time series context have been recently derived in Ruiz-Medina, (2011, 2012a), and Ruiz-Medina and Espejo (2012).

Mortality (or incidence) risks are usually represented in maps summarizing the spatial or spatio-temporal distribution of mortality patterns over a region. The statistical methodology used in the epidemiological literature for estimating mortality risks is mainly based on extensions of the well-known autoregressive model of Besag, York and Mollié (BYM) Besag, York and Molié, (1991). Alternative Bayesian spatio-temporal models extending the BYM model are evaluated by Ugarte *et al.* (2009b). P-spline models are also being used for smoothing risks in space-time disease mapping (see for instance, Ugarte, Goicoa and Militino, 2010 and Ugarte *et al.*, 2012). A comparative study between spatial Conditional Autoregressive and P-spline modeling in disease mapping is presented in Goicoa *et al.* (2012). A flexible local-EM kernel smoothing algorithm is proposed in Nguyen, Brown and Stafford, (2012).

In a mathematical-probabilistic context, the spatio-temporal random field (S/TRF) modeling is proposed to integrate space and time in the construction of incidence maps (see Christakos,

1992 and Christakos, Olea and Yu, 2007). A quantitative description of incidence variation in relation to breast cancer analysis can be found in Christakos and Laf, (1997). The non-stationary nature of incidence variation can be suitably represented by considering the S/TRF approach. Prediction and extrapolation, or in general, estimation at any unsampled location and time can also be performed. Another potentially interesting feature of S/TRF-ν/μ is its connection with the theory of Stochastic Partial Differential Equations (SPDE), for modeling natural phenomena (see Christakos and Raghu, 1996, pp. 361-362). Recently, Yu and Christakos, (2011) considers the more flexible framework of generalized spatio-temporal random fields, which avoids the consideration of restrictive assumptions like linearity and normality. The functional approach adopted in this paper shares the above-referred features, in particular, the connection with the theory of SPDE (see Brown *et al.*, 2000).

In this paper, ARH(1) processes are considered for the construction of dynamical spatial epidemiological maps, capturing spatio-temporal interaction, and assuming a functional nature of the log-risk magnitude in space. In this context, penalized functional estimation is achieved following the approach given by Angelini, De Canditiis and Leblanc, (2003), based on the RKHS norm for definition of the penalty term, and extending it to the Hilbert-valued random process framework. The assumption that the autocorrelation operator is symmetric, and admits a spectral decomposition in terms of the auto-covariance eigenvector system is considered (see Bosq, 2000). Projection is then performed into orthogonal bases of the RKHS generated by the kernel of the auto-covariance operator of the ARH(1) process. The Kalman filtering algorithm is implemented for estimation of the projected random effect involved in the definition of a solution to the penalized functional estimation problem addressed (see Angelini, De Canditiis and Leblanc, 2003). In this implementation, moment-based estimates are considered for the autocorrelation and Hilbert-valued innovation auto-covariance operators (see Bosq, 2000). Heterogeneous spatial dependence modeling is achieved from the functional non-parametric framework adopted in

the representation of the log-relative-risk values in space. Its performance is compared with the classical Conditional Autoregressive (CAR) modeling.

As mentioned in Ugarte, Goicoa and Militino, (2010), the problem of detecting areas with high risks is also crucial to discover possible inequalities in health among regions and then, to implement appropriate health policies concerning allocation of health funds. To detect raised-risk areas in a spatio-temporal context, functional confidence intervals are constructed from the computation of the ARH(1) plug-in estimator and the posterior projection estimate of its covariance operator, obtained in the Kalman filtering recursion.

The rest of this paper is organized as follows. In Section 2, the Gaussian Hilbertian approximation to our observation model is derived. Section 3 describes the estimation methodology. In Section 4, the approach presented is illustrated using breast cancer mortality data from Spain. In Section 5, a simulation study is conducted to compare the CAR and ARH(1) frameworks in terms of bias and error of the relative risk estimates. Section 6 provides final comments and some open research lines.

A6.2 The model

The following assumptions are considered (see, for example, Ugarte, Goicoa and Militino, 2009a). The region under study is divided into l contiguous small areas, labelled by $i = 1, \ldots, l$. Conditional on the relative risk values $r_{i,t}$ at each area i and time t, $t = 1, \ldots, T$, the variables providing the observed number of cases $C_{i,t}$ at such area i and time t are assumed to be independent Poisson random variables with means $\mu_{i,t} = e_{i,t} r_{i,t}$. $i = 1, \ldots, l$, and $t = 1, \ldots, T$. Cases represent deaths or disease incidences, and $e_{i,t}$ is the expected number of cases in area i at time t. Hence,

$$\log(\mu_{i,t}) = \log(e_{i,t}) + \log(r_{i,t}), \quad i = 1, \ldots, l, \quad t = 1, \ldots, T.$$

Furthermore, in the subsequent development, we assume that

$$\log(r_{i,t}) = \beta(i) + u_{i,t} = \beta(i) + u_t(i)$$

with $\beta(\cdot)$ representing a spatial functional fixed effect, and $\{u_t(\cdot),\ t \in \mathbb{Z}\}$ being a zero-mean Gaussian ARH(1) process satisfying the state equation (A6.9) introduced in the Appendix. The observation model considered is then derived as follows

$$\begin{aligned} Z_{i,t} &= \log\left(\frac{C_{i,t}}{e_{i,t}}\right) = \log\left(C_{i,t}\right) - \log(e_{i,t}) = \log(\mu_{i,t}) + \frac{C_{i,t} - \mu_{i,t}}{\mu_{i,t}} - \log(e_{i,t}) \\ &= \beta(i) + u_t(i) + \frac{C_{i,t} - \mu_{i,t}}{\mu_{i,t}} \sim \beta(i) + u_t(i) + \varepsilon_{i,t}, \quad i = 1, \ldots, l, \quad t = 1, \ldots T, \end{aligned} \tag{A6.1}$$

where we have used the normal approximation to the Poisson distribution. Specifically, $\{\varepsilon_{i,t},\ i = 1, \ldots, l,\ t = 1, \ldots, T\}$ is a spatio-temporal Gaussian white noise, with $\varepsilon_{i,t}$ having variance $1/\mu_{i,t}$, for $i = 1, \ldots, l$, and $t = 1, \ldots, T$. Also, the needed conditions for interpretation of the underlying spatio-temporal white noise $\{\varepsilon_{i,t} = \varepsilon_t(i) = \varepsilon(\mathbf{x}_i, t), \quad \mathbf{x}_i \in D \subseteq \mathbb{R}^2,\ t \in \mathbb{Z}\}$ as a Hilbert-valued white noise in the strong sense in time $\{\varepsilon_t(\cdot) \in H,\ t \in \mathbb{R}_+\}$ are also assumed (see, for example, Bosq, 2000). In the subsequent development we work under the usual assumption that u_t and ε_t are independent, for $t = 1, \ldots, T$. Our observation model (A6.1) can be rewritten in a Hilbert-valued framework as

$$Z_t(i) = \beta(i) + u_t(i) + \varepsilon_t(i), \quad i = 1, \ldots, l, \quad t = 1, \ldots T, \tag{A6.2}$$

where $\beta, Z_t, u_t, \varepsilon_t \in H$, for $t = 1, \ldots T$, $H = L^2(D)$, $D \subset \mathbb{R}^2$. The index i refers to specific spatial coordinates $\mathbf{x}_i = (x_{i1}, x_{i2}) \in D$ associated with the i region studied, for $i = 1, \ldots, l$. Note that we are assuming that the support of the spatial functions in H includes the locations $\mathbf{x}_i = (x_{i1}, x_{i2})$, $i = 1, \ldots, l$, whose coordinates define the centroid associated with each one of the small areas studied.

Remark 19. *Under the conditions assumed here for the interpretation of a spatio-temporal process as a temporal Hilbert-valued process (see, for example, Bosq, 2000), the approach presented in Wikle and Cressie, (1999), can be interpreted within the ARH(1) framework, considering the special case where operator $\mathcal{A}$ introduced in (A6.9) (see Appendix) is an integral operator with kernel given by the interaction function defined in Wikle and Cressie, (1999), and denoted as w.*

A6.3 Penalized ARH(1)-based estimation

Let us consider the functional observation model (A6.2). Following the approach presented in Angelini, De Canditiis and Leblanc, (2003), the penalized least-squares estimation problem can be solved in a non-parametric mixed-effect framework, by projection into an orthonormal eigenvector system generating the Reproducing Kernel Hilbert Space (RKHS) of the Gaussian random effect. Thus, the norm of the RKHS of the random effect is selected for the definition of the penalty term. In our ARH(1) context, the following extended functional formulation of the minimization problem associated with the approach adopted in Angelini, De Canditiis and Leblanc, (2003) is considered:

$$\min_{F(\cdot)\in\mathcal{H}(Z)} \frac{1}{T}\sum_{t=1}^{T} \|\beta(\cdot)+u_t(\cdot)-F(Z_1,\ldots,Z_t)\|^2_{\mathcal{H}_0(Z)} + \gamma\|\mathbf{\Phi}_{\mathcal{H}_1(Z)}\left(F(Z_1,\ldots,Z_t)\right)\|^2_{\mathcal{H}_1(Z)}, \quad \text{(A6.3)}$$

where $\mathcal{H}(Z)$ denotes the RKHS generated by the kernel $\mathcal{K}_Z$ of the auto-covariance operator $R_0^Z = E[(Z_t-\beta)\otimes(Z_t-\beta)] = E[(Z_1-\beta)\otimes(Z_1-\beta)]$, for $t=1,\ldots,T$ (see Appendix). It is well-known (see, for example, Bosq, 2000), that $\mathcal{H}(Z)$ can be isometrically identified with the closed subspace $H(Z)$ of $\mathcal{L}^2_H(\Omega,\mathcal{A},P)$ generated by the zero-mean Gaussian Hilbert-valued random variables $Z_t-\beta$, $t=1,\ldots,T$, and their limits in the mean-square sense, where $(\Omega,\mathcal{A},P)$ denotes, as usual, the basic probability space (see Appendix). Thus, $H(Z)\equiv\mathcal{H}(Z)$. The inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}(Z)}$ in $\mathcal{H}(Z)$ is defined from the closed bilinear form generated by operator $[R_0^Z]^{-1}$,

the inverse operator of R_0^Z, that is,

$$\langle f, g \rangle_{\mathcal{H}(Z)} = [R_0^Z]^{-1}(f)(g), \quad \forall f, g \in \mathcal{H}(Z). \tag{A6.4}$$

In equation (A6.3), F is an operator from the Hilbert space $[\mathcal{H}(Z)]^{\otimes T} \equiv [H(Z)]^{\otimes T}$ into $\mathcal{H}(Z)$, where $[H]^{\otimes T}$ denotes the Tth self-tensorial product of the Hilbert space H. In addition, we have considered the following decomposition

$$\mathcal{K}_Z = \mathcal{K}_Z^0 + \mathcal{K}_Z^1$$

generating the direct sum

$$\mathcal{H}(Z) = \mathcal{H}_0(Z) \oplus \mathcal{H}_1(Z),$$

where functions in $\mathcal{H}_0(Z)$ are considered for the approximation of β, and the functions in $\mathcal{H}_1(Z)$ for the approximation of the small-scale local variation of Z, i.e., the roughness displayed by the functional values of Z (see Angelini, De Canditiis and Leblanc, 2003). The parameter γ is the smoothing parameter taking values in the interval $[0, 1]$. Note that, when $\gamma = 0$, the solution to the minimization problem (A6.3) will approximate the functional values of $\beta + u_t$, $t = 1, \ldots, T$, minimizing bias, in terms of an element of $\mathcal{H}(Z)$ having a huge norm in $\mathcal{H}_1(Z)$ (i.e., the corresponding estimator will displays high local variability). When $\gamma = 1$, the solution to (A6.3) will provide a coarser approximation to the functional values $\beta + u_t$, $t = 1, \ldots, T$, but $\mathbf{\Phi}_{\mathcal{H}_1(Z)}F$ will have small norm in the space $\mathcal{H}_1(Z)$ (low small-scale local variability). The solution to the minimization problem (A6.3) is computed from projection into the auto-covariance eigenvector system $\{\phi_i,\ i \geq 1\}$, under the following assumption, also considered in Bosq, (2000).

A.1 $\mathcal{A}$ is symmetric, thus, $\mathbf{\Psi} = \mathbf{\Phi}$ and $\mathcal{A} = \mathbf{\Phi}\mathbf{\Lambda}(\mathcal{A})\mathbf{\Phi}^*$, with $R_0^u = E[u_t \otimes u_t] = \mathbf{\Phi}\mathbf{\Lambda}(R_0^u)\mathbf{\Phi}^*$, for every $t \in \mathbb{Z}$ (see Appendix). Here, $\mathbf{\Lambda}(\mathcal{A})$ and $\mathbf{\Lambda}(R_0^u)$ denote the diagonal operators respectively containing the eigenvalues of $\mathcal{A}$ and R_0^u.

Note that condition **A.1** allows the computation of the solution to (A6.3) in terms of the common eigenvector system $\{\phi_i,\ i \in \mathbb{N}\}$ of $\mathcal{A}$ and R_0^u (see Ruiz-Medina and Salmerón, 2010).

Specifically, a finite-dimensional approximation to the forward Kalman filtering equations is implemented in terms of the projection into a finite set of eigenvectors $\{\phi_i,\ i=1,\ldots,M\}$ as given in Ruiz-Medina and Salmerón, 2010 (see Appendix). In this finite-dimensional approximation, the M eigenvectors selected correspond to the M highest eigenvalues. Hence, $\{\phi_i,\ i=1,\ldots,M\}$ generate the space $\mathcal{H}_1(Z)$, and $\{\phi_{M+1},\ldots\phi_{M+k},\ldots\}$ generate the space $\mathcal{H}_0(Z)$. The non-parametric kernel-based estimation of β is then obtained from the estimation of its coefficients with respect to $\{\phi_{M+1},\ldots\phi_{M+L}\}$, as given in Angelini, De Canditiis and Leblanc, (2003), for certain L defining the coarsest level. In the real-data example analyzed in the next section, a local polynomial kernel wavelet-based estimator $\widehat{\mathcal{K}}_Z^0$ (for high-dimensional data) is computed for approximation of $\mathcal{K}_Z^0$ (see, for example Rincón and Ruiz-Medina, 2012a, Appendix A.1).

Following the approach presented in Angelini, De Canditiis and Leblanc, (2003), γ can be chosen such that $T\gamma=\sigma^2/b$, with $\sigma^2=E\|\varepsilon_t\|_H^2$ and $b=E\|u_t\|_H^2$, for all $t\in\mathbb{Z}$, since ε and u are H-valued stationary processes. In the next section, we compute $\gamma=\sigma^2/Tb$ by cross validation.

For each time $t=1,\ldots,T$, equation (A6.16) below (see Appendix) provides a finite-dimensional approximation $\mathbf{\Phi}_M\mathbf{P}_{t|t}\mathbf{\Phi}_M^*$ of the a-posteriori covariance operator of the computed functional estimator, whose trace provides its a-posteriori functional variance. Indeed, for each time $t=1,\ldots,T$, its diagonal elements are estimates of the variance of the predictor at each spatial location analyzed, for such a time t. In the Gaussian case considered, from equations (A6.15) and (A6.16) below (see Appendix), the following finite-dimensional functional confidence intervals can then be computed for the ARH(1) process u :

$$\left[\mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t)+Z_{\alpha/2}\mathrm{diag}\left(\left[\mathbf{\Phi}_M\mathbf{P}_{t|t}\mathbf{\Phi}_M^*\right]^{1/2}\right),\mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t)+Z_{1-\alpha/2}\mathrm{diag}\left(\left[\mathbf{\Phi}_M\mathbf{P}_{t|t}\mathbf{\Phi}_M^*\right]^{1/2}\right)\right], \quad \text{(A6.5)}$$

for $t=1,\ldots,T$, where $Z_{\alpha/2}$ and $Z_{1-\alpha/2}$ denote the corresponding percentile values of the standard normal distribution, and $\mathrm{diag}(B^{1/2})$ denotes the diagonal matrix defined by the square root of the entries at the diagonal of the finite-dimensional operator B.

An approximation $\widehat{\mathbf{P}}_{t|t}$ to $\mathbf{P}_{t|t}$ can be computed from the projection estimates (A6.18) given in the Appendix. Considering the asymptotic properties of such projection estimates (A6.18) (see, for example, Bosq, 2009), the following *plug-in* finite-dimensional approximation of the functional confidence intervals for H-valued process u is computed

$$\left[\mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t) + Z_{\alpha/2}\text{diag}\left([\mathbf{\Phi}_M\widehat{\mathbf{P}}_{t|t}\mathbf{\Phi}_M^*]^{1/2}\right), \mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t) + Z_{1-\alpha/2}\text{diag}\left([\mathbf{\Phi}_M\widehat{\mathbf{P}}_{t|t}\mathbf{\Phi}_M^*]^{1/2}\right)\right], \quad \text{(A6.6)}$$

for each $t = 1, \ldots, T$. For the estimation of H-valued process $Y_t(\cdot) = \beta(\cdot) + u_t(\cdot)$, the following intervals are then considered:

$$\left[\mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t) + \widehat{\beta} + Z_{\alpha/2}\widehat{\mathbf{SE}}(\mathbf{\Phi}_M, \widehat{\mathbf{P}}_{t|t}), \mathbf{\Phi}_M\widehat{\mathbf{a}}(t|t) + \widehat{\beta} + Z_{1-\alpha/2}\widehat{\mathbf{SE}}(\mathbf{\Phi}_M, \widehat{\mathbf{P}}_{t|t})\right], \quad \text{(A6.7)}$$

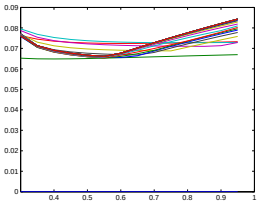
where $\widehat{\mathbf{SE}}(\mathbf{\Phi}_M, \widehat{\mathbf{P}}_{t|t}) = \text{diag}\left([\mathbf{\Phi}_M\widehat{\mathbf{P}}_{t|t}\mathbf{\Phi}_M^*]^{1/2}\right)$, for $t = 1, \ldots, T$.

Remark 20. *Note that in the above-described estimation procedure, kernel $\mathcal{K}_Z^0$ allows the reference to specific spatial locations in the estimation of the functional parameter β. Furthermore, in the Kalman filtering recursion (see Appendix), the composition with the projection operator $\mathbf{\Phi}_M$ by the left-hand side also allows to locate in space the functional estimation of the random effect u_t. Specifically, in the Kalman filtering recursion, the diagonal auto-covariance operator of the Gaussian observation noise $\mathbf{R}_{\varepsilon_t}$, appearing in the definition of the innovation or residual covariance operator $\left(T\gamma\mathbf{R}_{\varepsilon_t} + \mathbf{\Phi}_M\mathbf{P}_{t|t-1}\mathbf{\Phi}_M^*\right)$, involves the reference to the mean number of cases at each region, for every time t (year) analyzed. Namely, for every function $f \in H$, with $f(\mathbf{z}) = \sum_{k=1}^{\infty} f_k\phi_k(\mathbf{z})$, where $f_k = \langle f, \phi_k\rangle_H$, $k \geq 1$, with respect to the orthogonal eigenvector system $\{\phi_k,\ k \geq 1\}$ (see Appendix), operator $\mathbf{R}_{\varepsilon_t}$ can be approximated, for a given truncation level M, as follows*

$$\mathbf{R}_{\varepsilon_t}(f)(\mathbf{z}) = \sum_{k=1}^{M}\alpha_k\phi_k(\mathbf{z}), \quad \alpha_k = \sum_{i=1}^{l}\frac{1}{\mu_{i,t}}f_k\phi_k(\mathbf{x}_i),\ k = 1, \ldots, M,$$

where, as before, $\mathbf{x}_i = (x_{i1}, x_{i2})$ denotes the spatial coordinates of the centroid associated with the i-region, for $i = 1, \ldots, l,$, $\mu_{i,t} = e_{i,t}r_{i,t}$, $i = 1, \ldots, l$, and $t = 1, \ldots, T$.

Figure A6.1: Averaged 5- fold CV functional standard error estimates for $\gamma \in [0.3, 1)$, with discretization step size 0.05.



A6.4 Illustration

An application of the proposed penalized ARH(1)-based estimation methodology to the analysis of breast cancer mortality data is now presented. Specifically, breast cancer mortality data are collected from 50 Spanish provinces during the period 1975-2005. The observed relative risk over each region is computed as the observed number of cases divided by the number of expected cases. Expected cases are calculated standardizing by age and using Spain as the reference population. From the 31 breast cancer mortality log-relative-risk maps observed (one per year), the first one is considered as a realization of the functional random initial condition, and the remaining 30 maps constitute the original functional sample. The 50 provinces are randomly partitioned into 5 subsamples of the same size. As usual, at each iteration of the 5-fold cross validation (5- fold CV) procedure applied, a single subsample is retained for validation, i.e., for implementation of equations (A6.15)-(A6.17) through the 30 years analyzed. The remaining subsamples are considered for fitting the ARH(1) model, i.e., for computing the moment-based estimates (A6.18), from the sample constituted by the 30 maps of the training provinces. In the validation, at each one of the five rounds of the 5- fold CV, an estimation

$$\left[\text{diag}\left([\widehat{\mathbf{\Phi}}_M \widehat{\mathbf{P}}_{1|1} \widehat{\mathbf{\Phi}}_M^*]^{1/2}\right), \ldots, \text{diag}\left([\widehat{\mathbf{\Phi}}_M \widehat{\mathbf{P}}_{31|31} \widehat{\mathbf{\Phi}}_M^*]^{1/2}\right)\right] \tag{A6.8}$$

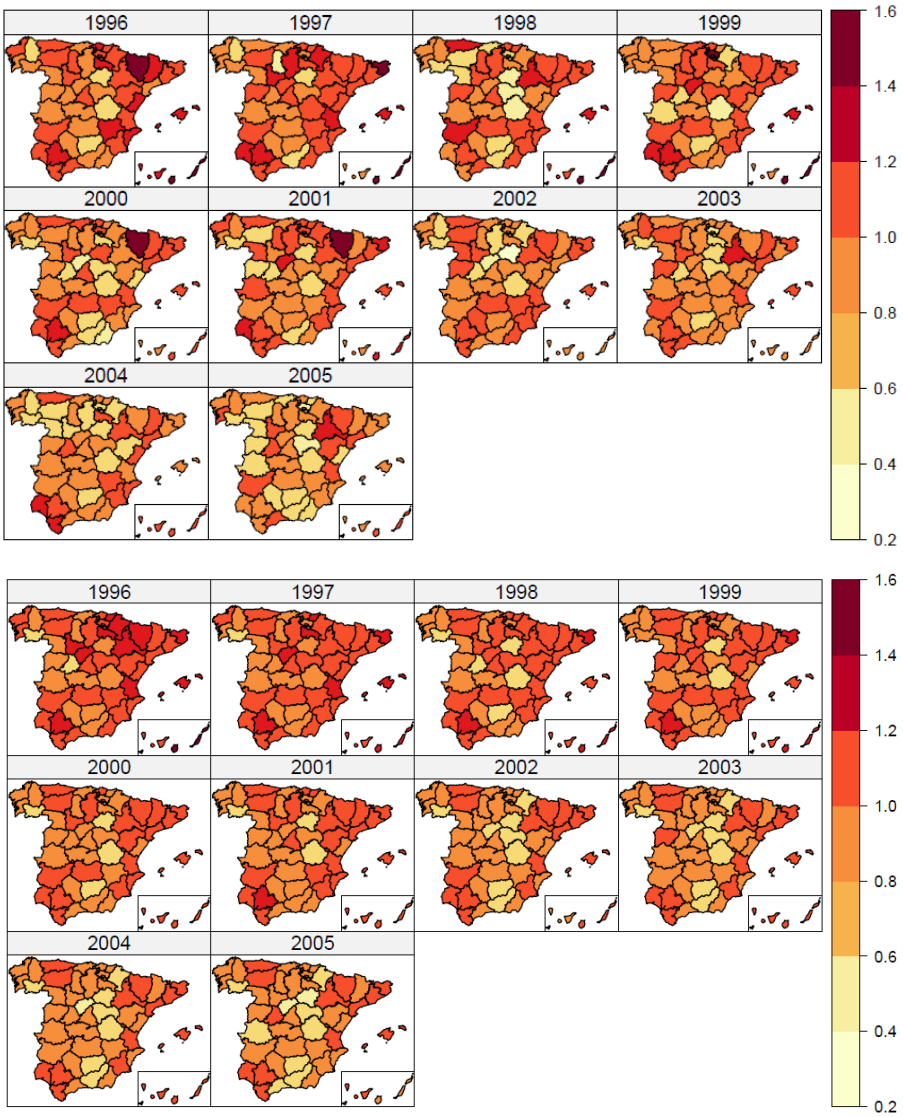
of the functional vector $\left[\text{diag}\left([\mathbf{\Phi}_M\mathbf{P}_{1|1}\mathbf{\Phi}_M^*]^{1/2}\right),\ldots,\text{diag}\left([\mathbf{\Phi}_M\mathbf{P}_{31|31}\mathbf{\Phi}_M^*]^{1/2}\right)\right]$ is computed, by applying the Kalman filtering recursion from the validation sample, considering the associated empirical eigenvector system $\{\widehat{\phi}_1,\ldots,\widehat{\phi}_M\}$ with $M = 29$. These estimates are averaged over the five rounds. The parameter γ selected corresponds to the value $\gamma = 0.55$, which provides the minimum L^∞ norm of the averaged 5-fold CV functional standard error vector estimate, computed from (A6.8) at each round.

Figure A6.1 shows the exponential of the averaged components over the five rounds of (A6.8) represented with respect to the values of γ tested, which correspond to a partition of the interval $(0,1)$ with discretization step size 0.05. For $\gamma < 0.3$, numerical inversion problems appear in the computation of the gain operators $\mathbf{K}_t$, $t = 1,\ldots,T$, when the truncation order $M = 29$ is considered. This is the reason why Figure A6.1 displays the results obtained for $\gamma \geq 0.3$.

The resulting penalized functional projection estimation of the breast cancer mortality relative risk maps is first analyzed, regarding the approximation it provides in relation to the geographical distribution of mortality risks. For illustration purposes, Figure A6.2 only displays the breast cancer mortality relative risk maps for the period 1996-2005. This period shows a decreasing mortality risk trend. It can be appreciated that the functional non-parametric fitting of the spatial dependence structure displayed by log-relative risks, in term of the ARH(1) framework, is more flexible allowing a higher order of local heterogeneity in the evolution of maps. During the last recuperation period analyzed, it seems that the penalized ARH(1) framework detects high risk areas more accurately than the CAR approach (see Figures A6.2-A6.3). Summarizing, the penalized ARH(1) framework seems to provide an interpolation between the singular spatial local variation of the observed log-relative risks, and the spatial local smoothing provided by the CAR model.

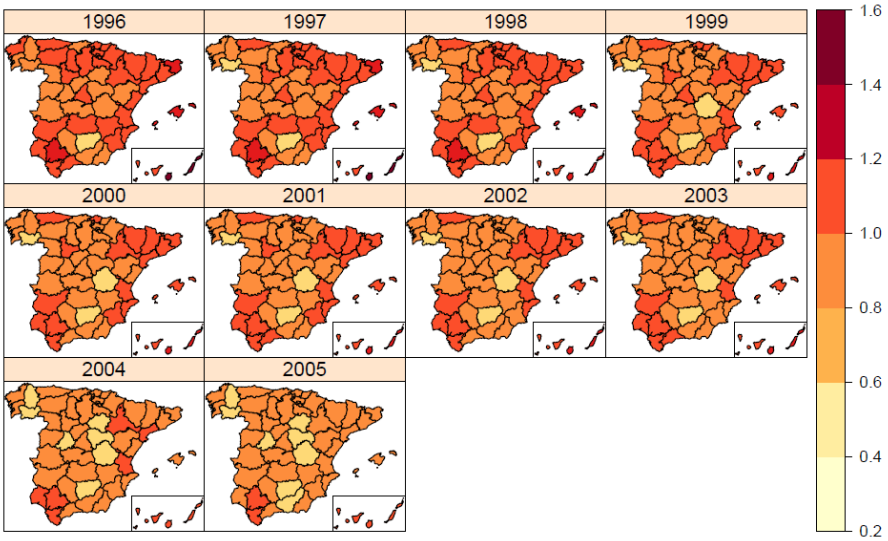
Confidence intervals for the relative risk values are constructed by applying the exponential function to the lower and upper confidence bounds in (A6.7). For a maximum amplitude of

Figure A6.2: Standardized mortality ratio maps (top figure) and functional estimated breast cancer mortality risk maps based on the ARH(1) approach (bottom figure)



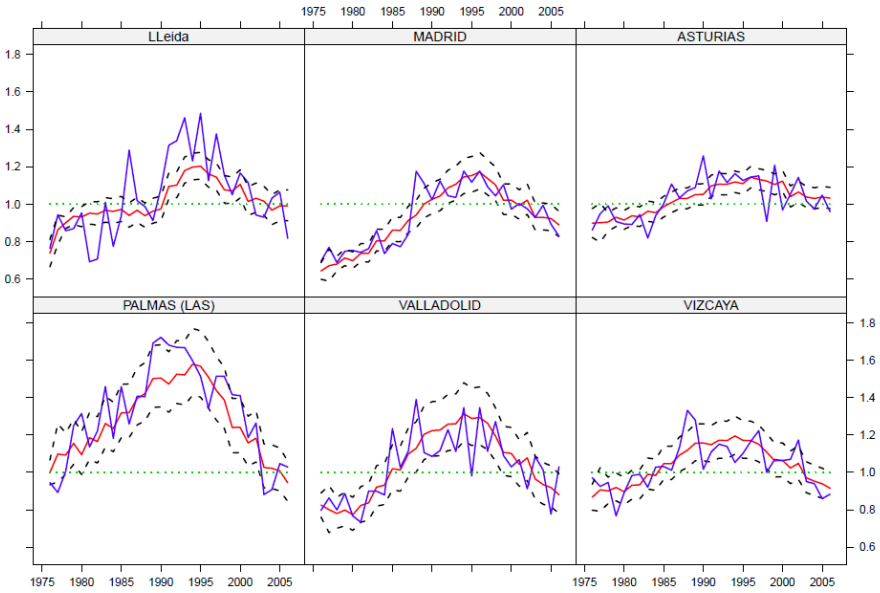
such intervals of order 0.2, the value $1 - \alpha = 0.8696$ is selected in their construction. As we said before, the estimation of spatial function β is based on a local polynomial kernel wavelet-based approximation of $\mathcal{K}_Z^0$, obtained from the wavelet transform of Epanechnikov kernel (see

Figure A6.3: Relative risk estimates using a spatio-temporal CAR model



Appendix A.1 of Rincón and Ruiz-Medina, 2012a, for its computation, as well as Müller, 1987, Wu and Zhang, 2006). To detect high risk areas, curves providing lower and upper confidence bounds at each province in the period analyzed are constructed, by considering the corresponding time-cross-sections in the sequence of lower and upper confidence bounds associated with the functional interval sequence (A6.7). A region is classified as a raised-risk area if the lower confidence bound of the interval is greater than 1 (see, for example, Ugarte, Goicoa and Militino, 2009a). Following this procedure, the estimated breast cancer relative risk curves, and the corresponding confidence intervals, for $1 - \alpha = 0.8696$, are displayed in Figures A6.4, for some Spanish provinces since 1976. It can be observed that the penalized ARH(1) estimates of relative risk curves at each Spanish province display a higher degree of local singularity than the ones obtained with the CAR model. Edge effects must be corrected in some provinces considering, for example, suitable wavelet bases. Again, the penalized ARH(1) temporal smoothing interpolates the local variability displayed by the observed relative risks and the CAR estimates.

Figure A6.4: SMR (blue line), ARH(1) estimated risks (red line) and confidence bounds for relative risks at some provinces in Spain over the period 1975-2005



A6.5 Simulation study

A simulation study has been conducted to compare the proposed penalized functional estimation methodology with a spatio-temporal CAR model (see Section 2.2 in Ugarte, Goicoa and Militino, 2010) in terms of bias and error of the relative risk estimates. CAR and ARH(1) models are considered as data scenarios in the 100 simulations performed for this comparison. CAR model is firstly generated in terms of the neighborhood structure, expected values, and parameter values obtained from the analysis of the Spanish breast cancer mortality data. Both models, the CAR and the ARH(1) model, are then fitted to the generated data. The integrated nested Laplace approximation technique was used to fit the CAR model, because it is computationally very fast and stable (see Rue, Martino and Chopin, 2009). Secondly, data are generated under an ARH(1) model with functional innovations displaying weak-dependence in time, and having

spatial integral auto-covariance operator constructed from a Gamma-type kernel (see Ruiz-Medina and Salmerón, 2010). The non-parametric estimate $\widehat{\beta}(\cdot)$ of $\beta(\cdot)$ derived in the analysis of the Spanish breast cancer mortality data is considered in the generation of the log-relative risk values. Risk estimates are then computed applying both, CAR and ARH(1), estimation methodologies.

Let r_{jit} be the jth-simulated risk of breast cancer death ($j = 1, \ldots, J$) by small area i ($i = 1, \ldots, I$) in year t ($t = 1, \ldots, T$), and $\hat{r}_{ijt}$ its corresponding estimate. We first define the Mean Relative Bias and the Relative Root Mean Prediction Error as follows

$$\text{MARB}_{it} = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{\hat{r}_{jit} - r_{jit}}{r_{jit}}\right), \quad \text{RRMPE}_{it} = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(\frac{\hat{r}_{jit} - r_{jit}}{r_{jit}}\right)^2}.$$

The mean relative absolute bias and mean relative root mean prediction error by provinces are computed averaging the previous indicators by years. Namely

$$\text{MARB}_i = \frac{1}{T}\sum_{t=1}^{T}|MARB_{it}|, \quad \text{MRRMPE}_i = \frac{1}{T}\sum_{t=1}^{T} RRMEP_{it}.$$

We finally compute the Mean Relative Absolute Bias (MARB) and the Mean Relative Root Mean Prediction Error (MRRMPE) as global measures

$$\text{MARB} = \frac{1}{I}\sum_{i=1}^{I} MARB_i, \quad \text{MRRMPE} = \frac{1}{I}\sum_{i=1}^{I}\text{MRRMPE}_i.$$

Results obtained from the simulation study are given in Tables A6.1 and A6.3. When the ARH(1) risk generation is considered, better results are obtained from the ARH(1)-based estimation than from the CAR-based estimation in most of the provinces, in terms of both relative bias and error. In addition, the ARH(1) approach slightly outperforms the CAR approach in terms of the Mean Relative Root Mean Prediction Square Error when simulating from the CAR model. Small differences in favor of the CAR model are observed when using the MARB as

a global measure of bias. Indeed, precision, measured in terms of the Mean Relative Absolute Bias, could be improved in the ARH(1) estimates obtained from the CAR-based simulations by replacing the auto-covariance eigenvectors generating the space $\mathcal{H}_0$ by suitable wavelet bases (see, for example, Angelini, De Canditiis and Leblanc, 2003). This improvement should be reflected in a better approximation of the spatial functional parameter $\beta(\cdot)$ in equation (A6.1). However, it should be noted that the selected auto-covariance eigenvector system is optimal in the approximation of the local variability of the process, that is, in the construction of the space $\mathcal{H}_1$, whose norm is related to the inverse of the operator associated with the kernel $\mathcal{K}_Z^1$. This fact allows us to control the mean-square variability of the estimator, and, in particular, the Mean Relative Root Mean Prediction Error through the smoothing parameter γ in equation (A6.3). Indeed, the model selection problem associated with this parameter could be solved adopting some information criteria (see, for example, Araki *et al.*, 2009; Kawano and Konishi, 2009; Rincón Ruiz-Medina, 2012b, among others). Alternatively, this parameter could be approximated from the estimation of the functional variance of the ARH(1) process u_t, and of the Gaussian innovation process ν_t involved in the ARH(1) state equation (A6.9) (see Angelini, De Canditiis and Leblanc, 2003). These topics will be addressed in a subsequent paper.

Table A6.1: Average of the Mean Relative Absolute Bias by province ($MARB_i$) and Mean Relative Absolute Bias (MARB) obtained with two sets of 100 simulations from the CAR and ARH(1) models respectively

Cod	Prov	CAR-based simulations		ARH(1)-based simulations	
		CAR est	ARH(1) est	CAR est	ARH(1) est
1	Alava	0.029	0.012	0.054	0.014
2	Albacete	0.045	0.009	0.042	0.009
3	Alicante	0.038	0.007	0.013	0.007
4	Almería	0.057	0.138	0.031	0.040
5	Ávila	0.034	0.261	0.065	0.071
6	Badajoz	0.045	0.007	0.025	0.012
7	Balears (Illes)	0.073	0.006	0.017	0.012
8	Barcelona	0.039	0.006	0.005	0.001
9	Burgos	0.032	0.024	0.035	0.042
10	Cáceres	0.056	0.009	0.031	0.011
11	Cádiz	0.050	0.009	0.019	0.004
12	Castellon	0.037	0.008	0.028	0.005
13	Ciudad Real	0.041	0.066	0.029	0.009
14	Córdoba	0.039	0.010	0.024	0.006
15	Coruña (A)	0.046	0.008	0.011	0.011
16	Cuenca	0.043	0.017	0.051	0.010
17	Girona	0.068	0.005	0.023	0.005
18	Granada	0.039	0.011	0.022	0.023
19	Guadalajara	0.026	0.030	0.081	0.007
20	Guipúzcoa	0.038	0.032	0.027	0.024
21	Huelva	0.049	0.018	0.031	0.008
22	Huesca	0.036	0.019	0.056	0.024
23	Jaén	0.044	0.007	0.025	0.005
24	León	0.032	0.012	0.027	0.008
25	Lleida	0.048	0.196	0.034	0.005
26	Rioja (La)	0.039	0.033	0.050	0.008
27	Lugo	0.046	0.675	0.029	0.006
28	Madrid	0.035	0.004	0.006	0.004
29	Málaga	0.046	0.017	0.020	0.003
30	Murcia	0.040	0.014	0.018	0.008

Table A6.2: (continue)

Cod	Prov	CAR-based simulations		ARH(1)-based simulations	
		CAR est	ARH(1) est	CAR est	ARH(1) est
31	Navarra	0.044	0.011	0.027	0.014
32	Ourense	0.032	0.012	0.030	0.006
33	Asturias	0.046	0.004	0.012	0.006
34	Palencia	0.044	0.022	0.062	0.007
35	Palmas (Las)	0.036	0.010	0.035	0.004
36	Pontevedra	0.041	0.282	0.017	0.018
37	Salamanca	0.037	0.030	0.032	0.008
38	SC Tenerife	0.042	0.007	0.030	0.014
39	Cantabria	0.044	0.009	0.026	0.020
40	Segovia	0.034	0.034	0.068	0.007
41	Sevilla	0.041	0.004	0.017	0.011
42	Soria	0.041	0.016	0.110	0.014
43	Tarragona	0.044	0.007	0.026	0.005
44	Teruel	0.034	0.024	0.075	0.007
45	Toledo	0.039	0.019	0.029	0.008
46	Valencia	0.041	0.006	0.010	0.004
47	Valladolid	0.036	0.030	0.034	0.007
48	Vizcaya	0.030	0.009	0.016	0.004
49	Zamora	0.043	0.027	0.050	0.015
50	Zaragoza	0.040	0.021	0.023	0.008
	MARB	0.042	0.045	0.033	0.012

Table A6.3: Average of the Mean Relative Root Mean Prediction Square Error by province ($MRRMPE_i$) and Mean Relative Root Mean Prediction Error (MRRMPE) obtained with two sets of 100 simulations from the CAR and ARH(1) models respectively.

Cod	Prov	CAR-based simulations		ARH(1)-based simulations	
		CAR est	ARH est	CAR est	ARH est
1	Alava	0.256	0.116	0.242	0.095
2	Albacete	0.260	0.102	0.192	0.055
3	Alicante	0.295	0.068	0.075	0.044
4	Almería	0.329	1.384	0.132	0.375
5	Ávila	0.255	0.261	0.299	0.668
6	Badajoz	0.279	0.075	0.135	0.091
7	Balears (Illes)	0.396	0.066	0.078	0.094
8	Barcelona	0.286	0.055	0.029	0.006
9	Burgos	0.231	0.257	0.184	0.388
10	Cáceres	0.317	0.101	0.146	0.081
11	Cádiz	0.282	0.091	0.097	0.025
12	Castellon	0.269	0.086	0.136	0.026
13	Ciudad Real	0.255	0.675	0.147	0.061
14	Córdoba	0.270	0.089	0.123	0.030
15	Coruña (A)	0.287	0.081	0.052	0.094
16	Cuenca	0.248	0.178	0.231	0.070
17	Girona	0.380	0.064	0.110	0.020
18	Granada	0.274	0.111	0.125	0.202
19	Guadalajara	0.230	0.324	0.311	0.039
20	Guipúzcoa	0.285	0.321	0.111	0.210
21	Huelva	0.262	0.195	0.148	0.042
22	Huesca	0.240	0.208	0.218	0.196
23	Jaén	0.265	0.078	0.123	0.031
24	León	0.238	0.126	0.138	0.043
25	Lleida	0.288	0.199	0.170	0.025
26	Rioja (La)	0.253	0.434	0.220	0.051
27	Lugo	0.274	0.676	0.153	0.035
28	Madrid	0.245	0.044	0.034	0.029
29	Málaga	0.264	0.173	0.103	0.017
30	Murcia	0.285	0.151	0.098	0.054

Table A6.4: (continue)

Cod	Prov	CAR-based simulations		ARH(1)-based simulations	
		CAR est	ARH est	CAR est	ARH est
31	Navarra	0.270	0.126	0.141	0.097
32	Ourense	0.263	0.147	0.147	0.030
33	Asturias	0.293	0.045	0.066	0.041
34	Palencia	0.290	0.230	0.272	0.035
35	Palmas (Las)	0.255	0.104	0.199	0.019
36	Pontevedra	0.277	0.282	0.088	0.159
37	Salamanca	0.249	0.294	0.149	0.049
38	SC Tenerife	0.284	0.074	0.171	0.107
39	Cantabria	0.269	0.095	0.127	0.154
40	Segovia	0.251	0.342	0.276	0.038
41	Sevilla	0.262	0.048	0.097	0.096
42	Soria	0.256	0.235	0.517	0.094
43	Tarragona	0.274	0.084	0.138	0.027
44	Teruel	0.249	0.230	0.302	0.032
45	Toledo	0.255	0.226	0.146	0.056
46	Valencia	0.270	0.065	0.058	0.030
47	Valladolid	0.243	0.325	0.164	0.044
48	Vizcaya	0.258	0.089	0.081	0.018
49	Zamora	0.277	0.257	0.213	0.111
50	Zaragoza	0.258	0.248	0.118	0.050
	MRRMPE	0.272	0.207	0.157	0.090

A6.6 Final Comments

The penalized ARH(1) estimation approach presented in this paper allows the global estimation of mortality risk maps in disease mapping, displaying weak-dependence in time. The results derived here show that the temporal directional nature of breast cancer data, and the spatial local variability of the associated mortality log-risk can be suitably fitted under the functional statistical approach presented here. It can be appreciated that the amount of smoothing provided by the penalized ARH(1) projection estimation method interpolates the SMR variation and the CAR smoothing. The empirical orthogonal eigenfunction bases are selected for projection. They reproduce the spatial heterogeneous local variability properties displayed by the functions in $\mathcal{H}_1(Z)$ (see, for example, Ruiz-Medina and Espejo, 2012). However, some edge effects are observed in some provinces that could be removed by the selection of alternative bases, like suitable compactly supported wavelet bases (see, for example, Ruiz-Medina and Espejo, 2013a). The methodology given in this paper could be also extended to the framework of missing data following the approach derived in Ruiz-Medina and Salmerón, (2010), based on the combination of EM algorithm with forward and backward Kalman filtering. But, for this extension, the moment-based parameter estimation considered here must be replaced by the maximum likelihood approach.

The simulation study conducted in this article supports the fact that the considered functional penalized ARH(1) estimation methodology permits a suitable control of the spatial local variability of the log-risk values allowing for spatial heterogeneities. Additionally, the autoregressive dynamics of order one assumed in time seems to fit the temporal evolution and correlation of mortality risk maps accurately, as it occurs in the classical conditional autoregressive scenario. However, we have to note that further research is needed with regard to the selection problem associated with the functional basis, the truncation order, and the smoothing parameter γ. Regarding the truncation order, a first criterion considered here has been the percentage of ex-

plained empirical variability computed in terms of the approximated trace of the auto-covariance operator of the ARH(1) process. However, the truncation criterion could change depending on the characteristics of the orthogonal basis chosen for projection (see, for example, Ruiz-Medina and Espejo, 2012, 2013a). In relation to the smoothing parameter, although it has been selected in the illustration by cross-validation, alternative approaches could be adopted based on estimation of the functional components of variance (see, for example, Ugarte, Goicoa and Militino, 2010) or on the application of information criteria (see Araki *et al.*, 2009; Kawano and Konishi, 2009; Rincón and Ruiz-Medina, 2012b, among others). Some improvements could be achieved by including an additional term for penalization of the spatial random variability, allowing its localization by provinces. All these aspects will be pursued in a subsequent paper.

Appendix. Autoregressive Hilbertian processes of order one

Let us consider H to be a separable Hilbert space. The random Hilbert-valued sequence $\{Y_t,\ t \in \mathbb{Z}\}$ defined on the basic probability space $(\Omega, \mathcal{A}, P)$ is said to be an ARH(1) process if it satisfies (see Bosq, 2000)

$$X_t = Y_t - \mu = \mathcal{A}\left(X_{t-1}\right) + \nu_t, \quad t \in \mathbb{Z}, \tag{A6.9}$$

where $\mu \in H$, and $Y_t, X_t, \nu_t \in H$. The zero-mean H-valued process ν can be a martingale difference sequence or a Hilbertian white noise in the strong sense. The autocovariance operator of the Hilbert-valued innovation process ν will be denoted as $R_\nu = E[\nu_t \otimes \nu_t]$, for $t \in \mathbb{Z}$. Here, $\otimes$ represents the tensorial product $f \otimes g$ of two functions f and g in H, which defines a Hilbert-Schmidt operator on this space. The autocorrelation operator $\mathcal{A}$ belongs to $\mathcal{L}(H)$, the space of bounded linear operators on H. In the following, we will refer to the *spatial case*, where H is constituted by functions having spatial support contained in a bounded set $D \subseteq R^n$. Note that, in the referred spatial case, the autocorrelation operator $\mathcal{A}$ reflects the spatio-temporal interaction.

Operator $\mathcal{A}$ can be a non-symmetric compact operator satisfying

$$\mathcal{A}\psi_i = \lambda_i(\mathcal{A})\psi_i, \quad i \in \mathbb{N},$$

and

$$\mathcal{A}^*\phi_i = \lambda_i(\mathcal{A}^*)\phi_i, \quad i \in \mathbb{N},$$

where $\lambda_i(\mathcal{A}^*) = \lambda_i(\mathcal{A})$, with $\mathcal{A}^*$ denoting the adjoint operator of $\mathcal{A}$, and $\{\psi_i,\ i \in \mathbb{N}\}$ and $\{\phi_i,\ i \in \mathbb{N}\}$ being the left and right eigenvector systems associated with $\mathcal{A}$ and $\mathcal{A}^*$, respectively (see Ruiz-Medina and Salmerón, 2010 and Salmerón and Ruiz-Medina, 2009). By projection into the right eigenvector system, we obtain a diagonal infinite-dimensional version of the autoregressive equation (A6.9), given by

$$a_j(t) = \lambda_j(\mathcal{A})a_j(t-1) + \nu_j(t), \quad j \in \mathbb{N}, \quad t \in \mathbb{Z}, \tag{A6.10}$$

where

$$a_j(t) = \langle X_t(\cdot), \phi_j(\cdot)\rangle_H, \quad j \in \mathbb{N}, \quad t \in \mathbb{Z}, \tag{A6.11}$$

and

$$\nu_j(t) = \langle \nu_t(\cdot), \phi_j(\cdot)\rangle_H, \quad j \in \mathbb{N}, \quad t \in \mathbb{Z}.$$

The second-order structure of the H-valued process $\{X_t = Y_t(\mathbf{x}) - \mu,\ t \in \mathbb{Z}\}$ is characterized in terms of the auto-covariance and cross-covariance operators

$$R_0 = R_{X_tX_t} = E\left[X_t \otimes X_t\right], \quad t \in \mathbb{Z}, \tag{A6.12}$$

$$R_1 = R_{X_{t+1}X_t} = E\left[X_{t+1} \otimes X_t\right], \quad t \in \mathbb{Z}, \tag{A6.13}$$

which, from the observation of the functional random variables $X_1, \ldots, X_T$, can be respectively approximated in terms of the following empirical covariance operators

$$\hat{R}_0 = \frac{1}{T}\sum_{t=1}^{T} X_t \otimes X_t$$
$$\hat{R}_1 = \frac{1}{T-1}\sum_{t=1}^{T-1} X_{t+1} \otimes X_t. \tag{A6.14}$$

As derived in Ruiz-Medina and Salmerón, (2010), from equation (A6.10), for a given truncation level M, the following finite-dimensional approximation to the Kalman filtering equations in an ARH(1) framework is obtained by applying the projection operator $\mathbf{\Phi}_M^*$ into the finite-dimensional eigenvector system $\{\phi_i,\ i = 1, \ldots, M\}$

$$\widehat{\mathbf{a}}(t|t) = \widehat{\mathbf{a}}(t|t-1) + \mathbf{K}_t \left(Z_t - \mathbf{\Phi}_M \widehat{\mathbf{a}}(t|t-1)\right), \quad t = 1, \ldots, T, \tag{A6.15}$$

where

$$\widehat{\mathbf{a}}(t|t) = E\left(\mathbf{a}(t)|Z_t, \ldots, Z_1\right)$$

is the updated (a posteriori) projection estimate of the random coefficients $a_1(t), \ldots, a_M(t)$ at time t as given in (A6.11) and

$$\widehat{\mathbf{a}}(t|t-1) = E\left(\mathbf{a}(t)|Z_{t-1}, \ldots, Z_1\right)$$

is the corresponding projection predictor (a priori) estimate, previously computed for time t. Here, $\mathbf{K}_t$ is the optimal Kalman gain operator

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{\Phi}_M^* \left(T\gamma \mathbf{R}_{\varepsilon_t} + \mathbf{\Phi}_M \mathbf{P}_{t|t-1}\mathbf{\Phi}_M^*\right)^{-1},$$

defined from the innovation (or residual) covariance operator $\left(T\gamma \mathbf{R}_{\varepsilon_t} + \mathbf{\Phi}_M \mathbf{P}_{t|t-1}\mathbf{\Phi}_M^*\right)$, in terms of the predicted (a priori) projection estimate of the covariance operator

$$\mathbf{P}_{t|t-1} = \mathrm{Cov}\left(\mathbf{a}(t)|Z_{t-1}, \ldots, Z_1\right) = \mathbf{\Lambda}(\mathcal{A})\mathbf{P}_{t-1|t-1}\mathbf{\Lambda}(\mathcal{A}) + \mathbf{Q}.$$

Here, $\mathbf{R}_{\varepsilon_t}$ denotes the auto-covariance operator of the observation noise ε_t. The projected auto-covariance operator $\mathbf{Q}$ of the functional innovations process ν is given by

$$\mathbf{Q} = \mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M.$$

The updated (a posteriori) projection estimate of the covariance operator is computed as follows

$$\begin{aligned} \mathbf{P}_{t|t} &= E\left((\mathbf{a}(t) - \hat{\mathbf{a}}(t|t))\,(\mathbf{a}(t) - \hat{\mathbf{a}}(t|t))^*\right) \\ &= \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{\Phi}_M \mathbf{P}_{t|t-1}. \end{aligned} \tag{A6.16}$$

Therefore, the one-step-ahead (a priori) projection predictor is obtained from

$$\widehat{\mathbf{a}}(t|t-1) = \mathbf{\Lambda}(\mathcal{A})\widehat{\mathbf{a}}(t-1|t-1). \tag{A6.17}$$

The initial values considered are

$$\widehat{\mathbf{a}}(0|0) = \mathbf{0}, \quad \mathbf{P}_{0|0} = \mathbf{E}\left(\mathbf{a}(0)\mathbf{a}(0)^*\right),$$

where $\mathbf{a}(0) = \mathbf{\Phi}_M^* \mathbf{Z}_0$.

In the implementation of the above Kalman filtering equations, the following moment-based parameter estimates $\widehat{Q}$ and $\widehat{\mathcal{A}}$ of operators $\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M$ and $\mathcal{A}$, respectively, have been considered (see Bosq, 2000)

$$\begin{aligned} \widehat{Q} &= \mathbf{\Phi}_M^* \widehat{R}_\nu \mathbf{\Phi}_M = \mathbf{\Phi}_M^* \widehat{R}_0 \mathbf{\Phi}_M - \mathbf{\Phi}_M^* \widehat{R}_1 \mathbf{\Phi}_M \widehat{\Lambda} - \widehat{\Lambda} \mathbf{\Phi}_M^* \widehat{R}_1 \mathbf{\Phi}_M + \widehat{\Lambda} \mathbf{\Phi}_M^* \widehat{R}_0 \mathbf{\Phi}_M \widehat{\Lambda} \\ \widehat{\mathcal{A}} &= \widehat{R}_1 \widehat{R}_0^{-1}, \quad \widehat{\mathbf{\Lambda}}(\mathcal{A}) = \mathbf{\Phi}_M^* \widehat{\mathcal{A}} \mathbf{\Phi}_M, \end{aligned} \tag{A6.18}$$

where $\widehat{R}_0$ and $\widehat{R}_1$ are defined as in equation (A6.14).

Appendix A7

Maximum-likelihood asymptotic inference for autoregressive Hilbertian process

Ruiz-Medina, M. D. and Espejo, R. M. (2013b).

Maximum-Likelihood Asymptotic Inference for Autoregressive Hilbertian Processes.

Methodology and Computing Applied Probability,

DOI: 10.1007/s11009-013-9329-8.

Abstract

The autoregressive Hilbertian process framework has been introduced in Bosq (2000). This book provides the nonparametric estimation of the autocorrelation and covariance operators of the autoregressive Hilbertian processes. The asymptotic properties of these estimators are also provided. The maximum likelihood approach still remains unexplored. This paper obtains the asymptotic distribution of the maximum likelihood (ML) estimators of the auto-covariance operator of the Hilbert-valued innovation process, and of the autocorrelation operator of a Gaussian autoregressive Hilbertian process of order one. A real data example is analyzed in the financial context for illustration of the performance of the projection maximum likelihood estimation methodology in the context of missing data.

A7.1 Introduction

Recent contributions in the context of functional linear models are derived, for example, in Basse, Diop and Dabo-Niang (2008); Bosq (2000); Bosq and Blanke (2007); Ferraty and Vieu (2006); Guillas and Lai (2010); Ramsay and Silverman (2005); Ruiz-Medina (2011, 2012a), among others.

In the functional time series framework, the papers by Ruiz-Medina and Salmerón (2010, 2011) and Salmerón and Ruiz-Medina (2009) provide the maximum likelihood estimation of autoregressive Hilbertian processes of order one (ARH(1)processes) from incomplete functional data, by implementation of the forward and backward Kalman filtering methodology in combination with the Expectation-Maximization (EM) algorithm in terms of projections. The numerical projection methodology applied is based on the spectral decomposition of the autocorrelation operator, under the assumption of its compactness, in terms of the left and right eigenvector systems. In addition, Ruiz-Medina and Salmerón (2011) implement the SEM algorithm for the approximation of the asymptotic variance of the computed ML projection parameter estimators, from incomplete ARH(1) data. Note that this estimation methodology provides an extension, allowing the incorporation of interaction between different subjects, of the classical approaches to handling missing data in longitudinal data analysis (see, for example, Aalen and Gunnes, 2010, on reconstruction of missing longitudinal data considering the linear increment model, and Nakai and Ke, 2011, on a review on longitudinal data analysis with missing data). Recently, the spatial autoregressive Hilbertian processes of order one (SARH(1) processes) have been introduced in Ruiz-Medina (2011). The non-parametric estimation of the operators involved in the SARH(1) state equation has been derived in Ruiz-Medina (2012a). The extension of the presented ML-based parameter estimation approach to this context will constitute the topic of a subsequent paper.

In this paper, the asymptotic distribution of the maximum likelihood estimators of the auto-

covariance operator of the Hilbert-valued innovation process, and of the autocorrelation operator of a Gaussian ARH(1) process is derived, under suitable conditions. An invariance principle for the Robbins-Monro process in a Hilbert space is applied (see Walk, 1977). Recent results on *L^p-m-approximability*, derived by Hörmann and Kokoszka (2010), are also considered in the application of this invariance principle. The illustration of the performance of the maximum likelihood projection estimation methodology in the ARH(1) framework, from incomplete functional data, is achieved in terms of a real-data example in the financial context. Specifically, the indebtedness of 638 Spanish companies in the period 1999-2007 is estimated by combination of the forward and backward Kalman filtering with the EM algorithm, in terms of projections.

The outline of the paper is the following. Section A7.2 provides the preliminary elements and results used in the derivation of the Central Limit Theorem of this paper: the theory of ARH(1) processes, the invariance principle applied, as well as some results on *L^p-m-approximability* by Hörmann and Kokoszka (2010). The weak-sense formulation of the functional log-likelihood of the Gaussian ARH(1) innovation process is established from the theory of Gaussian measures on Hilbert spaces, as described in Section A7.6. The derivation of the asymptotic distribution of the ML parameter estimators in the ARH(1) framework can be found in Section A7.7. The computation algorithm for ML projection estimation from missing data, proposed in Ruiz-Medina and Salmerón (2010), is described in Section A7.8. The performance of this algorithm is illustrated with a real-data example, where functional estimation of indebtedness is achieved from the analysis of a panel of 638 small Spanish companies in the period 1999-2007.

A7.2 Preliminaries

The ARH(1) process definition, and key preliminary results applied in the derivation of Theorem 10 below are now introduced.

A7.3 The ARH(1) model

The ARH(1) processes satisfy the following functional state equation (see Bosq, 2000):

$$Z_t = Y_t(\cdot) - \mu(\cdot) = \mathcal{A}\left(Z_{t-1}\right)(\cdot) + \nu_t(\cdot), \quad t \in \mathbb{Z}, \tag{A7.1}$$

where $\mu, Y_t, \nu_t \in H$, $t \in \mathbb{Z}$. The innovation process ν is assumed to be a zero-mean H-valued martingale difference process, or a zero-mean strong Hilbertian white noise. Here, H is a separable Hilbert space. The auto-covariance operator of the Hilbert-valued innovation process will be denoted as $R_\nu = E[\nu_t \otimes \nu_t]$, $t \in \mathbb{Z}$. The infinite-dimensional parameter $\mathcal{A}$ is the autocorrelation operator, which is assumed to belong to the space $\mathcal{L}(H)$ of bounded linear operators on H (see Bosq, 2000). Since R_ν belongs to the class of Hilbert-Schmidt operators on H, our functional parameter space where the vector of unknown operators $(\mathcal{A}, R_\nu)$ lies is $\mathcal{L}(H) \times \mathcal{S}(H)$, with $\mathcal{S}(H)$ denoting the space of Hilbert-Schmidt operators on H.

Remark 21. *In Section A7.8, in the implementation of the proposed Maximum Likelihood Estimation algorithm, initialization is achieved by considering H-valued random variables uncorrelated with the innovation process ν. Also, it is assumed that the effect of these random initial conditions is negligible at times where observations are available.*

A7.4 Invariance principle and L^p-m-*approximation*

The following invariance principle for the Robbins-Monro process in a Hilbert space (see Walk, 1977) is applied in the derivation of the main result in the next section.

Lemma 2. *Let $\{V_n\}_{n\in\mathbb{N}}$ be a martingale difference sequence of H-valued random variables with $E\|V_n\|^2 < \infty$. Let further $\mathcal{R} : H \longrightarrow H$ be a trace operator, and let $\mathcal{R}^n$ be the covariance operator of V_n given $V_1, \ldots, V_{n-1}$ ($n \in \mathbb{N}$). Suppose*

$$E\left\|\frac{1}{n}\sum_{j=1}^{n}\mathcal{R}^j - \mathcal{R}\right\| \longrightarrow 0, \quad n \longrightarrow \infty,$$

$$\frac{1}{n}\sum_{j=1}^{n} E\left\|V_j\right\|^2 \longrightarrow trace(\mathcal{R}), \quad n \longrightarrow \infty,$$

and, for any $r > 0$,

$$\frac{1}{n}\sum_{j=1}^{n} E\left(\|V_j\|^2 \chi\{\|V_j\|^2 \geq rj\} | V_1, \ldots, V_{j-1}\right) \underset{P}{\longrightarrow} 0, \quad (n \longrightarrow \infty).$$

Then, the H-valued series $S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j$ *converges in distribution to a Gaussian* H*-valued random variable with zero-mean and covariance operator* $\mathcal{R}$*, as* $n \longrightarrow \infty$.

In the verification of the conditions assumed in the above lemma on the conditional covariance operator of the martingale difference sequence, the following definition and results on L^p*-m-approximation* will be applied (see Hörmann and Kokoszka, 2010).

Definition 10. *A Hilbert-valued sequence on the probability space* $(\Omega, \mathcal{A}, P)$ *is said to be* L^p*-m-*approximable *if each element* X_n *of the sequence admits the following representation:*

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots),$$

where ε_i *are* $S-$*valued (possibly* $S = H$*) and* f *is a measurable function* $f : S^{\infty} \longrightarrow H$*. Assume also that, for each* n*, a copy* $\{\varepsilon_i^{(n)}\}$ *independent of* $\{\varepsilon_i\}$ *is available on the same probabilistic space* $(\Omega, \mathcal{A}, P)$*. Then, we have*

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon_{n-m}^{(n)}, \varepsilon_{n-m-1}^{(n)}, \ldots)$$

$$\sum_{m=1}^{\infty} \left(E\|X_m - X_m^{(m)}\|^p\right)^{1/p} < \infty.$$

Process $\{X_n^{(m)},\ n \in \mathbb{Z}\}$ *is stationary, and* $X_n \underset{d}{=} X_n^{(m)}$*, which is m-dependent, i.e., for every* k*,* $\mathcal{F}_k^-$ *and* $\mathcal{F}_{k+m}^+$ *are independent, with*

$$\mathcal{F}_k^- = \sigma\{\ldots X_{k-2}^{(m)}, X_{k-1}^{(m)}, X_k^{(m)}\}, \quad and \quad \mathcal{F}_k^+ = \sigma\{X_k^{(m)}, X_{k+1}^{(m)}, X_{k+2}^{(m)} \ldots\}.$$

Lemma 3. *Let $\{Z_n\}$ be a zero-mean ARH(1) sequence satisfying:*

(i) $\mathcal{A} \in \mathcal{L}(H)$

(ii) $\|\mathcal{A}\|_{\mathcal{L}(H)} < 1$

(iii) ε_n *is a zero-mean* L_H^2 *innovation process*

(iv) $[E\|\varepsilon_0\|^p]^{1/p} < \infty, \quad p \geq 2.$

Then, the sequence $\{Z_n\}$ *is* L^p-m-approximable

Lemma 4. *Let* $\{X_n\} \subset L_H^4$ *be* L^4-m-approximable *with auto-covariance operator* C*. Then, there exists a constant* $U_X < \infty$*, which does not depend on* N*, such that*

$$NE\|\widehat{C}_N - C\|^2_{\mathcal{S}(H)} \leq U_X,$$

where, as before, $\mathcal{S}(H)$ *is the Hilbert space of Hilbert-Schmidt operators on* H*. Note that if* X_n *has zero mean, then, it can be considered*

$$U_X = E\|X\|^4 + 4\sqrt{2}[E\|X\|^4]^{3/4} \sum_{r=1}^{\infty} [E\|X_r - X_r^{(r)}\|^4]^{1/4}.$$

Here, $\widehat{C}_N$ *denotes the empirical auto-covariance operator of* $\{X_n\}$ *given by* $\widehat{C}_N = \frac{1}{N}\sum_{i=1}^N Z_i \otimes Z_i$*, with* $Z_i = X_i - E[X_i]$*,* $i = 1, \ldots, N$*.*

A7.5 Asymptotic normality of ML parameter estimators

This section provides the asymptotic Gaussian distribution of the maximum likelihood estimator of the auto-covariance operator R_ν of the innovation process ν, under suitable conditions. The ML estimator of the autocorrelation operator $\mathcal{A}$ is then derived as a function of the ML estimator $\widehat{R}_\nu$ of R_ν, and its asymptotic distribution as well.

A7.6 Gaussian measures on Hilbert spaces

Let us first consider some fundamental definitions and elements from the theory of Gaussian measures on Hilbert spaces (see, for example, Da Prato and Zabczyk, 2002, Chapter 1). In the following, we adopt Da Prato and Zabczyk's (2002) notation for Gaussian measures. Specifically, for the real-valued case, the Gaussian measure with mean a and variance zero is formally represented as $N_{a,0}(dx) = \delta_a(dx)$, with δ_a being the Dirac measure at a. Moreover, for the case of positive variance λ, the corresponding gaussian measure is denoted as $N_{a,\lambda}(dx) = \frac{1}{\sqrt{2\pi\lambda}} \exp\left(\frac{-(x-a)^2}{2\lambda}\right) dx$, with characteristic function $\widehat{N_{a,\lambda}}(h) = \int_{\mathbb{R}} \exp(ihx)\, N_{a,\lambda}(dx) = \exp\left(iah - \frac{1}{2}\lambda h^2\right)$, for $h \in \mathbb{R}$.

In the Hilbert-valued context, for a given non-negative symmetric operator Q on H in the trace class, i.e., $Q \in L_1^+(H)$, with H, as before, being a separable Hilbert space, and for a function $a \in H$, we will denote as $N_{a,Q}$ the measure on H satisfying

$$\widehat{N_{a,Q}}(h) = \int_H \exp(ihx)\, N_{a,Q}(dx) = \exp\left(i\langle a, h\rangle_H - \frac{1}{2}\langle Qh, h\rangle_H\right), \quad h \in H. \tag{A7.2}$$

Let us now consider $\mathcal{B}(H)$ to be the σ-algebra of all Borel subsets of H, and the identification $x \in H \longrightarrow \gamma(x) = (x_k) \in l^2$, where $x_k = \langle x, e_k\rangle_H$, $k \geq 1$, are the projections of function $x \in H$ into the eigenvectors e_k, $k \geq 1$, of Q, satisfying $Qe_k = \lambda_k(Q)e_k$, $k \geq 1$, which constitute a complete orthonormal system in H. The cylindrical subsets of H can be defined via such an identification as $I = \{x \in H :\ P_n x = (x_1 \ldots x_n) \in B\}$, where $B \in \mathcal{B}(\mathbb{R}^n)$, with $\mathcal{B}(\mathbb{R}^n)$ being the σ-algebra of all Borel subsets of $\mathbb{R}^n$. It can be easily proved that the σ-algebra generated by all cylindrical subsets of H coincides with $\mathcal{B}(H)$.

Theorem 9. *Let $a \in H$, and $Q \in L_1^+(H)$, then, there exists a unique probability measure μ on $(H, \mathcal{B}(H))$ such that*

$$\int_H \exp\langle h, x\rangle_H\, \mu(dx) = \exp\left(i\langle a, h\rangle_H - \frac{1}{2}\langle Qh, h\rangle_H\right), \quad h \in H.$$

Moreover μ is the restriction to H (identified with l^2) of the product measure $\prod_{k=1}^{\infty} \mu_k =$

$\prod_{k=1}^{\infty} N_{a_k,\lambda_k(Q)}$, *defined on* $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, *with* $a_k = \langle a, e_k \rangle_H$, $Qe_k = \lambda_k(Q)e_k$, *for* $k \geq 1$. *Thus,* $\mu = N_{a,Q}$, *and for* $a \equiv 0$, $\mu = N_Q$.

A7.7 Central limit theorem

Assume now that ν, the zero-mean Gaussian H-valued innovation process, have trace auto-covariance operator $R_\nu \in L_1^+(H)$, from Theorem 9, there exists a unique probability measure N_{R_ν} on $(H, \mathcal{B}(H))$ such that $N_{R_\nu} = \prod_{k=1}^{\infty} N_{0,\lambda_k(R_\nu)}$, with $R_\nu e_k = \lambda_k(R_\nu)\, e_k$, $k \geq 1$. Thus, formally, the functional log-likelihood associated with ν admits the following expression, with respect to the Gaussian measure N_{R_ν} on the Hilbert space H : For $\nu_t \sim \exp(\Theta)$, with $t \in \mathbb{Z}$ fixed, and with $\Theta = R_\nu \in \mathcal{S}(H)$, denoting by $\nu \in H$ the values of ν_t,

$$\begin{aligned} \log f(\nu|R_\nu) &= \log\left(\prod_{k=1}^{\infty} \frac{1}{\sqrt{2\pi\lambda_k(R_\nu)}}\right) - \sum_{k=1}^{\infty} \frac{\nu_k^2}{2\lambda_k(R_\nu)} \\ &= \log\left(\left|\frac{1}{\sqrt{2\pi}} R_\nu^{-1/2}\right|\right) - \frac{1}{2} R_\nu^{-1}(\nu)(\nu) \\ &= \log\left(\left|\frac{1}{\sqrt{2\pi}} R_\nu^{-1/2}\right|\right) - \frac{1}{2} \langle \nu \otimes \nu, \varphi(R_\nu) \rangle_{\mathcal{S}(H)} , \end{aligned} \tag{A7.3}$$

where $\nu_k = \langle \nu, \phi_k \rangle_H$, $k \geq 1$, and, for an operator $\mathcal{C}$ on H, $|\mathcal{C}| = \prod_{k=1}^{\infty} \lambda_k(\mathcal{C})$. Here, as before, $\lambda_n(R_\nu)$, $n \geq 1$, denote the eigenvalues of operator R_ν, and $\varphi(R_\nu) = R_\nu^{-1}$. From now on, we will denote by Φ^* the projection operator into the eigenvector system $\{\phi_k\}_{k\geq 1}$, i.e., $\Phi^* x = \gamma(x) = (x_k) \in l^2$. Note that in the case where R_ν^{-1} is bounded on a dense subspace H_1 of H $\sum_{k=1}^{\infty} \frac{x_k^2}{2\lambda_k} < \infty$, for all $x \in H_1$. In particular, H_1 can be generated from $\{\phi_k,\ k \geq 1\}$. In general, equation (A7.3) must be understood in the weak-sense, i.e., in the sense of tempered distributions, as the inverse Fourier transform of its characteristic functional (A7.2), given, in this case, by

$$\widehat{N_{0,R_\nu}}(h) = \exp\left(-\frac{1}{2} \langle R_\nu h, h \rangle_H\right) = \exp\left(-\frac{1}{2} \sum_{k=1}^{\infty} \lambda_k(R_\nu) h_k^2\right) < \infty,$$

for $k \geq 1$, with $h_k = \langle h, \phi_k \rangle_H$, in view of the Cauchy—Schwarz inequality, since $\sum_{k=1}^{\infty} \lambda_k(R_\nu) < \infty$, by the trace property of R_ν, and $\sum_{k=1}^{\infty} h_k^2 = \|h\|_H^2 < \infty$, by Parseval identity, since $h \in H$.

Remark 22. *In the case where ν_t, $t \geq 0$, are independent, equation (A7.3) leads to the following formal expression, understood in the weak sense:*

$$\begin{aligned} \log f(\nu|\nu_1, \ldots, \nu_T, R_\nu) &= T \log \left(\left| \frac{1}{\sqrt{2\pi}} R_\nu^{-1/2} \right| \right) - \frac{T}{2} R_\nu^{-1}(\nu)(\nu) \\ &= T \log \left(\left| \frac{1}{\sqrt{2\pi}} R_\nu^{-1/2} \right| \right) - \frac{T}{2} \langle \nu \otimes \nu, \varphi(R_\nu) \rangle_{\mathcal{S}(H)} . \end{aligned}$$

Theorem 10. *Let Z be a zero-mean Gaussian ARH(1) process, satisfying the conditions assumed in Lemma 3 with $p = 4$, and with $\{\nu_t \otimes \nu_t\}_{t \in \mathbb{N}}$, being an $\mathcal{S}(H)$-valued martingale difference sequence. When $T \longrightarrow \infty$, suppose that*

$$\frac{1}{T-1} \sum_{t=2}^{T} \left\| E\left[diag(\Phi^*(\nu_{t-1}) \otimes \Phi^*(\nu_{t-1}))^{\otimes 4} \right] \right\|_{l^1(\mathbb{N}^4)} \longrightarrow 0,$$

where $f^{\otimes 4}$ denotes the 4th self-tensorial product of f. Furthermore, ν is assumed to admit the following factorization: $\nu_t = g(\nu_{t-1})X_t$, where $X_t \in H$ are i.i.d. zero-mean H-valued Gaussian random variables with finite functional variance $E\|X_t\|^2 < \infty$. In this factorization, X_t is independent of $\nu_{t-1}, \ldots, \nu_1$, for all $t > 0$, and, for a certain orthonormal basis $\{\phi_k\}_{k \in \mathbb{N}_*}$ of H, the random operator $g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1})$ admits the representation:

$$\begin{aligned} &g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1}) \\ &\qquad = \frac{\sum_{k,l,p,q} \nu_{t-1}(\phi_k)\nu_{t-1}(\phi_l)\nu_{t-1}(\phi_p)\nu_{t-1}(\phi_q)[\phi_k \otimes \phi_l \otimes \phi_p \otimes \phi_q]}{\sum_k E[X_t(\phi_k)]^4 [\phi_k \otimes \phi_k \otimes \phi_k \otimes \phi_k]}, \end{aligned}$$

in the space $\mathcal{L}^1(\mathcal{S}(H \otimes H))$ of absolute integrable $\mathcal{S}(H \otimes H)$-valued random variables with respect to the corresponding probability measure. Here, $\mathcal{S}(H \otimes H)$ denotes, as before, the space of Hilbert-Schmidt operators on $H \otimes H$. Then,

$$\sqrt{T}\left(\widehat{\Theta} - \Theta\right) \underset{d}{\longrightarrow} Y \sim \mathcal{N}\left(0, \mathcal{R}_{T_\nu}\right), \quad T \to \infty,$$

where, as before, $\Theta = R_\nu$, $\widehat{\Theta} = \widehat{R}_\nu = \frac{1}{T}\sum_{i=1}^{T} \nu_i \otimes \nu_i$, *and* $\mathcal{R}_{T_\nu}$ *is the covariance operator of the sufficient statistics* $T_\nu = \nu \otimes \nu$.

Proof In the verification of the conditions assumed in Lemma 2, consider the $\mathcal{S}(H)$-valued martingale difference sequence $\{V_n = T_n(\nu) = \nu_n \otimes \nu_n\}_{n\in\mathbb{N}}$. The conditional covariance operator of the sequence is computed as follows:

$$
\begin{aligned}
& E[\nu_t \otimes \nu_t \otimes \nu_t \otimes \nu_t | \nu_{t-1} \otimes \nu_{t-1} \ldots \nu_1 \otimes \nu_1] \\
= \; & E[g(\nu_{t-1})X_t \otimes g(\nu_{t-1})X_t \otimes g(\nu_{t-1})X_t \otimes g(\nu_{t-1})X_t | \nu_{t-1} \otimes \nu_{t-1} \ldots \nu_1 \otimes \nu_1] \\
& = \sum_{k,l,p,q,m,r,u,v} [\phi_k \otimes \phi_l \otimes \phi_p \otimes \phi_q][\phi_m \otimes \phi_r \otimes \phi_u \otimes \phi_v] \\
& \times E\left[g(\nu_{t-1})(\phi_k) g(\nu_{t-1})(\phi_l) g(\nu_{t-1})(\phi_p) g(\nu_{t-1})(\phi_q)\right. \\
& \left. \times X_t(\phi_m) X_t(\phi_r) X_t(\phi_u) X_t(\phi_v) | \nu_{t-1} \otimes \nu_{t-1} \ldots \nu_1 \otimes \nu_1\right] \\
& = \sum_{k,l,p,q} g(\nu_{t-1})(\phi_k) g(\nu_{t-1})(\phi_l) g(\nu_{t-1})(\phi_p) g(\nu_{t-1})(\phi_q) [\phi_k \otimes \phi_l \otimes \phi_p \otimes \phi_q] \\
& \times \sum_{m,r,u,v} E[X_t(\phi_m) X_t(\phi_r) X_t(\phi_u) X_t(\phi_v)][\phi_m \otimes \phi_r \otimes \phi_u \otimes \phi_v] \\
& = g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1}) \otimes g(\nu_{t-1}) \\
& \times \sum_{m,r,u,v} \left[\delta_{m,r}\delta_{u,v} + \delta_{m,u}\delta_{r,v} + \delta_{m,v}\delta_{r,u}\right] [\phi_m \otimes \phi_r \otimes \phi_u \otimes \phi_v] \\
& = \frac{\sum_{k,l,p,q} \nu_{t-1}(\phi_k)\nu_{t-1}(\phi_l)\nu_{t-1}(\phi_p)\nu_{t-1}(\phi_q)[\phi_k \otimes \phi_l \otimes \phi_p \otimes \phi_q]}{\sum_p E[X_t(\phi_p)]^4 [\phi_p \otimes \phi_p \otimes \phi_p \otimes \phi_p]} \\
& \times \sum_p E[X_t(\phi_p)]^4 [\phi_p \otimes \phi_p \otimes \phi_p \otimes \phi_p] = \nu_{t-1} \otimes \nu_{t-1} \otimes \nu_{t-1} \otimes \nu_{t-1}.
\end{aligned}
$$

Furthermore, from the above equation,

$$
\frac{1}{T}\sum_{t=2}^{T+1} E[\nu_t \otimes \nu_t \otimes \nu_t \otimes \nu_t | \nu_{t-1} \otimes \nu_{t-1} \ldots \nu_1 \otimes \nu_1] = \frac{1}{T}\sum_{t=2}^{T+1} \nu_{t-1} \otimes \nu_{t-1} \otimes \nu_{t-1} \otimes \nu_{t-1} = \frac{1}{T}\sum_{t=1}^{T} \mathcal{R}^t.
$$

Considering Lemma 2.1 of Hörmann and Kokoszka (2010), $\nu_t = Z_t - \mathcal{A}Z_{t-1}$ is L^p-m-approximable. Furthermore, $Z_t \otimes Z_t$ is an AR$\mathcal{S}(H)(1)$ process (see Bosq, 2000), then, from Lemma 3, is L^p-m-approximable, in particular, with $p = 4$. Therefore, applying again Lemma 2.1

of Hörmann and Kokoszka (2010), $\nu_t \otimes \nu_t = Z_t \otimes Z_t - \mathcal{A}Z_{t-1} \otimes \mathcal{A}Z_{t-1}$ is also L^p-m-approximable, and for $p = 4$, we obtain from Lemma 4, the convergence to zero, as $T \to \infty$, of

$$E \left\| \frac{1}{T} \sum_{t=1}^{T} \mathcal{R}^t - \mathcal{R}_{T_\nu} \right\|_{\mathcal{S}(H \otimes H)} = E \left\| \frac{1}{T} \sum_{t=1}^{T} \nu_t \otimes \nu_t \otimes \nu_t \otimes \nu_t - E[\nu_t \otimes \nu_t \otimes \nu_t \otimes \nu_t] \right\|_{\mathcal{S}(H \otimes H)}.$$

Hence, $\frac{1}{T} \sum_{t=1}^{T} E \left\| \nu_t \otimes \nu_t \right\|^2 \longrightarrow \text{trace}(E[\nu_t \otimes \nu_t \otimes \nu_t \otimes \nu_t])$, when $T \longrightarrow \infty$.

Finally, the convergence to zero, as $T \longrightarrow \infty$, of

$$\frac{1}{T-1} \sum_{t=2}^{T} \left\| E \left[\text{diag}(\Phi^*(\nu_{t-1}) \otimes \Phi^*(\nu_{t-1}))^{\otimes 4} \right] \right\|_{l^1(\mathbb{N}^4)}$$

implies the convergence to zero in probability of

$$\frac{1}{T} \sum_{t=1}^{T} E \left(\|\nu_t \otimes \nu_t\|^2 \chi\{\|\nu_t \otimes \nu_t\|^2 \geq rt\} | \nu_1 \otimes \nu_1, \ldots, \nu_{t-1} \otimes \nu_{t-1} \right).$$

Thus, all the conditions formulated in Lemma 2 are satisfied, and hence, the $\mathcal{S}(H)$-valued series $S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \nu_t \otimes \nu_t$ converges in distribution to a Gaussian $\mathcal{S}(H)$-valued random variable with zero-mean and covariance operator $\mathcal{R}_{T_\nu}$, as $T \longrightarrow \infty$.

Corollary 2. *Assume that $\mathcal{A}$ is a symmetric positive compact operator with $\|\mathcal{A}\|_{\mathcal{L}(H)} < 1$, where $\|\cdot\|_{\mathcal{L}(H)}$ denotes the norm in the space of linear bounded operators. Assume also that the eigenvectors of $\mathcal{A}$ coincide with the eigenvectors of R_ν and R_Z, which are also positive symmetric operators. Then, the maximum likelihood estimator of the autocorrelation operator $\mathcal{A}$ is given by:*

$$\begin{aligned} \widehat{\mathcal{A}}_T &= \left[I - \frac{1}{T} \sum_{t=1}^{T} \left\langle \nu_t \otimes \nu_t, R_Z^{-1} \right\rangle_{\mathcal{S}(H)} \right]^{1/2} = \left[I - \left\langle \frac{1}{T} \sum_{t=1}^{T} \nu_t \otimes \nu_t, R_Z^{-1} \right\rangle_{\mathcal{S}(H)} \right]^{1/2} \\ &= \left[I - \left\langle \widehat{R}_\nu^T, R_Z^{-1} \right\rangle_{\mathcal{S}(H)} \right]^{1/2}, \end{aligned} \tag{A7.4}$$

where I denotes the identity operator on H, $\widehat{R}_\nu^T$ is the ML of R_ν, and $R_Z = E[Z_t \otimes Z_t]$, for every $t \geq 0$, with $Z_t = Y_t - \mu$, and Y_t being an ARH(1) process satisfying equation (A7.1).

Remark 23. *Under the conditions of Theorem 10, replacing R_Z by the empirical auto-covariance operator $\widehat{R}_Z^T = \frac{1}{T}\sum_{t=1}^T Z_t \otimes Z_t$, the maximum likelihood estimator of the auto-covariance operator R_Z of the zero-mean Gaussian process Z. The asymptotic distribution of*

$$\widetilde{\mathcal{A}} = \left[I - \left\langle \frac{1}{T}\sum_{t=1}^{T} \nu_t \otimes \nu_t, [\widehat{R}_Z^T]^{-1} \right\rangle_{\mathcal{S}(H)} \right]^{1/2}$$

can be obtained from the asymptotic distribution of $\widehat{R}_Z^T$ (see Bosq, 2000, Theorem 4.9, p. 119), and the derived Gaussian distribution of the ML estimator $\widehat{R}_\nu^T = \frac{1}{T}\sum_{t=1}^T \nu_t \otimes \nu_t$ of R_ν obtained in Theorem 10.

Proof Note that $\{\phi_k,\ k \geq 1\}$ constitutes the common orthonormal system of eigenvectors of $\mathcal{A}$, R_Z and R_ν, that is, $\mathcal{A}\phi_k = \lambda_k(\mathcal{A})\phi_k$, $R_Z\phi_k = \lambda_k(R_Z)\phi_k$ and $R_\nu\phi_k = \lambda_k(R_\nu)\phi_k$, for $k \geq 1$. Furthermore, from equation (A7.1), $R_\nu = E[\nu_0 \otimes \nu_0] = E[\nu_t \otimes \nu_t] = R_Z - \mathcal{A}R_Z\mathcal{A}$. Applying the Spectral Theorem on Spectral Calculus (see, for example, Dautray and Lions, 1985, pp. 119 and 126), we then obtain, for $k \geq 1$, $\lambda_k(R_\nu) = \lambda_k(R_Z) - \lambda_k^2(\mathcal{A})\lambda_k(R_Z) = \lambda_k(R_Z)(1 - \lambda_k^2(\mathcal{A}))$. Therefore, $\widehat{\lambda_k(\mathcal{A})} = \sqrt{1 - \frac{\widehat{\lambda_k(R_\nu)}}{\lambda_k(R_Z)}}$, $k \geq 1$. Again, from spectral representation theorem, the maximum likelihood estimator $\widehat{\mathcal{A}}^T$ of $\mathcal{A}$ is given by

$$\widehat{\mathcal{A}}^T = \sum_{k=1}^{\infty} \widehat{\lambda_{kT}(\mathcal{A})}[\phi_k \otimes \phi_k] = \left[I - \left\langle \widehat{R}_\nu^T, R_Z^{-1} \right\rangle_{\mathcal{S}(H)} \right]^{1/2},$$

where $\widehat{R}_\nu^T = \frac{1}{T}\sum_{t=1}^T \nu_t \otimes \nu_t$ is the ML estimator of R_ν.

A7.8 Computation algorithm

In several applied fields, functional estimation must usually be addressed with missing information. The estimation algorithm described in this section provides an approximation to the ML estimator of ARH(1) process Z from incomplete data. The main iterated steps executed comprise the following algorithms: Forward and backward Kalman filtering, and EM algorithm

(see Ruiz-Medina and Salmerón, 2010). We now proceed to provide some details on the implementation of these algorithms from projection into suitable biorthogonal bases of the space H.

Operator $\mathcal{A}$ is assumed to be a compact operator, possibly non-symmetric, admitting the following spectral decomposition: $\mathcal{A}\psi_i = \lambda_i(\mathcal{A})\psi_i$, $\mathcal{A}^*\phi_i = \lambda_i(\mathcal{A})\phi_i$, $i \geq 1$, where $\mathcal{A}^*$ denotes the adjoint operator of $\mathcal{A}$, $\{\psi_i,\ i \geq 1\}$ and $\{\phi_i,\ i \geq 1\}$ are the left and right eigenvector systems associated with $\mathcal{A}$ and $\mathcal{A}^*$, respectively, and with the point spectrum $\{\lambda_i(\mathcal{A}),\ i \geq 1\}$. They are biorthogonal, that is, $\mathbf{\Phi}\mathbf{\Psi}^* = I$, with I representing, as before, the identity operator on H. Here, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are the respective projection operators into $\{\phi_i :\ i \geq 1\}$ and $\{\psi_i :\ i \geq 1\}$. Indeed, applying operator $\mathbf{\Phi}^*$ to both left-hand-sides of equation (A7.1), we obtain the infinite-dimensional equation system: $a_j(t) = \lambda_j(\mathcal{A})a_j(t-1) + \nu_j(t)$, $j \geq 1$, in terms of $a_j(t) = \langle Z_t(\cdot), \phi_j(\cdot)\rangle_H$, and $\nu_j(t) = \langle \nu_t(\cdot), \phi_j(\cdot)\rangle_H$, for $t \geq 0$, and $j \geq 1$. Truncation at term M leads to the following finite-dimensional approximation (see Ruiz-Medina and Salmerón, 2010):

$$\mathbf{a}(t) = \mathbf{\Lambda}\mathbf{a}(t-1) + \boldsymbol{\nu}(t), \tag{A7.5}$$

where, for each time $t \in \mathbb{Z}$, $\mathbf{a}(t-k)_{M\times 1} = \mathbf{\Phi}_M^* Z_{t-k} = (a_1(t-k), \ldots, a_M(t-k))^*$, $k = 0, 1$, and $\boldsymbol{\nu}(t)_{M\times 1} = \mathbf{\Phi}_M^* \nu_t = (\nu_1(t), \ldots, \nu_M(t))^*$. Here, $\mathbf{\Lambda}$ denotes the diagonal $M \times M$ matrix with entries the eigenvalues $\lambda_j(\mathcal{A}),\ j = 1, \ldots, M$.

The implementation of forward and backward Kalman filtering algorithm is now briefly described, considering equation (A7.5), and the functional observation model:

$$O_t = Z_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{A7.6}$$

where ε represents zero-mean strong $H-$white noise, that is, ε_t, $t \geq 0$, are independent and identically distributed H-valued random variables, and $E(\|\varepsilon_t\|_H^2) = \sigma_\varepsilon^2 < \infty$, for all $t \geq 0$. Process ε is also assumed to be uncorrelated with Z. Forward and backward Kalman filtering

are combined with EM algorithm as described below, considering the initial parameter values:

$$\begin{aligned}\widehat{Q} &= \boldsymbol{\Phi}^*\widehat{R}_{ZZ}\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\widehat{R}_{Z_0Z_1}\boldsymbol{\Phi}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Phi}^*\widehat{R}_{Z_1Z_0}\boldsymbol{\Phi} + \boldsymbol{\Lambda}\boldsymbol{\Phi}^*\widehat{R}_{ZZ}\boldsymbol{\Phi}\boldsymbol{\Lambda} \\ &\simeq \widehat{\boldsymbol{\Phi}}^*\widehat{R}_{ZZ}\widehat{\boldsymbol{\Phi}} - \widehat{\boldsymbol{\Phi}}^*\widehat{R}_{Z_0Z_1}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Lambda}} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Phi}}^*\widehat{R}_{Z_1Z_0}\widehat{\boldsymbol{\Phi}} + \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Phi}}^*\widehat{R}_{ZZ}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Lambda}} \\ \widehat{\mathcal{A}} &= \widehat{R}_{Z_0Z_1}\widehat{R}_{ZZ}^{-1}, \quad \boldsymbol{\Phi}\mathcal{A}\boldsymbol{\Psi}^* = \boldsymbol{\Lambda}, \quad \widehat{\boldsymbol{\Phi}}\widehat{\mathcal{A}}\widehat{\boldsymbol{\Psi}^*} = \widehat{\boldsymbol{\Lambda}}, \end{aligned} \tag{A7.7}$$

where, as before, $\widehat{R}_{ZZ} = \frac{1}{T}\sum_{t=1}^{T} Z_t \otimes Z_t$ and $\widehat{R}_{Z_0Z_1} = \frac{1}{T-1}\sum_{t=1}^{T-1} Z_t \otimes Z_{t+1}$.

Given the observations of process O up to time t, the implementation of forward Kalman filtering is achieved by considering $\widehat{\mathbf{a}}(t|t) = \widehat{\mathbf{a}}(t|t-1) + \mathbf{K}_t\left(O_t - \boldsymbol{\Psi}\widehat{\mathbf{a}}(t|t-1)\right)$, where $\widehat{\mathbf{a}}(t|t) = E\left(\mathbf{a}(t)|O_t,\ldots,O_1\right)$ and $\widehat{\mathbf{a}}(t|t-1) = E\left(\mathbf{a}(t)|O_{t-1},\ldots,O_1\right)$. Here,

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\boldsymbol{\Phi}^*\left(R_\varepsilon + \boldsymbol{\Psi}\mathbf{P}_{t|t-1}\boldsymbol{\Psi}^*\right)^{-1},$$

in terms of

$$\mathbf{P}_{t|t-1} = \mathrm{var}\left(\mathbf{a}(t)|O_{t-1},\ldots,O_1\right) = \boldsymbol{\Lambda}\mathbf{P}_{t-1|t-1}\boldsymbol{\Lambda} + \widehat{Q},$$

with $\mathbf{P}_{t|t} = E\left(\left(\mathbf{a}(t) - \hat{\mathbf{a}}(t|t)\right)\left(\mathbf{a}(t) - \hat{\mathbf{a}}(t|t)\right)^*\right)$. The functional mean-square error is approximated from $\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t\boldsymbol{\Psi}\mathbf{P}_{t|t-1}$, and the one-step-ahead predictor is computed as $\widehat{\mathbf{a}}(t|t-1) = \boldsymbol{\Lambda}\widehat{\mathbf{a}}(t-1|t-1)$. The initial values considered are $\widehat{\mathbf{a}}(0|0) = \mathbf{0}$, and $\mathbf{P}_{0|0} = \mathbf{E}\left(\mathbf{a}(0)\mathbf{a}(0)^*\right)$, with $\mathbf{a}(0) = (a_1(0),\ldots,a_M(0))^*$ being the vector of projections of the random initial condition $Z_0(\cdot) \simeq \sum_{i=1}^{M} a_i(0)\psi_i(\cdot)$. Here, R_ε is the auto-covariance operator of the observation noise ε.

The previous projection estimates obtained from the implementation of the forward Kalman filtering are considered in the initialization of the backward Kalman smoothing, in terms of the following recursion:

$$\begin{aligned} &E\left(\mathbf{a}(t-1)|O_1,\ldots,O_T\right) = \widehat{\mathbf{a}}_{t-1|t-1} + \left(\mathbf{P}_{t-1|t-1}\widehat{\boldsymbol{\Lambda}^*}(\mathbf{P}_{t|t-1})^{-1}\right) \\ &\times \left(E\left(\mathbf{a}(t)|O_1,\ldots,O_T\right) - \widehat{\boldsymbol{\Lambda}}\widehat{\mathbf{a}}_{t-1|t-1}\right) \\ &\mathrm{Var}\left(\mathbf{a}(t-1)|O_1,\ldots,O_T\right) = \mathbf{P}_{t-1|t-1} + \left(\mathbf{P}_{t-1|t-1}\widehat{\boldsymbol{\Lambda}^*}(\mathbf{P}_{t|t-1})^{-1}\right) \\ &\times \left(\mathrm{Var}\left(\mathbf{a}(t)|O_1,\ldots,O_T\right) - \mathbf{P}_{t|t-1}\right)\left(\mathbf{P}_{t-1|t-1}\widehat{\boldsymbol{\Lambda}^*}(\mathbf{P}_{t|t-1})^{-1}\right)^*, \end{aligned}$$

for $t = T, \ldots, 1$, where, as before, * stands for transposition or the adjoint.

Finally, the Expectation and Maximization steps of the EM algorithm are derived as follows: Firstly, E-step is computed considering now as observations times, $t_1, \ldots, t_T$, such that the effect of the random initial condition is negligible. From standard results on quadratic forms, the conditional expectation of the 'projected complete data' given the 'projected incomplete data' is then obtained as (see Ruiz-Medina and Salmerón, 2010)

$$\begin{aligned} & C - \frac{T}{2}\log|\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M| - \frac{T}{2}\log|\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M| \\ & -\frac{1}{2}\mathrm{tr}\left\{(\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M)^{-1}(C_Z - \mathbf{\Lambda} B_Z^* - B_Z \mathbf{\Lambda} + \mathbf{\Lambda} A_Z \mathbf{\Lambda})\right\}, \\ & -\frac{1}{2}\mathrm{tr}\left\{(\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M)^{-1} C_\varepsilon\right\}, \end{aligned} \tag{A7.8}$$

where tr denotes the trace, and

$$\begin{aligned} C_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i)\mathbf{a}(t_i)^* | O_{t_1}, \ldots, O_{t_T}\right), \\ B_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i)\mathbf{a}(t_i-1)^* | O_{t_1}, \ldots, O_{t_T}\right), \\ A_Z &= \sum_{i=1}^{T} E\left(\mathbf{a}(t_i-1)\mathbf{a}(t_i-1)^* | O_{t_1}, \ldots, O_{t_T}\right), \\ C_\varepsilon &= \sum_{i=1}^{T} E\left(\mathbf{\Phi}_M^* \varepsilon_{t_i} (\mathbf{\Phi}_M^* \varepsilon_{t_i})^* | O_{t_1}, \ldots, O_{t_T}\right). \end{aligned}$$

Furthermore, the M-step is computed by differentiating (A7.8) with respect to $\mathbf{\Lambda}$, $\mathbf{\Phi}_M^* R_\nu \mathbf{\Phi}_M$, and $\mathbf{\Phi}_M^* R_\varepsilon \mathbf{\Phi}_M$, and applying basic differentiation properties of the trace, we obtain

$$\begin{aligned} \widehat{\mathbf{\Lambda}} &= \mathrm{diag}[B_Z](\mathrm{diag}[A_Z])^{-1}, \\ \mathbf{\Phi}_{M\times M}^* \widehat{R_\varepsilon \mathbf{\Phi}}_{M\times M} &= C_\varepsilon / T, \\ \mathbf{\Phi}_{M\times M}^* \widehat{R_\nu \mathbf{\Phi}}_{M\times M} &= \frac{1}{T}\left[C_Z - B_Z \widehat{\mathbf{\Lambda}} - \widehat{\mathbf{\Lambda}} B_Z^* + \widehat{\mathbf{\Lambda}} A_Z \widehat{\mathbf{\Lambda}}\right], \end{aligned}$$

where $\mathrm{diag}[A]$ denotes the diagonal of matrix A. Summarizing, the proposed estimation algorithm is implemented in terms of the following steps:

Step 1 Projection into the right eigenvector system of $\mathcal{A}$.

Step 2 Computation of the initial parameter values from the moment-based projection estimates (A7.7).

Step 3 Forward Kalman filtering.

Step 4 Backward Kalman smoothing.

Step 5 Expectation step of EM algorithm.

Step 6 Maximization step of EM algorithm.

Note that Steps 3-6 are iterated, since the output parameter estimates from the EM algorithm are considered for implementation of the forward Kalman filtering recursion to improving estimations of the ARH(1) process Z. Backward Kalman smoothing is then computed from the conditional moments obtained in the last iteration of forward Kalman filtering. Finally, parameter estimates defined from the maximization step are calculated in terms of the outputs provided by the T iterations of Backward Kalman smoothing. Step 1-2 are needed for initialization of the computational algorithm described in this section.

A7.9 Real data example

In this section, the financing decisions of firms during the period 1999-2007 are analyzed in a panel constituted by 638 small Spanish companies, belonging to 9 different industry sectors, and located at 17 autonomous Spanish communities. Data have been collected from the SABI (*Sistema de Análisis de Balances Ibéricos*) database. The autoregressive dynamics of order one in time is assumed for the evolution of indebtedness of the 638 Spanish companies studied through the period 1999-2007. Thus, our functional data sequence is constituted by indebtedness surfaces constructed from the 638 observed values at each year of the period analyzed. The forward

and backward Kalman filtering, and EM algorithms for functional estimation of indebtedness surfaces at each year are implemented as described in the previous section. Note that, in this real-data application, the initial value considered for the diagonal autocovariance operator of the observation noise ε is given by $\widehat{R}_{\varepsilon}^{0} = \left[\mathrm{tr}(\widehat{R}_{OO}^{T}) - \mathrm{tr}(\widehat{R}_{\widetilde{Z}\widetilde{Z}}^{T})\right]\mathbf{I}$, where $\mathbf{I}$ denotes the identity operator, $\widetilde{Z}$ stands for the denoising and detrended functional data, O for the detrended and noisy functional data, tr for the trace, and $\widehat{R}_{OO}^{T} = \frac{1}{T}\sum_{i=1}^{T} O_i \otimes O_i$ and $\widehat{R}_{\widetilde{Z}\widetilde{Z}}^{T} = \frac{1}{T}\sum_{i=1}^{T} \widetilde{Z}_i \otimes \widetilde{Z}_i$ are the corresponding empirical autocovariance operators.

Stopping of the projection estimation algorithm is achieved after considering a number of iterations large enough to obtain a stabilization of the sequence of estimates. In the example analyzed, after 30 iterations of the forward and backward Kalman filtering in combination with the EM algorithm, the order of magnitude of the L^{∞}-norms of the realizations of the corresponding projection absolute error curves, for a truncation level $M = 512$, decreases slowly in comparison with the increasing of the number of iterations considered, as it can be appreciated in Figures A7.1 - A7.4. Thus, we can conclude that a stabilization is appreciated since 30 iterations of Steps 3-6 of the projection estimation algorithm. Indeed, in Figures A7.1-A7.3, a slow decreasing (less than 0.1 from the top to the bottom) of the order of magnitude of the L^{∞}-norms of the realizations of the projection absolute error curves can be appreciated, when the number of iterations is increased from 30 to 404 iterations, at each one of the years analyzed. The same fact can be appreciated in Figure A7.4, which displays the realizations of the projection absolute error curves through the years $2000, 2002, 2004, 2005, 2006$, and 2007, after 2000 iterations of the forward and backward Kalman filtering in combination with the EM algorithm. We obtain a decreasing of the order of magnitude of the projection absolute error L^{∞}-norms less than 0.2, while the number of iterations is increased from 404 (see bottom of Figures A7.1-A7.3) to 2000 (see Figure A7.4).

Figure A7.1: Realizations of the projection absolute error curves $|\mathbf{\Phi}_M^*(Z_t - \widehat{Z}_t)| = (|a_1(t) - \widehat{a}_1(t)|, \ldots, |a_M(t) - \widehat{a}_M(t)|)$, for $M = 512$ (horizontal axis, $M = 1, \ldots, 512$). Years $t = 1999$ (left), $t = 2000$ (center), $t = 2001$ (right) are displayed, for 30 iterations of Steps 3-6 at the top, and for 404 iterations at the bottom.

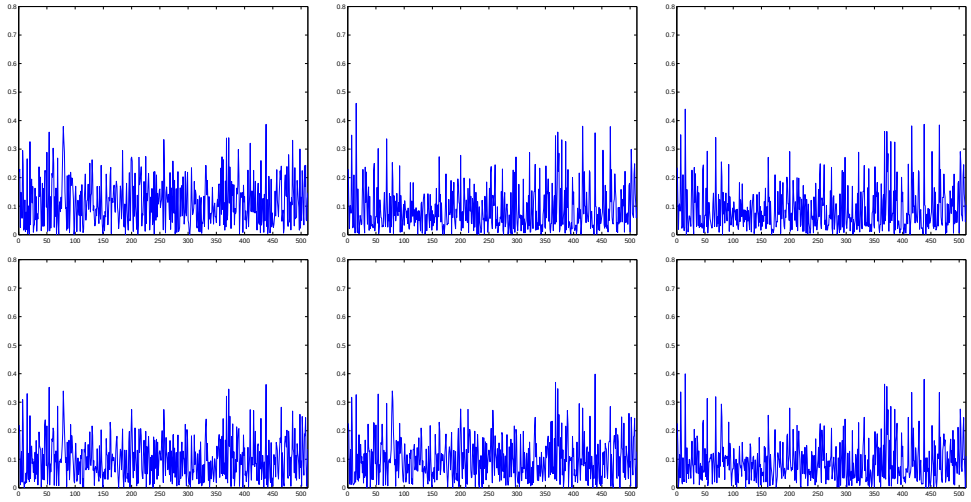


Figure A7.2: Realizations of the projection absolute error curves $|\mathbf{\Phi}_M^*(Z_t - \widehat{Z}_t)| = (|a_1(t) - \widehat{a}_1(t)|, \ldots, |a_M(t) - \widehat{a}_M(t)|)$, for $M = 512$ (horizontal axis, $M = 1, \ldots, 512$). Years $t = 2002$ (left), $t = 2003$ (center), $t = 2004$ (right) are displayed, for 30 iterations of Steps 3-6 at the top, and for 404 iterations at the bottom.

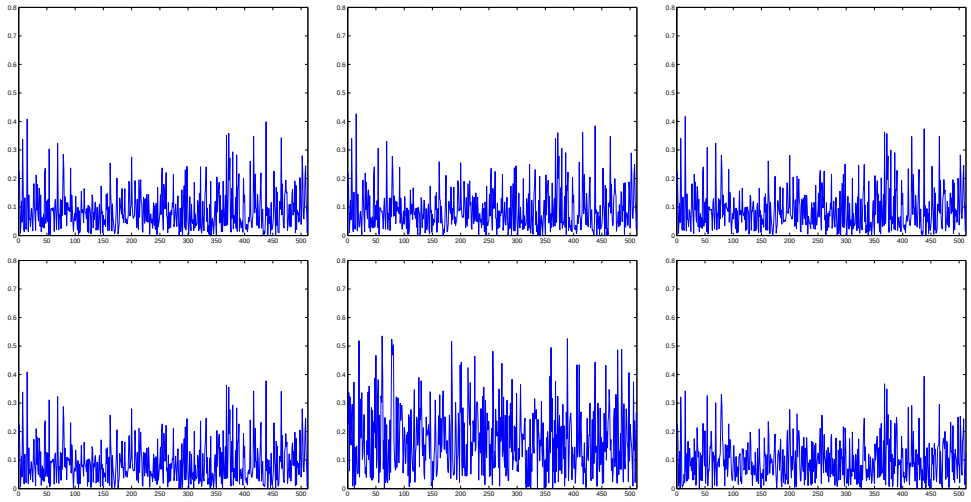


Figure A7.3: Realizations of the projection absolute error curves $|\mathbf{\Phi}_M^*(Z_t - \widehat{Z}_t)| = (|a_1(t) - \widehat{a}_1(t)|, \ldots, |a_M(t) - \widehat{a}_M(t)|)$, for $M = 512$ (horizontal axis, $M = 1, \ldots, 512$). Years $t = 2005$ (left), $t = 2006$ (center), $t = 2007$ (right) are displayed, for 30 iterations of Steps 3-6 at the top, and for 404 iterations at the bottom.

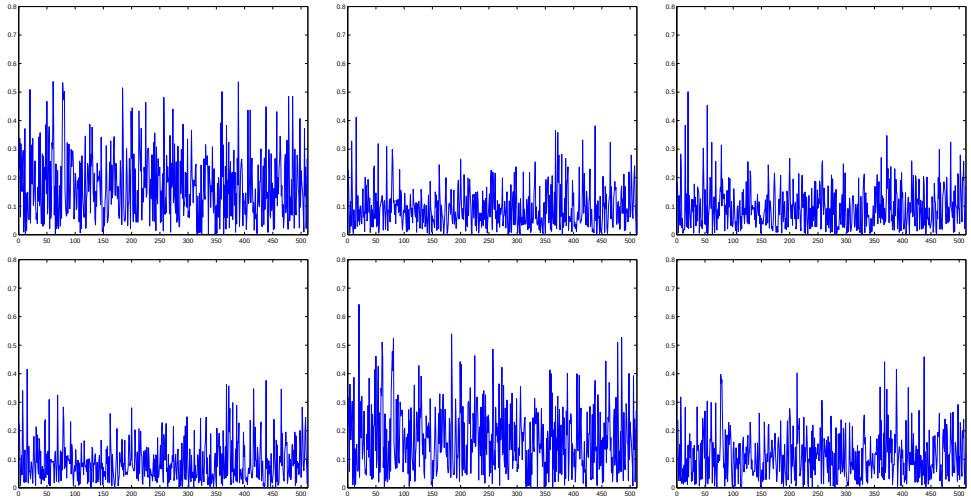
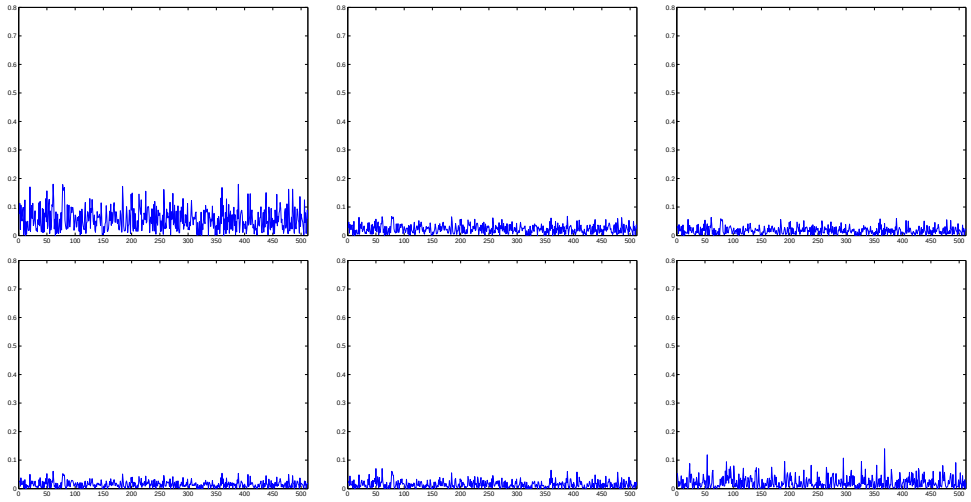


Figure A7.4: Realizations of the projection absolute error curves $|\mathbf{\Phi}_M^*(Z_t - \widehat{Z}_t)| = (|a_1(t) - \widehat{a}_1(t)|, \ldots, |a_M(t) - \widehat{a}_M(t)|)$, for $M = 512$ (horizontal axis, $M = 1, \ldots, 512$). Years $t = 2000, 2002, 2004, 2005, 2006, 2007$ are displayed from the top-left to the bottom-right, for 2000 iterations of Steps 3-6.



Appendix A8

Least-squares estimation of multifractional random fields in a Hilbert-valued context

Ruiz-Medina, M. D., Anh, V. V., Espejo, R. M., Angulo, J. M. and Frías, M. P. (2013)

Least-squares estimation of multifractional random fields in a Hilbert-valued context.

Journal of Optimization. Theory and Applications, DOI: 10.1007/s10957-013-0423-4.

Abstract

This paper derives conditions under which a stable solution to the least-squares estimation problem for multifractional random fields can be obtained. The observation model is defined in terms of a multifractional pseudodifferential equation. The weak and strong-sense formulations of this problem are studied through the spectral theory of pseudodifferential operators of variable order and their inverses. The theory of Reproducing Kernel Hilbert Spaces is applied. Numerical projection methods are also proposed based on this theory. A simulation study is developed to illustrate the estimation results derived.

This paper derives conditions under which a stable solution to the least-squares linear estimation problem for multifractional random fields can be obtained. The observation model is defined in terms of a multifractional pseudodifferential equation. The weak-sense and strong-

sense formulations of this problem are studied through the theory of fractional Sobolev spaces of variable order, and the spectral theory of multifractional pseudodifferential operators and their parametrix. The theory of Reproducing Kernel Hilbert Spaces is also applied to define a stable solution to the direct and inverse estimation problems. Numerical projection methods are proposed based on the construction of orthogonal bases of these spaces. Indeed, projection into such bases leads to a regularization, removing the ill-posed nature of the estimation problem. A simulation study is developed to illustrate the estimation results derived. Some open research lines in relation to the extension of the derived results to the multifractal process context are also discussed.

Data in many fields of applications display scaling, as in the case, for example, of the sample paths of fractional Brownian motion Mandelbrot and Van Ness (1968), or even multiscaling, for instance, of the type of multifractional Brownian motion (see Benassi, Jaffard, and Roux (1997); Ayache, and Lévy-Véhel (2000)). A way to characterize the multiscaling/multifractal behaviour of a random process or field is via the non-trivial singularity spectrum of its sample paths Jaffard (1999). A different, smoother, source of heterogeneity can be introduced through the theory of pseudodifferential operators of variable order, defining multifractional second-order random fields as given in Ruiz-Medina, Anh and Angulo (2004), extending multifractional Brownian motion (see also Anh and Leonenko, 2001; Kelbert, Leonenko, and Ruiz-Medina, 2005 and Leonenko, Ruiz-Medina and Taqqu, 2011 and the references therein, for the case of fractional-order pseudodifferential equations, including the spatiotemporal case).

The theory of random field estimation has been extensively developed in the last three decades (see, for example, Anh, Leonenko, Sakhno, 2004; Anh, Leonenko, Sakhno, 2007; Christakos, 2000; Ivanov, 1997; Leonenko, 1999; Ruiz-Medina, Angulo and Anh, 2003; Fernández-Pascual, Ruiz-Medina and Angulo, 2006). Recently, Functional Statistics is playing a crucial role in the framework of inference for stochastic processes and fields (see, for example, Bosq,

2000; Bosq and Blanke, 2007, Ruiz-Medina, 2011, 2012a; Ruiz-Medina and Espejo, 2012, 2013a; Ruiz-Medina and Salmerón, 2010). We adopt here this framework in the derivation of our main estimation results. In particular, we consider functional random variables with values in a fractional Sobolev space of variable order.

The theory of least-squares estimation of multifractional random fields in a general setting, extending the multifractional Brownian random field case, still remains an open problem, due to the difficulties arising from the heterogeneous local behavior of these random fields, reflected by the variable order characterizing the multifractional Sobolev space, which includes or can be isomorphically identified with the corresponding Reproducing Kernel Hilbert Space (RKHS). The results derived in this paper constitute an extension to the multifractional random field context of the results obtained in Ramm (2005).

Here, we consider the problem of minimizing the mean-square error in the estimation of the values of an output random field from the observation of a multifractional input random field, both being related by a multifractional pseudodifferential equation. Observations are assumed to be affected by additive noise with RKHS isomorphic to a fractional Sobolev space of variable order. The estimation problem admits a stable solution under some regularity assumptions on the multifractional bilinear form defining the inner product of the RKHS of the observation random field. These assumptions are satisfied when the norm of the covariance operator of the additive multifractional observation noise is dominated by the norm of the RKHS of the multifractional input random field, and the linear filter relating the input and output random fields admits a parametrix. Indeed, regularization of the problem is achieved when suitable multifractional Sobolev norms are considered in the application of the numerical projection methods proposed (see Ruiz-Medina and Fernández-Pascual, 2010). Reciprocally, similar results can also be formulated in the case of inverse estimation of a multifractional input random field from the observation of an output random field, both related by a pseudodifferential equation.

The outline of the paper is as follows. Section A8.1 provides the ordinary formulation of the direct and inverse variable order least-squares linear estimation problems. The estimation system equation and the observation model in the multifractional generalized context are given in Section A8.2. A unique stable solution to the variable order least squares estimation problems is formulated under the regularity conditions assumed on the RKHS norms involved. In Section A8.3, the results derived are applied to the class of multifractional random fields with covariance operator defined by a continuous function of a self-adjoint and elliptic multifractional pseudo-differential operator. In Section A8.4, numerical projection methods respectively based on covariance eigenfunction system, and on dual Riesz bases, both diagonalizing the covariance structure, are applied in the finite-dimensional approximation of the least-squares linear direct and inverse estimation problems. For illustration purposes, a simulation study is developed in Section A8.6. Finally, conclusions are given in Section A8.7.

A8.1 Random Field Formulation

Let $\left\{\mathcal{Y}_{\beta(\cdot)}(\mathbf{z}) :\ \mathbf{z} \in S \subseteq \mathbb{R}^n\right\}$ be a zero-mean second-order random field on the probability space $(\Omega, \mathcal{A}, P)$, with RKHS included in or equal to a fractional Sobolev space $H^{\beta(\cdot)}(S)$ of variable order $\beta(\cdot)$, with $H^{\beta(\cdot)}(S)$ being constituted, as usual, by the weak-sense restrictions to set S of the functions in $H^{\beta(\cdot)}(\mathbb{R}^n)$ (see Kikuchi and Negoro, 1995, and Leopold, 1989, 1991, 1999, for its proper definition). Assume that $\mathcal{Y}_{\beta(\cdot)}$ defines the output of a linear system given in terms of a pseudodifferential operator of variable order A, with random input $\mathcal{S}_{\gamma(\cdot)}$ being a multifractional zero-mean second-order random field, that is,

$$\mathcal{Y}_{\beta(\mathbf{z})}(\mathbf{z}) := \mathrm{A}\mathcal{S}_{\gamma(\cdot)}(\mathbf{z}) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i\langle \mathbf{z}, \boldsymbol{\lambda} \rangle} p_{\mathrm{A}}(\mathbf{z}, \boldsymbol{\lambda}) \widehat{\mathcal{S}}_{\gamma(\cdot)}(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \tag{A8.1}$$

for all $\mathbf{z} \in S \subseteq \mathbb{R}^n$, where $\gamma(\cdot)$ denotes the variable order of the fractional Sobolev space related to the RKHS of the multifractional random signal $\mathcal{S}_{\gamma(\cdot)}$. Here, the integral is understood in the

mean-square sense, and $p_{\mathrm{A}} \in \mathcal{B}^{\infty}\left(\mathbb{R}^n_{\mathbf{z}} \times \mathbb{R}^n_{\boldsymbol{\lambda}}\right)$ denotes the symbol of pseudodifferential operator A, satisfying that, for any multi-indices α and β, there exists a positive constant $C_{\alpha,\beta}$ such that

$$|\mathrm{D}^{\alpha}_{\boldsymbol{\xi}}\mathrm{D}^{\beta}_{\mathbf{x}}p(\mathbf{x},\boldsymbol{\xi})| \leq C_{\alpha,\beta}\langle\boldsymbol{\xi}\rangle^{\sigma(\mathbf{x})-\rho|\alpha|+\delta|\beta|}, \tag{A8.2}$$

where $\mathrm{D}^{\alpha}_{\boldsymbol{\xi}}$ and $\mathrm{D}^{\beta}_{\mathbf{x}}$ respectively denote the derivatives with respect to $\boldsymbol{\xi}$ and $\mathbf{x}$, and $\langle\boldsymbol{\xi}\rangle = (1 + |\boldsymbol{\xi}|^2)^{1/2}$. That is, p_{A} belongs to $\mathcal{S}^{\sigma}_{\rho,\delta}$, with $0 \leq \delta < \rho \leq 1$, and σ being a real-valued function in $\mathcal{B}^{\infty}(\mathbb{R}^n)$. Here, $\mathcal{B}^{\infty}(\mathbb{R}^n_{\mathbf{z}} \times \mathbb{R}_{\boldsymbol{\lambda}})$ denotes the space of all C^{∞}-functions on $\mathbb{R}^n_{\mathbf{z}} \times \mathbb{R}_{\boldsymbol{\lambda}}$ (respectively on $\mathbb{R}^n$), whose derivatives of all orders are bounded (see Kikuchi and Negoro, 1995 and Leopold, 1989, 1991, 1999).

We address the problem of deriving the best functional linear estimator $\widehat{\mathcal{Y}}_{\beta(\cdot)} = \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\cdot)$ of $\mathcal{Y}_{\beta(\cdot)}$ in the closed subspace of $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ generated by the random variables

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{x}) := \mathcal{S}_{\gamma(\cdot)}(\mathbf{x}) + \mathcal{N}_{\theta(\cdot)}(\mathbf{x}), \quad \mathbf{x} \in S_{\mathcal{X}} \subseteq S \subseteq \mathbb{R}^n,$$

where $\mathcal{N}_{\theta(\cdot)}$ is multifractional additive noise, with RKHS included in or equal to a fractional Sobolev space of variable order $\theta(\cdot)$, mutually uncorrelated with $\mathcal{S}_{\gamma(\cdot)}$, and $S_{\mathcal{X}}$ denotes the observation domain. For simplicity, here, we consider that the supports of $\mathcal{S}_{\gamma(\cdot)}$ and $\mathcal{Y}_{\beta(\mathbf{z})}$ coincide. The space $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ is the Hilbert space of zero-mean second-order random variables with inner product

$$\langle \mathcal{X}, \mathcal{Y} \rangle_{\mathcal{L}^2(\Omega,\mathcal{A},P)} := E[\mathcal{X}\mathcal{Y}],$$

and the induced norm

$$\|\mathcal{X}\|^2_{\mathcal{L}^2(\Omega,\mathcal{A},P)} := E[\mathcal{X}]^2.$$

The set $S_{\mathcal{X}}$ can be bounded or unbounded, with C^{∞}- or fractal boundary.

In the Gaussian case, the best functional linear estimator $\widehat{\mathcal{Y}}_{\beta(\cdot)}(\cdot) = \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\cdot)$ of $\mathcal{Y}_{\beta(\cdot)}$ in the closed subspace of $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ generated by

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{x}), \quad \mathbf{x} \in S_{\mathcal{X}} \subseteq S \subseteq \mathbb{R}^n,$$

coincides with

$$\widetilde{\mathcal{Y}}_{\beta(\cdot)}(\cdot) = E\left[\mathcal{Y}_{\beta(\cdot)}(\cdot)|\mathcal{X}_{\alpha(\cdot)}(\mathbf{x}),\ \mathbf{x} \in S_{\mathcal{X}}\right].$$

Under the condition $\underline{\alpha} = \inf_{\mathbf{x}\in S_{\mathcal{X}}} \alpha(\mathbf{x}) > n/2$ (see embedding theorems between Besov spaces of variable order, for example, in Leopold, 1999), the auto-covariance operator

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}} := E\left[\mathcal{X}_{\alpha(\cdot)} \otimes \mathcal{X}_{\alpha(\cdot)}\right]$$

of $\mathcal{X}_{\alpha(\cdot)}$ admits an integral representation in the strong sense (pointwise). Here, for a Hilbert space H, $f \otimes g \in H \otimes H$, for $f, g \in H$, defines a Hilbert-Schmidt operator on H, with

$$[f \otimes g](v) = f\left\langle g, v\right\rangle_H, \quad \forall v \in H.$$

Thus, we can introduce the covariance kernel

$$B_{\mathcal{X}_{\alpha(\cdot)}}(\mathbf{x}, \mathbf{y}) = E[\mathcal{X}_{\alpha(\cdot)}(\mathbf{x})\mathcal{X}_{\alpha(\cdot)}(\mathbf{y})],$$

for $\mathbf{x}, \mathbf{y} \in S_{\mathcal{X}}$, as well as the cross-covariance function $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{Y}_{\beta(\cdot)}}(\mathbf{x}, \mathbf{z}) = E[\mathcal{X}_{\alpha(\cdot)}(\mathbf{x})\mathcal{Y}_{\beta(\cdot)}(\mathbf{z})]$, for $\mathbf{x} \in S_{\mathcal{X}}$ and $\mathbf{z} \in S$, which are assumed to be known, or approximated from the spectral decomposition of the empirical auto-covariance and cross-covariance operators defined a follows:

$$\widehat{\mathrm{R}}_{\mathcal{X}_{\alpha(\cdot)}} := \frac{1}{N}\sum_{i=1}^{N} \mathcal{X}^i_{\alpha(\cdot)} \otimes \mathcal{X}^i_{\alpha(\cdot)},\ \widehat{\mathrm{R}}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{Y}_{\beta(\cdot)}} := \frac{1}{N}\sum_{i=1}^{N} \mathcal{X}_{\alpha(\cdot)} \otimes \mathcal{Y}_{\beta(\cdot)}, \tag{A8.3}$$

for N independent detrended realizations $\mathcal{X}^i_{\alpha(\cdot)}$ of $\mathcal{X}_{\alpha(\cdot)}$ over set $S_{\mathcal{X}}$, respectively for $\mathcal{Y}_{\beta(\cdot)}$ over S.

From the Orthogonal Projection Theorem, for each $\mathbf{z} \in S$, the least-squares linear estimate $\mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\mathbf{z})$ of $\mathcal{Y}_{\beta(\cdot)}(\mathbf{z})$ based on the information provided by $\mathcal{X}_{\alpha(\cdot)}$ in $(A8.1)$ is defined as

$$\widehat{\mathcal{Y}}_{\beta(\mathbf{z})}(\mathbf{z}) = \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) := (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i\langle \mathbf{z}, \boldsymbol{\xi}\rangle} p_{\mathrm{K}}(\mathbf{z}, \boldsymbol{\xi})\widehat{\mathcal{X}}_{\alpha(\cdot)}(\boldsymbol{\xi})d\boldsymbol{\xi}, \tag{A8.4}$$

with p_{K} being the symbol of the pseudodifferential operator K, satisfying, for each $\mathbf{z} \in S$,

$$\begin{aligned} & E\left[\left(\mathcal{Y}_{\beta(\cdot)}(\mathbf{z}) - \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\mathbf{z})\right)\mathcal{X}_{\alpha(\cdot)}(\mathbf{x})\right] = 0 \\ \Longleftrightarrow \quad & B_{\mathcal{Y}_{\beta(\cdot)}\mathcal{X}_{\alpha(\cdot)}}(\mathbf{z},\mathbf{x}) = \mathrm{K}B_{\mathcal{X}_{\alpha(\cdot)}}(\mathbf{z},\mathbf{x}) \\ & = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i\langle \mathbf{z}-\mathbf{x},\boldsymbol{\xi}\rangle} p_{\mathrm{K}}(\mathbf{z}-\mathbf{x},\boldsymbol{\xi}) f_{\mathcal{X}_{\alpha(\cdot)}}(\boldsymbol{\xi})d\boldsymbol{\xi}, \quad \forall \mathbf{x} \in S_{\mathcal{X}}. \end{aligned}$$

Here, $\mathcal{X}_{\alpha(\cdot)}$ is assumed to admit the integral variable order representation

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle} g_{\mathcal{X}_{\alpha(\cdot)}}(\mathbf{z},\boldsymbol{\xi})\hat{\varepsilon}(\boldsymbol{\xi})d\boldsymbol{\xi},$$

in terms of the symbol $g_{\mathcal{X}_{\alpha(\cdot)}}$ factorizing the symbol $f_{\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}} = g_{\mathcal{X}_{\alpha(\cdot)}}\overline{g_{\mathcal{X}_{\alpha(\cdot)}}}$ associated with the autocovariance operator $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ of $\mathcal{X}_{\alpha(\cdot)}$, where $\widehat{\varepsilon}(\boldsymbol{\xi})d\boldsymbol{\xi}$ defines a complex white noise random measure satisfying

$$E[\widehat{\varepsilon}(\boldsymbol{\xi})] = 0, \quad E[\widehat{\varepsilon}(\boldsymbol{\xi})\overline{\widehat{\varepsilon}(\boldsymbol{\omega})}] = \sigma^2\delta(\boldsymbol{\xi}-\boldsymbol{\omega}),$$

with δ denoting the Dirac delta distribution. For simplicity, we take $\sigma^2 = 1$. In particular, $E|\widehat{\varepsilon}(\boldsymbol{\xi})d\boldsymbol{\xi}|^2 = \sigma^2 d\boldsymbol{\xi}$.

We now consider the case where our interest lies in the computation of the best functional linear predictor of $\mathcal{S}_{\gamma(\cdot)}$, in the closed subspace of $\mathcal{L}^2(\Omega,\mathcal{A},P)$ generated by the random variables

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{x}) = \mathcal{Y}_{\beta(\cdot)}(\mathbf{x}) + \mathcal{N}_{\theta(\cdot)}(\mathbf{x}), \quad \mathbf{x} \in S_{\mathcal{X}} \subseteq S.$$

Again, the Orthogonal Projection Theorem leads to the system of estimation equations given for each $\mathbf{z} \in S$, by

$$\begin{aligned} & E\left[\left(\mathcal{S}_{\gamma(\cdot)}(\mathbf{z}) - \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\mathbf{z})\right)\mathcal{X}_{\alpha(\cdot)}(\mathbf{x})\right] = 0 \\ \Longleftrightarrow \quad & B_{\mathcal{S}_{\gamma(\cdot)}\mathcal{X}_{\alpha(\cdot)}}(\mathbf{z},\mathbf{x}) = \mathrm{K}B_{\mathcal{X}_{\alpha(\cdot)}}(\mathbf{z},\mathbf{x}) \\ & = (2\pi)^{-n}\int_{\mathbb{R}^n} e^{i\langle \mathbf{z}-\mathbf{x},\boldsymbol{\xi}\rangle} p_{\mathrm{K}}(\mathbf{z}-\mathbf{x},\boldsymbol{\xi}) f_{\mathcal{X}_{\alpha(\cdot)}}(\boldsymbol{\xi})d\boldsymbol{\xi}, \quad \forall \mathbf{x} \in S_{\mathcal{X}}. \end{aligned}$$

As before, p_{K} denotes the symbol defining the pseudodifferential operator K, which, in this case, provides the least-squares functional linear estimator of $\mathcal{S}_{\gamma(\cdot)}$ as follows:

$$\widehat{\mathcal{S}}_{\gamma(\mathbf{z})}(\mathbf{z}) = \mathrm{K}\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) := (2\pi)^{-n} \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle} p_{\mathrm{K}}(\mathbf{z},\boldsymbol{\xi}) \widehat{\mathcal{X}}_{\alpha(\cdot)}(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{A8.5}$$

In the above direct and inverse functional linear estimation problems, the pseudodifferential operator of variable order K is not usually continuous or bounded with respect to the $L^2(S_{\mathcal{X}})$-topology. Therefore, suitable fractional Sobolev spaces of variable order are considered, whose norms allow the definition of a stable (continuous) solution.

A8.2 Formulation in a Hilbert-Valued Context

In the following, we identify the random fields involved in the functional linear estimation problems with Hilbert-valued random variables, i.e., functional random variables having values in a suitable separable Hilbert space H. Indeed, in each case, H will coincide with the fractional Sobolev space of variable order, isomorphically identified with the RKHS of the random field considered. In the Gaussian case, the modulus of continuity of the trajectories of random fields $\mathcal{Y}_{\beta(\cdot)}$, $\mathcal{X}_{\alpha(\cdot)}$, $\mathcal{S}_{\gamma(\cdot)}$, and $\mathcal{N}_{\theta(\cdot)}$ is given in terms of the variable local Hölder exponent of the functions in their RKHSs (see, for example, Adler 1981). Assume that for each $\omega \in \Omega$, $\mathcal{Y}_{\beta(\cdot)}(\omega,\cdot) \in H^{\beta(\cdot)}(S)$, $\mathcal{X}_{\alpha(\cdot)}(\omega,\cdot) \in H^{\alpha(\cdot)}(S_{\mathcal{X}})$, $\mathcal{S}_{\gamma(\cdot)}(\omega,\cdot) \in H^{\gamma(\cdot)}(S)$, and $\mathcal{N}_{\theta(\cdot)}(\omega,\cdot) \in H^{\theta(\cdot)}(S_{\mathcal{X}})$ (see Bosq, 2000, Chapter 1).

From now on, we then consider the infinite-dimensional formulation of the two functional linear estimation problems, in terms of the projection of the functional random variables $\mathcal{Y}_{\beta(\cdot)}$, $\mathcal{X}_{\alpha(\cdot)}$, $\mathcal{S}_{\gamma(\cdot)}$ and $\mathcal{N}_{\theta(\cdot)}$ into suitable bases of the spaces $H^{\beta(\cdot)}(S)$, $H^{\alpha(\cdot)}(S_{\mathcal{X}})$, $H^{\gamma(\cdot)}(S)$, and $H^{\theta(\cdot)}(S_{\mathcal{X}})$. That is, we consider the generalized random fields involved

$$\left\{\mathcal{Y}_{\beta(\cdot)}(\phi),\ \phi \in H^{\beta(\cdot)}(S)\right\}, \quad \left\{\mathcal{X}_{\alpha(\cdot)}(\varphi),\ \varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}})\right\},$$

$$\left\{\mathcal{S}_{\gamma(\cdot)}(\psi),\ \psi \in H^{\gamma(\cdot)}(S)\right\}, \quad \left\{\mathcal{N}_{\theta(\cdot)}(\chi),\ \chi \in H^{\theta(\cdot)}(S_{\mathcal{X}})\right\},$$

respectively generated by the infinite-dimensional random variables

$$\left\{\mathcal{Y}_{\beta(\cdot)}(\phi_k)=\left\langle \mathcal{Y}_{\beta(\cdot)},\phi_k\right\rangle_{H^{\beta(\cdot)}(S)},k\in\mathbb{N}\right\},\ \left\{\mathcal{X}_{\alpha(\cdot)}(\varphi_k)=\left\langle \mathcal{X}_{\alpha(\cdot)},\varphi_k\right\rangle_{H^{\alpha(\cdot)}(S_{\mathcal{X}})},k\in\mathbb{N}\right\}$$

and

$$\left\{\mathcal{S}_{\gamma(\cdot)}(\psi_k)=\left\langle \mathcal{S}_{\gamma(\cdot)},\psi_k\right\rangle_{H^{\gamma(\cdot)}(S)},k\in\mathbb{N}\right\},\ \left\{\mathcal{N}_{\theta(\cdot)}(\chi_k)=\left\langle \mathcal{N}_{\theta(\cdot)},\chi_k\right\rangle_{H^{\theta(\cdot)}(S_{\mathcal{X}})},k\in\mathbb{N}\right\},$$

for suitable orthonormal bases $\{\phi_k,k\in\mathbb{N}\}$, $\{\varphi_k,k\in\mathbb{N}\}$, $\{\psi_k,k\in\mathbb{N}\}$, and $\{\chi_k,k\in\mathbb{N}\}$.

In the subsequent development, we establish the weak-sense formulation of the best functional linear predictor of $\mathcal{Y}_{\beta(\cdot)}$ given $\mathcal{X}_{\alpha(\cdot)}=\mathcal{S}_{\gamma(\cdot)}+\mathcal{N}_{\theta(\cdot)}$ over set $S_{\mathcal{X}}$, as well as of $\mathcal{S}_{\gamma(\cdot)}$ given $\mathcal{X}_{\alpha(\cdot)}=\mathcal{Y}_{\beta(\cdot)}+\mathcal{N}_{\theta(\cdot)}$ over set $S_{\mathcal{X}}$. The main result of this section, Theorem 11, provides the solution with minimum multifractional singularity order to the above estimation problems.

For simplicity, let us consider the case $S=\mathbb{R}^n$. (The case where $S\neq\mathbb{R}^n$ can be similarly treated by considering functions of H^∞ having support with suitable geometrical characteristics, with H^∞ denoting the intersection of all fractional Sobolev spaces of variable order). Let $\mathcal{S}$ and $\mathcal{Y}$ be defined as in (A8.1). Assume that $\mathcal{S}$ and $\mathcal{Y}$ have respective multifractional singularity orders $-\gamma(\cdot),\ -\beta(\cdot)\in\mathcal{B}^\infty(\mathbb{R}^n)$. The following weak-sense formulation of $(A8.1)$ is then considered:

$$\mathcal{Y}_{\beta(\cdot)}(\phi)=\mathcal{S}_{\gamma(\cdot)}(\mathrm{A}'\phi),\quad\forall\phi\in H^{\beta(\cdot)}(\mathbb{R}^n),$$

where A' represents the adjoint operator of A, defined from $H^{\beta(\cdot)}(\mathbb{R}^n)$ into $H^{\gamma(\cdot)}(\mathbb{R}^n)$. We consider the problem of computing the least-squares linear estimator of $\mathcal{Y}_{\beta(\cdot)}$ (respectively of $\mathcal{S}_{\gamma(\cdot)}$, in the inverse formulation) from the information provided by the multifractional generalized version of $(A8.1)$ given by

$$\mathcal{X}_{\alpha(\cdot)}(\varphi)=S_{\gamma(\cdot)}(\varphi)+\mathcal{N}_{\theta(\cdot)}(\varphi),\quad\forall\varphi\in H^{\alpha(\cdot)}(S_{\mathcal{X}})$$

(respectively,

$$\mathcal{X}_{\alpha(\cdot)}(\varphi)=\mathcal{Y}_{\beta(\cdot)}(\varphi)+\mathcal{N}_{\theta(\cdot)}(\varphi),\quad\varphi\in H^{\alpha(\cdot)}(S_{\mathcal{X}}),$$

in the inverse formulation). Here, the observation noise $\mathcal{N}_{\theta(\cdot)}$ is a generalized random field with minimum multifractional singularity order $-\theta(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$, and $-\alpha(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$ represents the minimum multifractional singularity order of the observation random field $\mathcal{X}$ defining $\mathcal{X}_{\alpha(\cdot)}$ on $H^{\alpha(\cdot)}(\mathbb{R}^n)$.

We assume that the covariance function $B_{\mathcal{X}_{\alpha(\cdot)}}$ of $\mathcal{X}_{\alpha(\cdot)}$ and the cross-covariance function $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{Y}_{\beta(\cdot)}}$ between $\mathcal{X}_{\alpha(\cdot)}$ and $\mathcal{Y}_{\beta(\cdot)}$ (respectively, the cross-covariance function $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}$ between $\mathcal{X}_{\alpha(\cdot)}$ and $\mathcal{S}_{\gamma(\cdot)}$, in the inverse formulation) are known, or approximated from the spectral decomposition of the empirical covariance operators in (A8.3), by projection into the corresponding empirical eigenfunction bases. The least-squares linear estimator

$$\widehat{\mathcal{Y}}_{\beta(\cdot)}(\phi) := \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi), \quad \forall \phi \in H^{\beta(\cdot)}(\mathbb{R}^n), \tag{A8.6}$$

of $\mathcal{Y}_{\beta(\cdot)}$ (respectively,

$$\widehat{\mathcal{S}}_{\gamma(\cdot)}(\phi) := \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi), \quad \forall \phi \in H^{\gamma(\cdot)}(\mathbb{R}^n)$$

of $\mathcal{S}_{\gamma(\cdot)}$) is then obtained by minimizing the mean-square error

$$\mathrm{MSE}(\mathrm{K}'\phi) := \mathcal{E}(\mathrm{K}'\phi) = E[\mathcal{Y}_{\beta(\cdot)}(\phi) - \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi)]^2, \tag{A8.7}$$

for all $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$, where $\mathrm{K}' : H^{\beta(\cdot)}(\mathbb{R}^n) \rightarrow H^{\alpha(\cdot)}(\mathbb{R}^n)$ is the adjoint of the operator K defined by (A8.4) (respectively defined by (A8.5), for the inverse estimation problem). Hence, a continuous definition of operator K providing the solution to the least-squares linear estimation problem can be found in the space $H^{\beta(\cdot)}(\mathbb{R}^n) \otimes H^{\alpha(\cdot)}(\mathbb{R}^n)$ (respectively in the space $H^{\gamma(\cdot)}(\mathbb{R}^n) \otimes H^{\alpha(\cdot)}(\mathbb{R}^n)$, in the inverse formulation). If A is a pseudodifferential operator, then, for all $\mathbf{x} \in \mathbb{R}^n$, $\beta(\mathbf{x}) \geq \gamma(\mathbf{x})$, and if A is the parametrix of a pseudodifferental operator, then $\gamma(\mathbf{x}) \geq \beta(\mathbf{x})$, for every point $\mathbf{x} \in \mathbb{R}^n$. Note that this order is well-defined, since the variable orders β and γ are real-valued functions with multidimensional support (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991).

The covariance functions $B_{\mathcal{X}_{\alpha(\cdot)}}$ of $\mathcal{X}_{\alpha(\cdot)}$ and $B_{\mathcal{S}_{\gamma(\cdot)}}$ of $\mathcal{S}_{\gamma(\cdot)}$, and the cross-covariance functions $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{Y}_{\beta(\cdot)}}$ between $\mathcal{X}_{\alpha(\cdot)}$ and $\mathcal{Y}_{\beta(\cdot)}$, and $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}$ between $\mathcal{X}_{\alpha(\cdot)}$ and $\mathcal{S}_{\gamma(\cdot)}$, are respectively given by the following expressions:

$$
\begin{aligned}
B_{\mathcal{X}_{\alpha(\cdot)}}(\varphi,\psi) &= \left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\varphi)]^*, \psi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}, \quad \forall \varphi,\psi \in H^{\alpha(\cdot)}(\mathbb{R}^n), \\
B_{\mathcal{S}_{\gamma(\cdot)}}(\varphi,\psi) &= \left\langle [\mathrm{R}_{\mathcal{S}_{\gamma(\cdot)}}(\varphi)]^*, \psi \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)}, \quad \forall \varphi,\psi \in H^{\gamma(\cdot)}(\mathbb{R}^n), \\
B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{Y}_{\beta(\cdot)}}(\varphi,\phi) &= E[\mathcal{X}_{\alpha(\cdot)}(\varphi)\mathcal{S}_{\gamma(\cdot)}(\mathrm{A}'\phi)] = B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\varphi,\mathrm{A}'\phi) \\
&= \left\langle [\mathrm{R}'_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\varphi)]^*, \mathrm{A}'\phi \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)} \\
&= \left\langle \varphi, [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi)]^* \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}, \\
&\quad \forall \varphi \in H^{\alpha(\cdot)}(\mathbb{R}^n),\ \phi \in H^{\beta(\cdot)}(\mathbb{R}^n), \\
B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\varphi,\phi) &= E[\mathcal{X}_{\alpha(\cdot)}(\varphi)\mathcal{S}_{\gamma(\cdot)}(\phi)] = B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\varphi,\phi) \\
&= \left\langle [\mathrm{R}'_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\varphi)]^*, \phi \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)} \\
&= \left\langle \varphi, [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\phi)]^* \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}, \\
&\quad \forall \varphi \in H^{\alpha(\cdot)}(\mathbb{R}^n),\ \forall \phi \in H^{\gamma(\cdot)}(\mathbb{R}^n),
\end{aligned}
$$

where $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}} : H^{\alpha(\cdot)}(\mathbb{R}^n) \longrightarrow H^{-\alpha(\cdot)}(\mathbb{R}^n)$, $\mathrm{R}_{\mathcal{S}_{\gamma(\cdot)}} : H^{\gamma(\cdot)}(\mathbb{R}^n) \longrightarrow H^{-\gamma(\cdot)}(\mathbb{R}^n)$, and $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}} : H^{\gamma(\cdot)}(\mathbb{R}^n) \longrightarrow H^{-\alpha(\cdot)}(\mathbb{R}^n)$ denote the covariance operators associated, respectively, with the generalized covariance functions $B_{\mathcal{X}_{\alpha}(\cdot)}$, $B_{\mathcal{S}_{\gamma}(\cdot)}$ and $B_{\mathcal{X}_{\alpha}(\cdot)\mathcal{S}_{\gamma(\cdot)}}$.

The mean-square error associated with K', for each $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$, is computed from ($A8.7$),

being given by

$$\begin{aligned}\mathcal{E}(\mathrm{K}'\phi) &= E[\mathcal{Y}_{\beta(\cdot)}(\phi) - \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi)]^2 = B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi, \mathrm{K}'\phi) \\ &+ B_{\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi, \mathrm{A}'\phi) - 2B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{K}'\phi, \mathrm{A}'\phi) \\ &= \left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi)]^*, \mathrm{K}'\phi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)} + \left\langle [\mathrm{R}_{\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi)]^*, \mathrm{A}'\phi \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)} \\ &- 2\left\langle \mathrm{K}'\phi, [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}\mathrm{A}'\phi]^* \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}\end{aligned}$$

(respectively

$$\begin{aligned}\mathcal{E}(\mathrm{K}'\phi) &= E[\mathcal{S}_{\gamma(\cdot)}(\phi) - \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi)]^2 = B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi, \mathrm{K}'\phi) \\ &+ B_{\mathcal{S}_{\gamma(\cdot)}}(\phi, \phi) - 2B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{K}'\phi, \phi) \\ &= \left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi)]^*, \mathrm{K}'\phi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)} + \left\langle [\mathrm{R}_{\mathcal{S}_{\gamma(\cdot)}}(\phi)]^*, (\phi) \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)} \\ &- 2\left\langle \mathrm{K}'\phi, [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\phi)]^* \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)},\end{aligned}$$

in the inverse formulation).

In the direct formulation, from the Orthogonal Projection Theorem, K′ minimizing the MSE must satisfy, for $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$,

$$E\left[\left(\mathcal{Y}_{\beta(\cdot)}(\phi) - \mathcal{X}_{\alpha(\cdot)}(\mathrm{K}'\phi)\right)\mathcal{X}_{\alpha(\cdot)}(\varphi)\right] = 0, \quad \forall \varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}}).$$

Then,

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}\left(\mathrm{A}'\phi\right)(\varphi) = \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi)(\varphi), \quad \forall \varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}}).$$

That is,

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}\left(\mathrm{A}'\phi\right) \underset{H^{\alpha(\cdot)}(S_{\mathcal{X}})}{=} \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi), \tag{A8.8}$$

for each $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$. Assuming that condition $(A8.8)$ holds, the partial Fréchet derivative $\partial\mathcal{E}(\mathrm{K}'\phi + \xi\mathrm{L})/\partial\xi$, with ξ being a small real number and L a linear operator from $H^{\beta(\cdot)}(\mathbb{R}^n)$ into

$H^{\alpha(\cdot)}(\mathbb{R}^n)$, is equal to zero for $\xi = 0$. That is,

$$\begin{aligned}\mathcal{E}((\mathrm{K}' + \xi \mathrm{L})(\phi)) &= B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi, \mathrm{K}'\phi) + 2\xi B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{L}\phi, \mathrm{K}'\phi) \\ &+ \xi^2 B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{L}\phi, \mathrm{L}\phi) \\ &+ B_{\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi, \mathrm{A}'\phi) - 2B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{K}'\phi, \mathrm{A}'\phi) \\ &- 2\xi B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{L}\phi, \mathrm{A}'\phi).\end{aligned}$$

Then,

$$\begin{aligned}\frac{\partial \varepsilon(\mathrm{K}' + \xi \mathrm{L})}{\partial \xi}(\phi) &= 2\left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi)]^*, \mathrm{L}\phi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)} \\ &- 2\left\langle \mathrm{L}\phi, [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi)]^* \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)} \\ &+ 2\xi\left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{L}\phi)]^*, \mathrm{L}\phi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}, \quad \forall \phi \in H^{\beta(\cdot)}(\mathbb{R}^n),\end{aligned} \tag{A8.9}$$

and (A8.9) is null for $\xi = 0$. The least-squares error is then given by

$$\mathrm{LSE}(\mathrm{K}'\phi) = \left\langle [\mathrm{R}_{\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi)]^*, \mathrm{A}'\phi \right\rangle_{H^{\gamma(\cdot)}(\mathbb{R}^n)} - \left\langle [\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi)]^*, \mathrm{K}'\phi \right\rangle_{H^{\alpha(\cdot)}(\mathbb{R}^n)}.$$

Since the dual RKHS of $\mathcal{X}_{\alpha(\cdot)}$ coincides, as a set of functions, with $H^{\alpha(\cdot)}(\mathbb{R}^n)$, with the multifractional geometry generated by its covariance function, we have $B_{\mathcal{X}_{\alpha(\cdot)}}(\mathrm{K}'\phi, \mathrm{K}'\phi) = \|\mathrm{K}'\phi)\|^2_{\mathcal{H}(\tilde{\mathcal{X}}_{\alpha(\cdot)})} < \infty$, for all $\phi \in H^{\beta(\cdot)}(\mathbb{R}^n)$.

From the Orthogonal Projection Theorem, a similar expression can be computed for the least-squares error in the definition of the optimal K in the inverse formulation. In the subsequent development we consider, for simplicity, the direct formulation. Similar assertions hold for the inverse formulation by considering the corresponding equations previously obtained.

The following result shows that under the existence of the pseudodual of variable order $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ of $\mathcal{X}_{\alpha(\cdot)}$, with $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ being defined on $\bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}}) = [H^{\alpha(\cdot)}(S_{\mathcal{X}})]^*$, ($A$8.8) admits a unique

continuous solution in the RKHS of $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$, which solves the estimation problem, and at least it can be defined, in the weak-sense, in terms of suitable test functions, hence, by projection into suitable orthonormal bases. Here, $\bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}})$ denotes the space of tempered distributions of $H^{-\alpha(\cdot)}(\mathbb{R}^n)$ with compact support contained in $S_{\mathcal{X}}$, and $[H]^*$ represents the dual Hilbert space of H.

Theorem 11. *Let $\mathcal{X}_{\alpha(\cdot)}$ be the generalized random field given in terms of $\mathcal{X}$ with generalized covariance function $B_{\mathcal{X}_{\alpha(\cdot)}}(\cdot,\cdot)$, and assume the existence of the pseudodual $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ of $\mathcal{X}_{\alpha(\cdot)}$. Then, the direct estimation problem admits a unique stable solution belonging to the dual RKHS of $\mathcal{X}_{\alpha(\cdot)}$ defined by*

$$\mathrm{K}'\phi \underset{H^{-\alpha(\cdot)}(\mathbb{R}^n)}{:=} \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}} \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}(\mathrm{A}'\phi), \quad \forall \phi \in H^{\beta(\cdot)}(\mathbb{R}^n), \tag{A8.10}$$

where $\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}} : H^{-\alpha(\cdot)}(\mathbb{R}^n) \longrightarrow H^{\alpha(\cdot)}(\mathbb{R}^n)$ is the covariance operator of $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$, and $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}} : H^{\gamma(\cdot)}(\mathbb{R}^n) \to H^{-\alpha(\cdot)}(\mathbb{R}^n)$ is the covariance operator associated with the cross-covariance function $B_{\mathcal{X}_{\alpha(\cdot)}\mathcal{S}_{\gamma(\cdot)}}$ between $\mathcal{X}_{\alpha(\cdot)}$ and $\mathcal{S}_{\gamma(\cdot)}$. Moreover,

$$\mathrm{K}\varphi \underset{H^{\beta(\cdot)}(\mathbb{R}^n)}{=} \mathrm{AR}_{\mathcal{S}_{\gamma(\cdot)}\mathcal{X}_{\alpha(\cdot)}} \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}(\varphi), \quad \forall \varphi \in H^{\alpha(\cdot)}(\mathbb{R}^n),$$

which provides a weak-sense multifractional solution to the corresponding ordinary estimation problem.

Proof Under the existence of $\widetilde{\mathcal{X}}_{\alpha(\cdot)} : \bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}}) \longrightarrow \mathcal{L}^2(\Omega, \mathcal{A}, P)$ bounded or continuous,

$$E\left[\widetilde{\mathcal{X}}_{\alpha(\cdot)}(\phi^*)\mathcal{X}_{\alpha(\cdot)}(\psi)\right] = (\mathrm{I}+\mathrm{R})(\psi)(\phi), \quad \phi, \psi \in H^{\alpha(\cdot)}(S_{\mathcal{X}}), \tag{A8.11}$$

where R denotes the parametrix of operator $\mathrm{T}_{\mathcal{X}_{\alpha(\cdot)}}$ satisfying (see, for example, Vakhania, Tarieladze and Chebonyan, 1987)

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}} = \mathrm{T}_{\mathcal{X}_{\alpha(\cdot)}} \mathrm{T}^*_{\mathcal{X}_{\alpha(\cdot)}}, \tag{A8.12}$$

and I denotes the identity operator. Therefore, the RKHS $\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$ of $\mathcal{X}_{\alpha(\cdot)}$ can be isomorphically identified with the space $H^{\alpha(\cdot)}(S_{\mathcal{X}})$, since, for $f, g \in \mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$, its inner product, given by

$$\langle f, g \rangle_{\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})} = \left\langle \mathrm{T}^{-1}_{\mathcal{X}_{\alpha(\cdot)}}(f), \mathrm{T}^{-1}_{\mathcal{X}_{\alpha(\cdot)}}(g) \right\rangle_{L^2(S_{\mathcal{X}})},$$

defines a norm which is equivalent to the one generated by the inner product in the space $H^{\alpha(\cdot)}(S_{\mathcal{X}})$.

Thus, the covariance operator $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ admits a continuous inverse

$$\mathrm{R}^{-1}_{\mathcal{X}_{\alpha(\cdot)}} : H^{\alpha(\cdot)}(S_{\mathcal{X}}) \longrightarrow \bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}}).$$

The assertion of this result then follows from the definition of $\widehat{\mathcal{Y}}_{\beta(\cdot)}$ in (A8.6), and from (A8.10) obtained from (A8.8), since $\mathrm{R}^{-1}_{\mathcal{X}_{\alpha(\cdot)}} \simeq \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}$ (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991). □

The following result provides a sufficient condition that ensures the existence of the pseudo-dual random field $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ of the observation random field $\mathcal{X}_{\alpha(\cdot)}$.

Theorem 12. *Let $\mathcal{S}_{\gamma(\cdot)}$ be the generalized input random multifractional signal. Assume that $\mathcal{X}_{\alpha(\cdot)}$ has minimum functional singularity order $\alpha(\cdot)$ which coincides with the duality order of $\mathcal{S}_{\gamma(\cdot)}$. Then, in the case where the covariance operator $\mathrm{R}_{\mathcal{N}_{\theta(\cdot)}}$ of $\mathcal{N}_{\theta(\cdot)}$ satisfies*

$$\|\mathrm{R}_{\mathcal{N}_{\theta(\cdot)}}\| < \|\mathrm{R}_{\widetilde{\mathcal{S}}_{\gamma(\cdot)}}\|^{-1},$$

the dual $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ of $\mathcal{X}_{\alpha(\cdot)}$ exists. Furthermore, the following inequality holds:

$$\|\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}\| \leq \left[\|\mathrm{R}_{\widetilde{\mathcal{S}}_{\gamma(\cdot)}}\|^{-1} - \|\mathrm{R}_{\mathcal{N}_{\theta(\cdot)}}\|\right]^{-1}.$$

The proof of this result follows directly from the perturbation theory for bounded linear operators in Banach and, in particular, Hilbert spaces (see Kato, 1995).

Remark 24. *Similar results hold for the best functional linear predictor of $\mathcal{S}_{\gamma(\cdot)}$ in the closed subspace of $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ generated by $\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) = \mathcal{Y}_{\beta(\cdot)}(\mathbf{z}) + \mathcal{N}_{\theta(\cdot)}(\mathbf{z})$,*

$\mathbf{z} \in S_{\mathcal{X}}$, *since under the existence of the continuous pseudodual* $\widetilde{\mathcal{X}}_{\alpha(\cdot)} : \bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}}) \longrightarrow \mathcal{L}^2(\Omega, \mathcal{A}, P)$, *the best functional linear estimator* $\widehat{\mathcal{S}}_{\gamma(\cdot)}$ *admits a continuous definition with respect to the norms in the spaces* $H^{\alpha(\cdot)}(S_{\mathcal{X}})$, $\bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}})$, *and* $H^{\gamma(\cdot)}(S)$ *involved in its formulation in a generalized random field framework.*

In particular, the following reformulation of Theorem 12 can be considered when the observation model is given by

$$\mathcal{X}_{\alpha(\cdot)} = \mathcal{Y}_{\beta(\cdot)} + \mathcal{N}_{\theta(\cdot)}.$$

Theorem 13. *Let* $\mathcal{Y}_{\beta(\cdot)}$ *be the generalized output random field. Assume that* $\mathcal{X}_{\alpha(\cdot)}$ *has minimum functional singularity order* $\alpha(\cdot)$, *which coincides with the duality order of* $\mathcal{Y}_{\beta(\cdot)}$. *Then, in the case where the covariance operator* $\mathrm{R}_{\mathcal{N}_{\theta(\cdot)}}$ *of* $\mathcal{N}_{\theta(\cdot)}$ *satisfies*

$$\|\mathrm{R}_{\mathcal{N}_{\theta(\cdot)}}\| < \|\mathrm{R}_{\widetilde{\mathcal{Y}}_{\beta(\cdot)}}\|^{-1},$$

the dual $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$ *of* $\mathcal{X}_{\alpha(\cdot)}$ *exists. Furthermore, the following inequality holds:*

$$\|\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}\| \leq \left[\|\mathrm{R}_{\widetilde{\mathcal{Y}}_{\beta(\cdot)}}\|^{-1} - \|R_{\mathcal{N}_{\theta(\cdot)}}\|\right]^{-1}.$$

Remark 25. *Theorems 12 and 13 admit similar formulation under the existence of the pseudodual of the generalized observation noise* $\mathcal{N}_{\theta(\cdot)}$.

A8.3 Heterogeneous Rational Covariance Spectra

In the previous section, the duality order $\alpha(\cdot)$ of the observation random field $\mathcal{X}_{\alpha(\cdot)}$ provides the functional order of the multifractional Sobolev spaces where its covariance operator admits a continuous inverse. Hence, this order establishes the Hilbert space norm that must be considered to compute K, or its adjoint K′, in a stable way. The spectral formulation of the conditions established before for a continuous inversion of the estimation problem will be derived through this section.

From now on, we will denote

$$\underline{\alpha} = \inf_{\mathbf{x}\in S_{\mathcal{X}}} \{\alpha(\mathbf{x})\}, \quad \overline{\alpha} = \sup_{\mathbf{x}\in S_{\mathcal{X}}} \{\alpha(\mathbf{x})\}.$$

Proposition 5. *Assume that the conditions of Theorem* 11 hold. Then, the following norms are equivalent: For $\varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}})$,

$$\|\varphi\|_{p_{\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}}} \simeq \|\varphi\|_{\langle\boldsymbol{\xi}\rangle^{\alpha(\cdot)}} \simeq \|\varphi\|_{H^{\alpha(\cdot)}(S_{\mathcal{X}})} \simeq \|\varphi\|_{\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})},$$

where $\|\varphi\|_{H^{\alpha(\cdot)}(S_{\mathcal{X}})}$ denotes the norm of φ in the fractional Sobolev space of variable order $\alpha(\cdot)$ (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991, for its proper definition), and

$$\begin{aligned}
\|\varphi\|_{p_{\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}}} &= \int_{S_{\mathcal{X}}} \left[\int_{\mathbb{R}^n} e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle} p_{\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}}(\mathbf{z},\boldsymbol{\xi})\widehat{\varphi}(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z} \\
\|\varphi\|_{\langle\boldsymbol{\xi}\rangle^{\alpha(\cdot)}} &= \int_{S_{\mathcal{X}}} \left[\int_{\mathbb{R}^n} e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{\alpha(\mathbf{z})}\widehat{\varphi}(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z} \\
\|\varphi\|_{\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})} &= \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}(\varphi)(\varphi).
\end{aligned}$$

Proof From Theorem 11, $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ is elliptic. Furthermore, $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ is a self-adjoint pseudodifferential operator on $L^2(S_{\mathcal{X}})$ with symbol $p_{\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}} \in \mathcal{S}^{-\alpha}_{\rho,\delta}$ (respectively, $p_{\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}} \in \mathcal{S}^{\alpha}_{\rho,\delta}$). Thus, $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ belongs to the space $\boldsymbol{\mathcal{S}}^{-\alpha}_{\rho,\delta}$ (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991). From the general theory of elliptic pseudodifferential operators of variable order (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991), the following inequalities then hold: There exist positive constants $C_1 \leq C_2$ such that

$$\begin{aligned}
&C_1 \int_{S_{\mathcal{X}}} \left[\int_{\mathbb{R}^n} e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{\alpha(\mathbf{z})}\widehat{\varphi}(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z} \\
&\leq \int_{S_{\mathcal{X}}} \left[\int_{\mathbb{R}^n} e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle} p_{\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}}(\mathbf{z},\boldsymbol{\xi})\widehat{\varphi}(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z} \\
&= \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}(\varphi)(\varphi) = \|\varphi\|^2_{\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})} \\
&\leq C_2 \int_{S_{\mathcal{X}}} \left[\int_{\mathbb{R}^n} e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{\alpha(\mathbf{z})}\widehat{\varphi}(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z}, \quad \forall\varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}}).
\end{aligned}$$

Hence, the result follows. □

Corollary 3. *Under the conditions assumed in Proposition* 5*, the following assertions hold:*

(i) $\phi \in \bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}})$ *iff* $\left\langle \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}\phi, \phi \right\rangle_{L^2(S_{\mathcal{X}})} < \infty$, *that is, iff* $\varphi \in \mathcal{H}(\widetilde{\mathcal{X}}_{\alpha(\cdot)})$.

(ii) $\varphi \in H^{\alpha(\cdot)}(S_{\mathcal{X}})$ *iff* $\left\langle \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}\varphi, \varphi \right\rangle_{L^2(S_{\mathcal{X}})} < \infty$, *that is, iff* $\varphi \in \mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$.

(iii) For each real-valued function $\beta(\cdot)$ *in* $\mathcal{B}^{\infty}(\mathbb{R}^n)$, *the space of all* C^{∞}*-functions on* $\mathbb{R}^n$ *whose derivatives of all orders are bounded, let* $\mathcal{F}_{\beta(\cdot)}$ *be the set of real-valued, positive and continuous functions* $F_{\beta(\cdot)}$ *satisfying*

$$C \leq F_{\beta(\cdot)}(\lambda)(1+|\lambda|^2)^{\beta(\cdot)/2} \leq C', \quad C, C' > 0,$$

and $\mathcal{R}_{\beta(\cdot)}$, $\beta(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$, *be the class of covariance operators* $\mathrm{R}_{\beta(\cdot)}$ *such that*

$$\mathrm{R}_{\beta(\cdot)} = F_{\beta(\cdot)}(\mathrm{L}_{\alpha(\cdot)}), \quad \textit{for some } F_{\beta(\cdot)} \in \mathcal{F}_{\beta(\cdot)},$$

where $\mathrm{L}_{\alpha(\cdot)}$ is a self-adjoint elliptic pseudodifferential operator of variable order $\alpha(\cdot)$ on $L^2(S_{\mathcal{X}})$. Then, the estimation problem can be solved in a stable way, as given in Theorem 11, for the class of Gaussian observation random fields with covariance operator $\mathrm{R}_{\beta(\cdot)} \in \mathcal{R}_{\beta(\cdot)}$.

Proof

Assertions (i) and (ii) follow directly from Proposition 5. In relation to (iii), since $\mathrm{L}_{\alpha(\cdot)}$ is a self-adjoint elliptic pseudodifferential operator of variable order $\alpha(\cdot)$ on $L^2(S_{\mathcal{X}})$, from condition (3), any $\mathrm{R}_{\beta(\cdot)} \in \mathcal{R}_{\beta(\cdot)}$ is a self-adjoint elliptic pseudodifferential operator belonging to the space $\boldsymbol{\mathcal{S}}_{\rho,\delta}^{-\beta\alpha}$ (see Kikuchi and Negoro, 1995; Leopold, 1989, 1991, in relation to the main results for continuous elliptic pseudodifferential operators). The rest of the proof follows from Theorem 11 and Proposition 5. □

A8.4 The Infinite-Dimensional Formulation

In this section, orthogonal bases of the RKHS $\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$ are considered, allowing the spectral diagonalization of the covariance operator $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$. Specifically, the representation based on the covariance operator eigenvector system is provided. Furthermore, $L^2(S_{\mathcal{X}})$ dual Riesz bases are constructed for a biorthogonal expansion.

A8.4.1 Eigenfunction-Based Orthogonal Projection

Let us consider the case where $\underline{\alpha} > n/2$. Under this condition, our observation random field $\mathcal{X}_{\alpha(\cdot)}$ is continuous in the mean-square sense (see, for example, Triebel 1978, for embeddings between Besov spaces). In the Gaussian case, we also have continuity in the sample-path sense. Furthermore, the covariance operator $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ belongs to the trace class, i.e.,

$$\sum_{k=1}^{\infty} \lambda_k(\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}) < \infty,$$

where $\lambda_k(\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}})$, $k \geq 1$, are the eigenvalues of $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ on $L^2(S_{\mathcal{X}})$, i.e.,

$$\lambda_k \phi_k(\mathbf{z}) = \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\phi_k), \quad k \geq 1.$$

Theorem 14. *The observation random field $\mathcal{X}_{\alpha(\cdot)}$ admits the following orthogonal expansion:*

$$\mathcal{X}_{\alpha(\cdot)}(\varphi) \underset{\text{m.s.}}{=} \sum_{k\in\mathbb{N}} \sqrt{\lambda_k}\mathcal{X}_{\alpha(\cdot)}(\phi_k)\phi_k(\varphi), \quad \varphi \in L^2(S_{\mathcal{X}}),$$

where $\{\phi_k\}_{k\in\mathbb{N}}$ and $\{\lambda_k\}_{k\in\mathbb{N}}$, respectively, denote the systems of eigenfunctions and eigenvalues associated with the covariance operator $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ of $\mathcal{X}_{\alpha(\cdot)}$ on $L^2(S_{\mathcal{X}})$.

Under the condition $\underline{\alpha} > n/2$,

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) = \sum_{k\in\mathbb{N}} \sqrt{\lambda_k}\mathcal{X}_{\alpha(\cdot)}(\phi_k)\phi_k(\mathbf{z}), \quad \mathbf{z} \in S_{\mathcal{X}}.$$

Proof The proof follows directly from the spectral decomposition of a self-adjoint and compact operator, i.e., from the following series representation of $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ as a compact operator on $L^2(S_{\mathcal{X}})$ (see, for example, Dunford and Schwartz, 1963):

$$R_{\mathcal{X}_{\alpha(\cdot)}}(\varphi)(\psi) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\varphi)\phi_k(\psi), \quad \forall \varphi, \psi \in L^2(S_{\mathcal{X}}). \tag{A8.13}$$

Furthermore, under the condition $\underline{\alpha} > n/2$, the convergence of series (A8.13) holds pointwise using the embedding of fractional Sobolev spaces into Hölder spaces (see Triebel, 1978), that is,

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\varphi)(\mathbf{z}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\varphi)\phi_k(\mathbf{z}), \quad \varphi \in L^2(S_{\mathcal{X}}), \quad \forall \mathbf{z} \in S_{\mathcal{X}}.$$

In particular, $\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}$ admits an integral representation in terms of the kernel

$$C_{\mathcal{X}_{\alpha(\cdot)}}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{x})\phi_k(\mathbf{y}), \quad \mathbf{z}, \mathbf{y} \in S_{\mathcal{X}},$$

where convergence holds pointwise. That is,

$$\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\varphi)(\mathbf{z}) = \int_{S_{\mathcal{X}}} \left[\sum_{k=1}^{\infty} \lambda_k \phi_k(\mathbf{z})\phi_k(\mathbf{y}) \right] \varphi(\mathbf{y}) d\mathbf{y}, \quad \forall \mathbf{z} \in S_{\mathcal{X}}, \quad \forall \varphi \in L^2(S_{\mathcal{X}}). \quad \square$$

A8.4.2 Biorthogonal Expansion in terms of Dual Riesz Bases

This section provides an alternative framework for the biorthogonal expansion of a multifractional random field in terms of uncorrelated (independent, in the Gaussian case) coefficients. Specifically, as it was proved in Proposition 5, the RKHS norm of the random field $\mathcal{X}_{\alpha(\cdot)}$ is equivalent to the norm of the fractional Sobolev space of variable order $H^{\alpha(\cdot)}(S_{\mathcal{X}})$. Thus, the two spaces contain the same set of convergent sequences. However, under the existence of the pseudodual $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$, the concept of orthogonality in the RKHS can then be more easily defined than in the space $H^{\alpha(\cdot)}(S_{\mathcal{X}})$, in terms of the autocovariance operator of $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$. That is, for $\varphi, \psi \in \mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$,

$$\varphi \perp \psi \Leftrightarrow \mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}}(\varphi)(\psi) = 0.$$

Moreover, orthogonality in $\mathcal{H}(\mathcal{X}_{\alpha(\cdot)})$ is equivalent to the orthogonality in the closed subspace of $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ generated by the zero-mean second-order random variables

$$\{\mathcal{X}_{\alpha(\cdot)}(\varphi),\ \varphi \in [H^{\alpha(\cdot)}(S_{\mathcal{X}})]^* \equiv \bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}})\}.$$

In the following, we will denote as $H(\mathcal{X}_{\alpha(\cdot)})$ this closed subspace of $\mathcal{L}^2(\Omega, \mathcal{A}, P)$. Consequently, we are interested in the construction of an orthonormal basis of $H(\mathcal{X}_{\alpha(\cdot)})$, leading to an orthogonal expansion of random field $\mathcal{X}_{\alpha(\cdot)}$, in terms of uncorrelated random coefficients.

We now go back to the Hilbert-valued random variable context. That is, we consider $\mathcal{X}_{\alpha(\cdot)}$ being defined as an $H^{\alpha(\cdot)}(S_{\mathcal{X}})-$valued random variable. Then, we can define the orthogonal (in the space $H(\mathcal{X}_{\alpha(\cdot)})$) random sequence

$$\{\mathcal{X}_{\alpha(\cdot)}(\gamma^k),\ k \geq 1\},$$

constructed by the projection of the values of $\mathcal{X}_{\alpha(\cdot)}$ into the orthonormal system

$$\left\{\gamma^k = \mathrm{T}^{-1}_{\mathcal{X}_{\alpha(\cdot)}}(\phi_k),\ k \geq 1, \quad \right\}$$

of $\mathcal{H}(\widetilde{\mathcal{X}}_{\alpha(\cdot)})$, the RKHS of the $\bar{H}^{-\alpha(\cdot)}(S_{\mathcal{X}})-$valued random variable $\widetilde{\mathcal{X}}_{\alpha(\cdot)}$, generated by its auto-covariance operator

$$\mathrm{R}_{\widetilde{\mathcal{X}}_{\alpha(\cdot)}} = E[\widetilde{\mathcal{X}}_{\alpha(\cdot)} \otimes \widetilde{\mathcal{X}}_{\alpha(\cdot)}],$$

with $\mathrm{T}_{\mathcal{X}_{\alpha(\cdot)}}$ given in (A8.12). Note that

$$\begin{aligned} E\left[\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\mathcal{X}_{\alpha(\cdot)}(\gamma^l)\right] &= \left\langle \mathcal{X}_{\alpha(\cdot)}(\gamma^k), \mathcal{X}_{\alpha(\cdot)}(\gamma^l)\right\rangle_{H(\mathcal{X}_{\alpha(\cdot)})} \\ &= \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\gamma^k)(\gamma^l) = \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}\mathrm{R}^{-1}_{\mathcal{X}_{\alpha(\cdot)}}(\phi_k)(\phi_l) = \delta_{k,l}, \end{aligned} \tag{A8.14}$$

for an orthonormal basis $\{\phi_k,\ k \geq 1\}$ of $L^2(S_{\mathcal{X}})$, in view of (A8.12), and considering $\mathrm{R} = \mathrm{I}$ in (A8.11), with I denoting, as before, the identity operator. Hence, we obtain the following result.

Proposition 6. *Assume that* (*A*8.11) *is satisfied with* $\mathrm{R} = \mathrm{I}$. *Then,* $\mathcal{X}_{\alpha(\cdot)}$ *admits the following orthogonal expansion:*

$$\mathcal{X}_{\alpha(\cdot)}(\varphi) = \sum_{k=1}^{\infty} \mathcal{X}_{\alpha(\cdot)}(\gamma^k)\gamma_k(\varphi), \quad \forall \varphi \in L^2(S_{\mathcal{X}}), \tag{A8.15}$$

where convergence holds in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, *and*

$$\begin{aligned} \gamma^k &= \mathrm{T}^{-1}_{\mathcal{X}_{\alpha(\cdot)}}(\phi_k),\ k \geq 1, \\ \gamma_k &= \mathrm{T}_{\mathcal{X}_{\alpha(\cdot)}}(\phi_k),\ k \geq 1, \end{aligned} \tag{A8.16}$$

for an orthonormal basis $\{\phi_k,\ k \geq 1\}$ *of* $L^2(S_{\mathcal{X}})$.

In the case where $\underline{\alpha} > n/2$, *equality (A8.15) holds pointwise, i.e,*

$$\mathcal{X}_{\alpha(\cdot)}(\mathbf{z}) = \sum_{k=1}^{\infty} \mathcal{X}_{\alpha(\cdot)}(\gamma^k)\gamma_k(\mathbf{z}), \quad \forall \mathbf{z} \in S_{\mathcal{X}}. \tag{A8.17}$$

Proof From (A8.14), the random Fourier coefficients appearing in expansions (A8.15) and (A8.17) are uncorrelated.

The proof is then obtained from the following identities: For a fixed M, from (A8.12), in

view of (A8.16),

$$\begin{aligned}
E\left[\mathcal{X}_{\alpha(\cdot)}(\varphi)-\sum_{k=1}^{M}\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\gamma_k(\varphi)\right]^2 &= E\left[\mathcal{X}_{\alpha(\cdot)}(\varphi)\right]^2 \\
&+ \sum_{k=1}^{M}\sum_{p=1}^{M}E\left[\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\mathcal{X}_{\alpha(\cdot)}(\gamma^p)\right]\gamma_k(\varphi)\gamma_p(\varphi) \\
&- 2\sum_{k=1}^{M}E\left[\mathcal{X}_{\alpha(\cdot)}(\varphi)\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\right]\gamma_k(\varphi) \\
&= \mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\varphi)(\varphi)+\sum_{k=1}^{M}\gamma_k^2(\varphi) \\
&- 2\sum_{k=1}^{M}\mathrm{R}_{\mathcal{X}_{\alpha(\cdot)}}(\gamma^k)(\varphi)\gamma_k(\varphi) \\
&= \sum_{k=1}^{\infty}\gamma_k^2(\varphi)+\sum_{k=1}^{M}\gamma_k^2(\varphi)-2\sum_{k=1}^{M}\gamma_k^2(\varphi) \\
&= \sum_{k=1}^{\infty}\gamma_k^2(\varphi)-\sum_{k=1}^{M}\gamma_k^2(\varphi),
\end{aligned}$$

which converges to zero when $M \longrightarrow \infty$. □

Corollary 4. *Assume that the conditions of Proposition 6 are satisfied. Let* $\widehat{\mathcal{Y}}_{\beta(\cdot)}$ *be the best linear estimator of* $\widehat{\mathcal{Y}}_{\beta(\cdot)}=\mathrm{K}\mathcal{X}_{\alpha(\cdot)}$ *of* $\mathcal{Y}_{\beta(\cdot)}$ *given*

$$\mathcal{X}_{\alpha(\cdot)}=S_{\gamma(\cdot)}+N_{\theta(\cdot)}.$$

Then, $\widehat{\mathcal{Y}}_{\beta(\cdot)}$ *admits the following weak-sense transformed orthogonal expansion: For all* $\varphi \in L^2(S_{\mathcal{X}})$,

$$\begin{aligned}
\widehat{\mathcal{Y}}_{\beta(\cdot)}(\varphi) &= \sum_{k=1}^{\infty}\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\mathrm{K}(\gamma_k)(\varphi) \\
&= \sum_{k=1}^{\infty}\mathcal{X}_{\alpha(\cdot)}(\gamma^k)\int_{S_{\mathcal{X}}}\left[\int_{\mathbb{R}^n}e^{i\langle\mathbf{z},\boldsymbol{\xi}\rangle}p_{\mathrm{K}}(\mathbf{z},\boldsymbol{\xi})\widehat{\gamma}_k(\boldsymbol{\xi})d\boldsymbol{\xi}\right]\varphi(\mathbf{z})d\mathbf{z},
\end{aligned}$$

and, if $\underline{\alpha} > n/2$, *for all* $\mathbf{z} \in S$,

$$\begin{aligned}\widehat{\mathcal{Y}}_{\beta(\cdot)}(\mathbf{z}) &= \sum_{k=1}^{\infty} \mathcal{X}_{\alpha(\cdot)}(\gamma^k)\mathrm{K}(\gamma_k)(\mathbf{z}) \\ &= \sum_{k=1}^{\infty} \mathcal{X}_{\alpha(\cdot)}(\gamma^k) \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle} p_{\mathrm{K}}(\mathbf{z},\boldsymbol{\xi})\widehat{\gamma}_k(\boldsymbol{\xi}) d\boldsymbol{\xi}.\end{aligned}$$

Remark 26. *Similar assertions hold for the best functional linear estimator of* $\mathcal{S}_{\gamma(\cdot)}$ *given the functional sample information* $\mathcal{X}_{\alpha(\cdot)} = \mathcal{Y}_{\beta(\cdot)} + N_{\theta(\cdot)}$.

The proof follows directly from Proposition 6 after applying the Orthogonal Projection Theorem to compute K.

A8.5 Examples

In this section, we formulate several examples of system equations involving pseudodifferential operators of variable order.

Example 8.1

We first consider the following integral equation, which generalizes the one defined by Bessel potential $(\mathrm{I} - \Delta)^{-\beta}$, $\beta > 0$, of fixed order β (see, for example, Anh, Angulo and Ruiz-Medina, 1999 and Anh and Leonenko, 2001):

$$\mathcal{Y}_{\beta(\cdot)} = (\mathrm{I} - \Delta)^{-\nu(\cdot)/2} S_{\gamma(\cdot)}, \quad \beta, \nu \in \mathcal{B}^{\infty}(\mathbb{R}^n),$$

where the identity is understood in the mean-square sense. Note that, as usual, $(-\Delta)$ represents the negative Laplacian operator, I denotes the identity operator, and

$$\mathcal{S}_{\gamma(\cdot)}(\mathbf{z}) = \mathrm{L}(-\Delta)^{-\gamma(\cdot)/2}\varepsilon(\mathbf{z}), \quad \mathbf{z} \in S \subseteq \mathbb{R}^n, \quad \underline{\gamma} > n/2,\ \overline{\gamma} \leq n, \quad \gamma \in \mathcal{B}^{\infty}(\mathbb{R}^n).$$

Thus, $\beta(\cdot) = \nu(\cdot) + \gamma(\cdot)$, since L is assumed to be an integral operator with homogeneous continuous kernel L, slowly varying at infinity, and with Fourier transform displaying the same behavior at zero frequency. Here, ε denotes white noise.

Equivalently,

$$\begin{aligned} \mathcal{Y}_{\beta(\mathbf{z})}(\mathbf{z}) &= \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{-\nu(\mathbf{z})/2}\widehat{S}_{\gamma(\cdot)}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{-\nu(\mathbf{z})/2}\frac{\widehat{L}(1/|\boldsymbol{\xi}|)}{|\boldsymbol{\xi}|^{\gamma(\mathbf{z})}}\widehat{\varepsilon}(\boldsymbol{\xi})d\boldsymbol{\xi}, \end{aligned}$$

where $\widehat{\varepsilon}$ represents a complex white noise measure in the spectral domain.

Example 8.2

Let us now consider the following extension of the Riesz-Bessel equation (see Anh, Angulo and Ruiz-Medina, 1999 and Anh and Leonenko, 2001)

$$\mathcal{Y}_{\beta(\cdot)}(\mathbf{z}) = (\mathrm{I}-\Delta)^{-\nu(\mathbf{z})/2}(-\Delta)^{-\rho(\mathbf{z})/2}S_{\gamma(\cdot)}(\mathbf{z}), \quad \beta,\rho \in \mathcal{B}^{\infty}(\mathbb{R}^n),$$

where

$$\mathcal{S}_{\gamma(\mathbf{z})}(\mathbf{z}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle}\frac{\widehat{\varepsilon}(\boldsymbol{\xi})}{1+|\boldsymbol{\xi}|^{\gamma(\mathbf{z})}}d\boldsymbol{\xi}, \quad \underline{\gamma} > n/2, \quad \gamma \in \mathcal{B}^{\infty}(\mathbb{R}^n).$$

That is,

$$\mathcal{Y}_{\beta(\mathbf{z})}(\mathbf{z}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle}(1+|\boldsymbol{\xi}|^2)^{-\nu(\mathbf{z})/2}|\boldsymbol{\xi}|^{-\rho(\mathbf{z})}\frac{\widehat{\varepsilon}(\boldsymbol{\xi})}{1+|\boldsymbol{\xi}|^{\gamma(\mathbf{z})}}d\boldsymbol{\xi}.$$

Example 8.3

This example introduces a class of random fields with variable order fractional rational spectra, extending the the family of homogeneous random fields with rational covariance spectrum (see Ramm, 2005). Indeed, this family constitutes a particular case of the one considered in Corollary 3(iii). Specifically, for $\mathrm{L}_{\alpha(\cdot)}$ being a self-adjoint elliptic pseudodifferential operator of variable order $\alpha(\cdot)$ on $L^2(S_{\mathcal{X}})$, we consider the following special class of functions:

$$F_{\beta(\cdot)}(\boldsymbol{\lambda},\mathbf{z}) = \frac{\mathrm{P}(\boldsymbol{\lambda},\mathbf{z})}{\mathrm{Q}(\boldsymbol{\lambda},\mathbf{z})},$$

with $\mathrm{P}(\boldsymbol{\lambda},\mathbf{z})$ and $\mathrm{Q}(\boldsymbol{\lambda},\mathbf{z})$ being positive elliptic polynomials of respective variable orders $p(\cdot)$ and $q(\cdot)$ satisfying

$$\begin{aligned} 0 < c_1 \leq \mathrm{P}(\boldsymbol{\lambda},\mathbf{z})(1+|\boldsymbol{\lambda}|^2)^{-p(\mathbf{z})/2} \leq c_2, \\ 0 < c_3 \leq \mathrm{Q}(\boldsymbol{\lambda},\mathbf{z})(1+|\boldsymbol{\lambda}|^2)^{-q(\mathbf{z})/2} \leq c_4, \end{aligned} \tag{A8.18}$$

for some positive constants c_1, c_2, c_3 and c_4. That is, $\beta(\cdot) = q(\cdot) - p(\cdot)$, and $\alpha(\cdot), p(\cdot), q(\cdot) \in \mathcal{B}^{\infty}(\mathbb{R}^n)$. In particular, (A8.18) is satisfied if P and Q are positive continuous polynomials of respective orders $p(\cdot)$ and $q(\cdot)$, with $\mathrm{P}(\boldsymbol{\lambda},\mathbf{z}) = \mathrm{P}(|\boldsymbol{\lambda}|,\mathbf{z})$ and $\mathrm{Q}(\boldsymbol{\lambda},\mathbf{z}) = \mathrm{Q}(|\boldsymbol{\lambda}|,\mathbf{z})$. For example, we can consider

$$\begin{aligned} \mathrm{P}(\mathbf{z},\boldsymbol{\lambda}) &= a_0 + \sum_{k=1}^{n} a_k |\boldsymbol{\lambda}|^k + (c+|\boldsymbol{\lambda}|)^{p(\mathbf{z})/2}, \quad a_k > 0, \quad k \geq 0, \quad c > 0, \\ \mathrm{Q}(\mathbf{z},\boldsymbol{\lambda}) &= b_0 + \sum_{k=1}^{l} b_k (1+|\boldsymbol{\lambda}|^2)^{k/2} + (d+|\boldsymbol{\lambda}|^2)^{q(\mathbf{z})/2}, \quad b_k > 0, \quad k \geq 0, \quad d > 0, \end{aligned}$$

for $\mathbf{z} \in \mathbb{R}^n$, taking n and l such that $\min_{\mathbf{z}\in\mathbb{R}^n} p(\mathbf{z}) > n$, and $\min_{\mathbf{z}\in\mathbb{R}^n} q(\mathbf{z}) > l$, and $p(\mathbf{z}) \leq q(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^n$.

Alternatively, we can also introduce examples of the referred class of fractional random fields with rational covariance spectra of variable order, in terms of the following identities in the mean-square sense:

$$\mathcal{Y}_{\beta(\mathbf{z})}(\mathbf{z}) = \int_{\mathbb{R}^n} e^{i\langle \mathbf{z},\boldsymbol{\xi}\rangle} \frac{\mathrm{P}\left((1+|\boldsymbol{\xi}|^2)^{\nu(\mathbf{z})}\right)}{\mathrm{Q}\left((1+|\boldsymbol{\xi}|^2)^{\varrho(\mathbf{z})}\right)} \frac{\widehat{\varepsilon}(\boldsymbol{\xi})}{(1+|\boldsymbol{\xi}|^{\alpha_1(\mathbf{z})})^{\alpha_2(\mathbf{z})}} d\boldsymbol{\xi},$$

where $\beta, \nu, \varrho, \alpha_1, \alpha_2 \in \mathcal{B}^{\infty}(\mathbb{R}^n)$, P and Q are positive elliptic polynomials of order p and q, respectively, with $q(\mathbf{z}) \geq p(\mathbf{z})$, for all $\mathbf{z} \in \mathbb{R}^n$. Moreover, it is assumed that $\nu(\mathbf{z}) \leq \varrho(\mathbf{z})$, for all $\mathbf{z} \in \mathbb{R}^n$, and $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are such that $\underline{\alpha}_1 = \inf_{\mathbf{z}} \alpha_1(\mathbf{z}) > n/2$, and $\underline{\alpha}_2 = \inf_{\mathbf{z}} \alpha_2(\mathbf{z}) > n/2$.

A8.6 Simulation Study

To illustrate the performance of the projection estimation methodology proposed, let us know consider some particular examples of the families of multifractional random field models introduced before. First, we summarize the main steps of the estimation algorithm implemented in these examples.

Step 1. The functional values of the signal $\mathcal{S}_{\gamma(\cdot)}$ and the observation noise $\mathcal{N}_{\theta(\cdot)}$, are interpolated from their discrete generations over a 10×10 regular spatial grid covering the region $S_{\mathcal{X}}$. We consider the Gaussian case.

Step 2. The kernel of the pseudodifferential operator of variable order A defining the equation system is evaluated on the region $S \times S_{\mathcal{X}} = [0, 15]^4$.

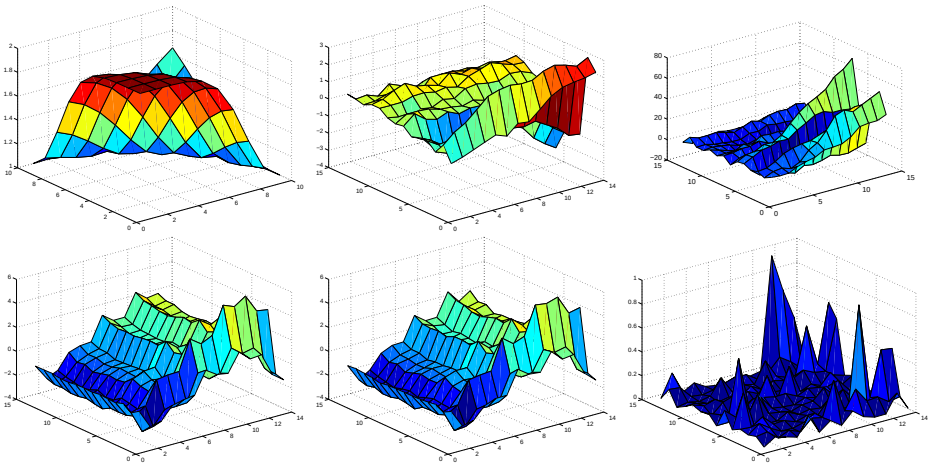
Step 3. The random output $\mathcal{Y}_{\beta(\cdot)}$ is then computed from the product of the symbols p_{A} and $p_{\mathcal{S}_{\gamma(\cdot)}}$ defining the pseudodifferential operator A, and the operator involved in the generation of $\mathcal{S}_{\gamma(\cdot)}$, respectively.

Step 4. The empirical autocovariance operator of the functional observations and the cross-covariance operator between the random output and the functional observations are then calculated from (A8.3), based on a random functional sample of size 100.

Step 5. The empirical eigenfunction system associated with the autocovariance operator computed in the previous step is then considered for projection. Truncation at term M is chosen by cross-validation.

Step 6. The linear filter defining K, computed in terms of the empirical autocovariance operator and the cross-covariance operator, is projected into the finite-dimensional empirical eigenfunction basis.

Figure A8.1: Example 8.1 Functional $\gamma(\cdot)$ exponent (top-left), the corresponding signal $S_{\gamma(\cdot)}$ (top-center) and one functional observation $\mathcal{X}_{\alpha(\cdot)} = \mathcal{S}_{\gamma(\cdot)} + N_{\theta(\cdot)}$ (top-right). Output functional value, $\mathcal{Y}_{\beta(\cdot)}$ (bottom-left), the corresponding functional estimation (bottom-center), and empirical functional mean quadratic errors based on 100 simulations (bottom-right)



Step 7. Numerical methods are applied for inversion of the empirical finite-dimensional estimation equation system.

Step 8. The inverse eigenfunction transform provides the approximation of K in the space domain.

Step 9. A new random functional sample is generated of size 100 for evaluating the linear mean-square predictor of the output, as well as the associated functional empirical mean quadratic errors.

A8.6.1 Numerical Examples

The following functional forms of the involved multifractional parameters are respectively considered in Examples 8.1-8.3. Specifically, for $(x, y) \in [0, 15] \times [0, 15]$, in Example 8.1, we establish

$$\begin{aligned}
\gamma(x,y) &= 1 + C_1(\cos((C_2\pi)/(1+\|(x,y)\|^2)))^2, \quad C_1 = C_2 = 1, \\
\nu(x,y) &= D_1 + \exp\left(-\|(x,y)\|/D_2\right), \quad D_1 = 1, \quad D_2 = 1/10 \\
\theta(x,y) &= K/(1+\|(x,y)\|^2), \quad K = 1.
\end{aligned}$$

In Example 8.2, the multifractional exponents involved are given by the following functional forms: For $(x,y) \in [0,15] \times [0,15]$,

$$\begin{aligned}
\gamma(x,y) &= C_1 + (\cos(\pi/(1+\|(x,y)\|^2))/C_2)^2, \quad C_1 = 1,\ C_2 = 10 \\
\rho(x,y) &= B_1 + \|(x,y)\|^{-2.4}, \quad B_1 = 1 \\
\nu(x,y) &= D_1 + \exp\left(-\|(x,y)\|^{1/2}/D_2\right), \quad D_1 = 1,\ D_2 = 1/10 \\
\theta(x,y) &= (\cos(K\pi/(1+\|(x,y)\|^2)))^2, \quad K = 1.
\end{aligned}$$

Finally, in Example 8.3, we consider

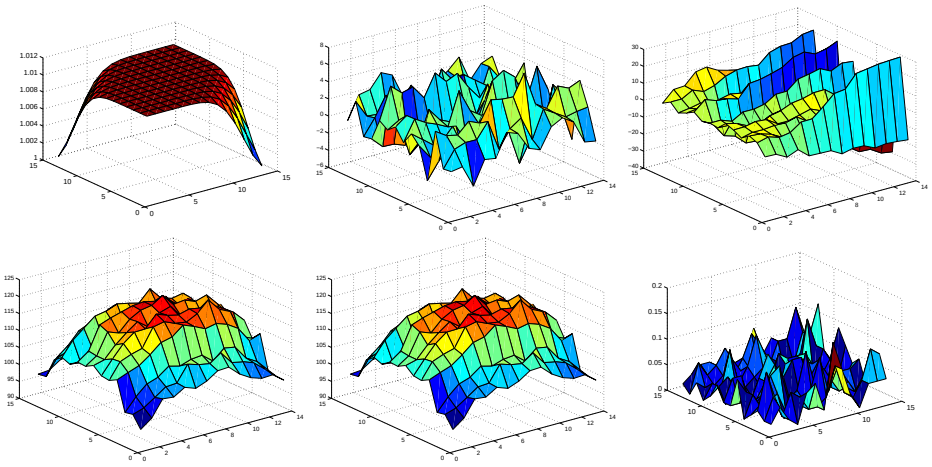
$$B_1 = C_1 = C_2 = E_1 = 1, \quad B_2 = D_2 = 1/10, \quad K = 1, \quad \alpha_2 \equiv 1 + \epsilon, \quad \epsilon > 0,$$

and for $(x,y) \in [0,15] \times [0,15]$,

$$\begin{aligned}
\gamma(x,y) &= C_1 + \exp\left(-(C_2/(1+\|(x,y)\|^2)\right) \\
\nu(x,y) &= D_1 + \exp\left(-\|(x,y)\|^{1/2}/D_2\right) \\
\rho(x,y) &= B_1 + \exp\left(-\|(x,y)\|/B_2\right) \\
\varrho(x,y) &= \cos(\pi/4\|(x,y)\|) \\
\alpha_1(x,y) &= E_1 + \cos(\pi + (\pi/2)\|(x,y)\|^{\theta/2}) \\
\theta(x,y) &= (\sin(C_2\pi/(1+\|(x,y)\|^2)))^2.
\end{aligned}$$

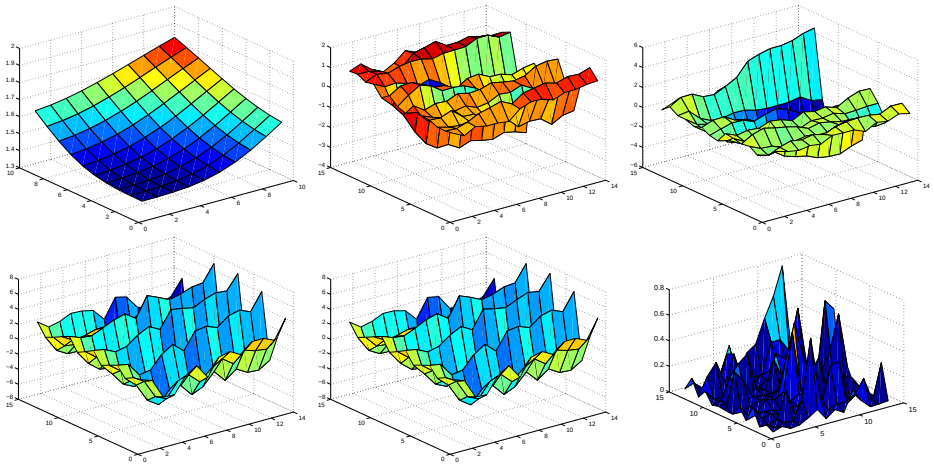
The functional exponent γ of the random signal $S_{\gamma(\cdot)}$ is displayed, for the three examples studied, in the top-left plots of Figures A8.1, A8.2 and A8.3. The sample functional values

Figure A8.2: Example 8.2 Functional $\gamma(\cdot)$ exponent (top-left), the corresponding signal $S_{\gamma(\cdot)}$ (top-center) and one functional observation $\mathcal{X}_{\alpha(\cdot)} = \mathcal{S}_{\gamma(\cdot)} + N_{\theta(\cdot)}$ (top-right). Output functional value, $\mathcal{Y}_{\beta(\cdot)}$ (bottom-left), the corresponding functional estimation (bottom-center), and empirical functional mean quadratic errors based on 100 simulations (bottom-right)



of the signal $\mathcal{S}_{\gamma(\cdot)}$ and of the functional observation $\mathcal{X}_{\alpha(\cdot)}$ are respectively displayed next to it. Additionally, a sample functional value of the output, its best linear functional mean-square estimate and the corresponding functional empirical mean-square errors (FEMSEs) are given in the bottom row of Figures A8.1-A8.3. Finally, the L^∞ norms of the FEMSEs computed for such examples are given after these figures. Note that the highest truncation level has been considered in Example 8.2, with $M = 150$ eigenfunctions in the finite-dimensional approximation, since a higher degree of local singularity is introduced by the system operator A, due to the functional form of its parameter ρ, and by the multifractional noise $\mathcal{N}_{\theta(\cdot)}$, in terms of its non-linear hyperbolic (cosine-type) parameter $\theta(\cdot)$. As it can be seen in Figures A8.2, the spatial structure of the random output $\mathcal{Y}_{\beta(\cdot)}$ is completely different from the signal, whose spatial patterns are, at the same time, completely distorted by the multifractional noise. Hence, a higher truncation level is required for the estimation of $\mathcal{Y}_{\beta(\cdot)}$ from the observation of $\mathcal{X}_\alpha$. In addition, the functional estimates of $\mathcal{Y}_{\beta(\cdot)}$ are very sensitive to the self-reference problem, i.e., to

Figure A8.3: Example 8.3 Functional $\gamma(\cdot)$ exponent (top-left), the corresponding signal $S_{\gamma(\cdot)}$ (top-center) and one functional observation $\mathcal{X}_{\alpha(\cdot)} = \mathcal{S}_{\gamma(\cdot)} + N_{\theta(\cdot)}$ (top-right). Output functional value, $\mathcal{Y}_{\beta(\cdot)}$ (bottom-left), the corresponding functional estimation (bottom-center), and empirical functional mean quadratic errors based on 100 simulations (bottom-right)



the presence of common functional data in the two samples randomly selected for the estimation of the functional parameter K and of the functional values of $\mathcal{Y}_{\beta(\cdot)}$, and for the computation of the functional empirical mean-square errors. Note that the two functional samples of size 100 are randomly selected from a general population constituted by 150 realizations of the random signal, observation noise, and output. Finally, for Examples 8.1 and 8.3, the same truncation level $M = 100$ is considered in the finite-dimensional approximation of the associated estimation problems. Slightly better performance is can be appreciated in Example 8.3, due to the higher level of self-reference in the randomly selected functional samples for the parameter and output values estimation. The smoother form of $\gamma(\cdot)$ in Example 8.3 also leads to a lower local heterogeneity level in the realizations, obtaining a more accurate estimation.

L^∞-norm of FEMSEs	
Example 8.1	0.9597
Example 8.2	0.1421
Example 8.3	0.7566

A8.7 Conclusions

In this paper, the problem of least squares estimation of an input random field from the observation of an output random field, and reciprocally, the estimation of an output random field from the observation of an input random field, when observations are affected by additive noise, is addressed in a multifractional random field framework. Regularization of this problem is achieved under the assumption of the existence of a pseudodual. Numerical projection methods into orthogonal and biorthogonal bases are considered for the suitable finite-dimensional approximation of both least-squares direct and inverse estimators of the random signal of interest.

The results derived allow one to define a stable solution to the least-squares multifractional estimation problem considered. However, the multifractal case still remains as an open problem. In the recent literature, different approaches have been adopted for the construction of multifractal processes and fields (see, for example, the proposal given in Anh, Leonenko and Shieh, 2008). An alternative way of introducing multifractal random field models can be found as an extension of the framework considered in this paper. Specifically, the regularity conditions assumed on the exponent function, defining the fractional-order of differentiation in the weak-sense, in multifractional Sobolev spaces can be relaxed by introducing it as the limit of regular functions in a suitable topology. This more flexible framework allows the characterization of limit spaces having functions with non-trivial singularity spectra.

List of Figures

List of Tables

Bibliography

[1] Aach, J. and Church, G. M. (2001). Alignment gene expression time series with time warping algorithms. *Bioinformatics,* **17**, 495–508.

[2] Aalen, O. O. and Gunnes, N. (2010). A dynamic approach for reconstructing missing longitudinal data using the linear increments model. *Biostatistics,* **11**, 453–472.

[3] Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications.

[4] Abraham, C., Cornillon, P. A., Matzner-Lober, E. and Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics,* **30**, 581—595.

[5] Abramovich, F. and Angelini, C. (2006). Testing in mixed-effects FANOVA models. *Journal of Statistical Planning and Inference,* **136**, 4326–4348.

[6] Adler, R. J. (1981). *The Geometry of Random Fields.* Wiley, Chichester.

[7] Aguilera, A. M., Escabias, M., Preda, C. and Saporta, G. (2010). Using basis expansion for estimating functional PLS regression. Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems,* **104**, 289–305.

[8] Aguilera-Morillo, M. C., Aguilera, A. M., Escabias, M. and Valderrama, M. J. (2013). Penalized spline approaches for functional logit regression. *TEST,* **22**, 251–277.

[9] Alonso, A. M., Casado, D. and Romo, J. (2012). Supervised classification for functional data: A weighted distance approach. *Computational Statistics & Data Analysis,* **56**, 2334—2346.

[10] Andel, J. (1986). Long memory time series models. *Kybernetika,* **22**, 105–123.

[11] Aneiros-Pérez, G., Cao, R. and Vilar-Fernández, J. M. (2011). Functional methods for time series prediction: a nonparametric approach. *Journal of Forecasting,* **30**, 377-–392.

[12] Angelini, C., De Canditiis, D. and Leblanc, F. (2003). Wavelet regression estimation in nonparametric mixed effect models. *Journal of Multivariate Analysis,* **85**, (2): 267–291.

[13] Angelini, C., De Canditiis, D. and Pensky, M. (2012). Clustering time-course Microarray data using functional Bayesian infinite mixture model. *Journal of Applied Statistics,* **39**, 129–149.

[14] Anh, V. V., Angulo, J. M. and Ruiz-Medina, M. D. (1999). Possible long-range dependence in fractional random fields. *Journal of Statistical Planning and Inference,* **80**, 95–110.

[15] Anh, V. V. and Leonenko, N. N. (2001). Spectral analysis of fractional kinetic equations with random data. *Journal of Statistical Physics,* **104**, 1349–1387.

[16] Anh, V. V. Leonenko, N. N. and Sakhno, L. M. (2004). On a class of minimum contrast estimators for fractional stochastic processes and fields. *Journal of Statistical Planning and Inference,* **123**, 161–185.

[17] Anh, V. V., Leonenko, N. N. and Sakhno, L. M. (2007). Minimum contrast estimation of random processes based on information of second and third orders. *Journal of Statistical Planning and Inference,* **137**, (4): 1302–1331.

[18] Anh, V., Leonenko, N. N. and Shieh, N. R. (2008). Multifractality of products of geometric Ornstein-Uhlenbeck-type processes. *Advances in Applied Probability,* **40**, 1129–1156.

[19] Anh, V. V. and Lunney, K. E. (1995). Parameter estimation of random fields with long-range depedence. *Mathematical and Computer Modelling,* **21**, 67–77.

[20] Antoniadis, A., Paparoditis, E. and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journals of the Royal Statistical Society. Series B,* **68**, (5): 837-–857.

[21] Antoniadis, A., Paparoditis, E. and Sapatinas, T. (2009). Bandwidth selection for functional time series prediction. *Statistics & Probability Letters,* **79**, (6): 733–740.

[22] Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis,* **87**, (1): 133—158.

[23] Aquaro, M. and Cižek, P. (2014). Robust estimation of dynamic fixed-effects panel data models. *Statistical Papers,* **55**, 169-–186.

[24] Araki Y., Konishi S., Kawano S. and Matsui H., (2009). Functional logistic discrimination via regularized basis expansions. *Communication in Statistics- Theory and Methods,* **38**, 2944–2957.

[25] Arellano, M. and BonHomme, S. (2012). Identifying distributional characteristics in random coefficients panel data models. *Economic Studies,* **79**, 987—1020.

[26] Arellano, M. and Honoré, B. (2001). *Panel data models: some recent developments.* In: J.J. Heckman and E.E. Leamer (Eds.). *Handbook of Econometrics.* Elsevier Science, Chapter **53**, 3229–3296.

[27] Arteche, J. and Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis,* **21**, (1): 1–25.

[28] Ayache, A. and Lévy-Véhel, J. (2000). The generalized multifractional Brownian motion. *Statistical Inference for Stochastic Processes,* **3**, 7—18.

[29] Badinger, H. and Egger, P. (2013). Estimation and testing of higher-order spatial autoregressive panel data error component models. *Journal of Geographical Systems*, **15**, 453-–489.

[30] Baíllo, A. and Cuevas, A. (2008). Supervised functional classification: a theoretical remark and some comparisons. *http://arxiv.org/abs/0806.2831*

[31] Baladandayuthapani, V., Mallick, B., Hong, M., Lupton, J., Turner, N. and Caroll, R. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinoginesis. *Biometrics,* **64**, (1): 64—73.

[32] Baltagi, B. H., Bresson, G. and Pirotte, A. (2012). Forecasting with spatial panel data. *Computational Statistics & Data Analysis,* **56**, 3381—3397.

[33] Baltagi, B. H., Fingleton, B. and Pirotte, A. (2014). Estimating and Forecasting with a Dynamic Spatial Panel Data Model. *Economics and Statistics,* **76**, (1): 112–138.

[34] Baltagi, B. H. and Li, D. (2004). Prediction in the panel data model with spatial correlation. *Advances in Spatial Science,* 283–295.

[35] Baltagi, B. H. and Pirotte, A. (2010). Panel data inference under spatial dependence. *Economic Modelling,* **27**, 1368—1381.

[36] Basse, M., Diop, A. and Dabo-Niang, S. (2008). *Mean square properties of a class of kernel density estimates for spatial functional random variables.* Annales De L'I.S.U.P. Publications de l'Institut de Statistique de l'Université de Paris, 91-108, Numéro Spécial-Volume LII, Fascicule 1-2, Paris.

[37] Basu, S. and Reinsel, G. C. (1993). Properties of the spatial unilateral first-order $ARMA$ model. *Advances in Applied Probability,* **25**, (3): 631–648.

[38] Benassi, A., Jaffard, S. and Roux, D. (1997). Elliptic Gaussian random processes. *Revista Matematica Iberoamericana,* **13**, 19—90.

[39] Bentkus, R. (1972). The error in estimating the spectral function of a stationary proces. *Litovskii Matematicheskii Sbornik,* **12**, (1): 55–71.

[40] Beran, J., Ghosh, S. and Schell, D. (2009). On least squares estimation for long-memory lattice processes. *Journal of Multivariate Analysis,* **100**, 2178–2194.

[41] Berlinet, A., Biau, G. and Rouvière, L. (2008). Functional supervised classification with wavelets. *Annales de l'ISUP,* **52**, 61—80.

[42] Besag, J., York, J. and Molié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics,* **43**, (1): 1–59.

[43] Besse, P. C., Cardot, H. and Stephenson, D. B. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics,* **27**, (4): 673–687.

[44] Bhaird, C. and Lucey, B. (2010). Determinants of capital structure in Irish SMEs. *Small Business Economics,* **35**, 357–375.

[45] Bhattacharyya, B. B., Khalil, T. M. and Richardson, G. D. (1996). Gauss–Newton estimation of parameters for a spatial autoregression model. *Statistics & Probability Letters,* **28**, 173–179.

[46] Biau, G., Bunea, F. and Wegkamp, M. H. (2003). Functional classification in Hilbert spaces. *IEEE Transactions on Information Theory,* **1**, 1—8.

[47] Boissy, Y., Bhattacharyya, B. B., Li, X. and Richardson, G. D. (2005). Parameter estimates for fractional autoregressive spatial process. *Annals of Statistics,* **33**, 2553–2567.

[48] Booth, L., Aivazian, V., Demirguç-Kunt, L. and Maksimovic, V. (2001). Capital structures in developing countries. *Journal of Finance,* **56**, 87–130.

[49] Bosq, D. (1996). Limit theorems for Banach-valued autoregressive processes. Applications to real continuous time processes. *Bulletin of the Belgian Mathematical Society,* **3**, (5): 537–555.

[50] Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer-Verlag, New York.

[51] Bosq, D. (2009). A note on asymptotic parametric prediction. *Journal of Statistical Planning and Inference,* **139**, (4): 1506–1513.

[52] Bosq, D. (2010). Tensorial products of functional $ARMA$ processes. *Journal of Multivariate Analysis,* **101**, (6): 1352–1363.

[53] Bosq, D. and Blanke, D. (2007). *Inference and Predictions in Large Dimensions.* John Wiley and sons, Paris.

[54] Brockwell, R. A. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

[55] Brown, P. E., Karesen, K. F., Roberts, G. O. and Tonellato, S. (2000). Blur-generated non-separable space-time models. *Journals of the Royal Statistical Society. Series B,* **62**, (5): 847–860.

[56] Brychkov, Y. A., Glaeske, H. J., Prudnikov, A. P. and Vũ, K. T. (1992). *Multidimensional integral transformations,* Gordon and Breach Science Publishers.

[57] Burnett, J. W., Bergstromb, J. C. and Dorfman, J. H. (2013). A spatial panel data approach to estimating U.S. state-level energy emissions. *Energy Economics,* **40**, 396-–404.

[58] Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics,* **34**, 2159–2179.

[59] Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics,* **30**, 241–255.

[60] Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters,* **45**, 11–22.

[61] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica,* **13**, 571–591.

[62] Cardot, H. and Sarda, P. (2005). Estimation in generalized linear model for functional data via penalized likelihood. *Journal of Multivariate Analysis,* **92**, 24–41.

[63] Cardot, H. and Sarda, P. (2011). Functional linear regression. *In The Oxford handbook of functional data analysis.* Oxford Univ. Press, Oxford, 21-–46.

[64] Case, A., Rosen, H. and Hines, J. (1993). Budget spillovers and fiscal policy interdependence: evidence from the states. *Journal of Public Economics,* **52**, 285–307.

[65] Chan, K. S. and Tsai, H. (2012). Inference of seasonal long-memory aggregate time series. *Bernoulli,* **4**, (18): 1448–1464.

[66] Chen, J., Gao, J. and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics,* **171**, 71-–85.

[67] Cheng, Q. C. (1993). *Transfer function model and GARMA II model.* Ph.D. thesis, Southern Methodist University.

[68] Chiou, J. M. and Li, P. L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journals of the Royal Statistical Society. Series B,* **69**, 679—699.

[69] Chiou, J. M. and Li, P. L. (2008a). Correlation-based functional clustering via subspace projection. *Journal of the American Statistical Association,* **103**, 1684—1692.

[70] Chiou, J. M. and Li, P. L. (2008b). *Functional clustering of longitudinal data.* In: S. Dabo-Niang, F. Ferraty (Eds.). *Functional and Operatorial Statistics*. Physica- Verlag, Heidelberg, 103—107.

[71] Chittenden, F., Hall, G. and Hutchinson, P. (1996). Small firm growth, access to capital markets and financial structure: review of issues and an empirical investigation. *Small Business Economics,* **8**, 59–67.

[72] Christakos, G. (1992). *Random field models in earth science.* Dover Publications, INC, Mineola, New York.

[73] Christakos, G. (2000). *Modern Spatiotemporal Geostatistics.* Oxford, University Press, New York .

[74] Christakos, G. and Laf, J. J. (1997). A study of the breast cancer dynamics in North Carolina. *Social Science & Medicine,* **45**, (10): 1503–1517.

[75] Christakos, G., Olea, R. A. and Yu, H. L. (2007). Recent results on the spatio-temporal modelling and comparative analysis of Black Death and bubonic plague epidemics. *Public Health,* **121**, (9): 700–720.

[76] Christakos, G. and Raghu, V. R. (1996). Dynamic stochastic estimation of physical variables. *Journal of Mathematical Geology,* **28**, (3): 341–365.

[77] Chung, C. F. (1996a). A generalized fractionally integrated autoregressive moving-average process. *Journal of Time Series Analysis,* **17**, 111–140.

[78] Chung, C. F. (1996b). Estimating a generalized long memory process. *Journal of Econometric,* **73**, 237–259.

[79] Chyzak, F., Paule, P., Scherzer, O., Schoisswohl, A. and Zimmermann, B. (2001). The construction of orthonormal wavelets using symbolic methods and a matrix analytical approach for wavelets on the interval. *Experimental mathematics,* **10**, 66—86.

[80] Cohen, A. Daubechies, I. and Vial, P. (1994). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis,* **1**, 54–81.

[81] Cohen, G. and Francos, J. M. (2002). Linear least squares estimation of regression models for two-dimensional random fields. *Journal of Multivariate Analysis,* **82**, 431–444.

[82] Collet, J. J. and Fadili, M. J. (2006). *Simulation of Gegenbauer Processes using Wavelet Packets.* Queensland University of Tech., School of Economics and Finance.

[83] Conley, T. and Ligon, E. (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth,* **7**, 157–187.

[84] Crambes, C., Kneip, A. and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *Annals of Statistics,* **37**, 35–72.

[85] Cuesta-Albertos, J. A. and Fraiman, R. (2007). Impartial trimmed k-means for functional data. *Computational Statistics & Data Analysis,* **51**, 4864—4877.

[86] Da Prato, G. and Zabczyk, J. (2002). *Second Order Partial Differential Equations in Hilbert spaces.* Cambridge University Press, New York.

[87] Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics,* **13**, (7): 759-–774.

[88] Damon, J. and Guillas, S. (2005). Estimation and simulation of autoregressive hilbertian processes with exogenous variables. *Statistical Inference for Stochastic Processes,* **8**, (2): 185—204.

[89] Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

[90] Dautray, R. and Lions, J. L. (1985). *Mathematical Analysis and Numerical Methods for Science and Technology.* Volume 3: Spectral Theory and Applications. Springer ISBN 978-3-540-66099-6 (2nd printing, 2000, X, 542 p).

[91] De Jong, A., Kabir, R. and Nguyen, T. (2008). Capital structure around the world: The roles of firms and country specific determinants. *Journal of Banking & Finance,* **32**, 1954–1969.

[92] Debarsy, N. and Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics,* **40**, 453—470.

[93] Debarsy, N., Ertur, C. and LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology,* **9**, 158—171.

[94] Degryse, H., De Goeij, P. and Kappert, K. (2012). The impact of firm and industry characteristics on small firms capital structure. *Small Business Economics,* **38**, (4): 431–447.

[95] Delaigle, A. and Hall, P. (2012a). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society. Series B,* **74**, 267–286.

[96] Delaigle, A. and Hall, P. (2012b). Methodology and theory for partial least squares applied to functional data. *Annals of Statistics,* **40**, 322–352.

[97] Delaigle, A., Hall, P. and Bathia, N. (2012). Componentwise classification and clustering of functional data. *Biometrika,* **99**, 299–313.

[98] Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics,* **21**, (3): 224—239.

[99] Delfiner, P. and Delhomme, J. P. (1975). *Optimum interpolation by kriging.* In: J. C. Davis, M. J. McCullagh (Eds.). *Display and Analysis of Spatial Data, NATO Advanced Studies,* 96–114.

[100] Donoho, D. L. (1993). Unconditional bases and bit-level compression. *Applied and Computational Harmonic Analysis,* 3388–3392.

[101] Dorogovtsev, A. (1989). *Elements of the General Theory of Measure and Integral.* Vyshcha shkola.

[102] Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics,* **80**, 549–560.

[103] Dunford, N. and Schwartz, J. T. (1963). *Linear Operators.* Part II: Spectral Theory. Interscience, NewYork.

[104] Durbin, J. and Koopman, S. J. (2001). *Time series analysis by state space methods.* Oxford University Press.

[105] Earls, C. and Hooker, G. (2014). Bayesian covariance estimation and inference in latent Gaussian process models. *Statistical Methodology,* **18**, 79-–100.

[106] Elhorst, J. P. (2010). *Spatial Panel Data Models.* In: M. M. Fischer, A. Getis (Eds.). *Handbook of Applied Spatial Analysis.* Springer Berlin Heidelberg, Chapter **2**, 377–407.

[107] Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics,* **16**, 365–384.

[108] Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2007). Functional PLS logit regression model. *Computational Statistics & Data Analysis,* **51**, 4891–4902.

[109] Espejo, R. M., Leonenko, N. N. and Ruiz-Medina, M. D. (2014). Gegenbauer random fields. *Random Operators and Stochastic Equations,* **22**, (1): 1—16. DOI: 10.1515/rose-2014-0001.

[110] Espejo, R. M., Leonenko, N. N., Olenko, A. and Ruiz-Medina, M. D. (2014). On a class of minimum contrast estimators for Gegenbauer random fields. *TEST.* (Submitted).

[111] Fan, J. and Zhang, J. T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journals of the Royal Statistical Society. Series B,* **62**, 303–322.

[112] Fernández-Pascual, R., Ruiz-Medina,M. D. and Angulo, J. M. (2006). Estimation of intrinsic processes affected by additive fractal noise. *Journal of Multivariate Analysis,* **97**, 1361–1381.

[113] Ferrara, L. and Guegan, D. (2001). *Comparison of Parameter Estimation Methods in Cyclical Long Memory Time Series.* In: C. Junis, J. Moody, A. Timmermann (Eds.). *Development in Forecast Combination and Portfolio Choice,* Wiley, New York, Chapter **8**.

[114] Ferraty, F., Kudraszow, N. and Vieu, P. (2012). Nonparametric estimation of a surrogate density function in infinite dimensional spaces. *Journal Nonparametric Statistics,* **24**, 447–464.

[115] Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis,* **44**, 161—173.

[116] Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination. *Journal Nonparametric Statistics,* **16**, 111–125.

[117] Ferraty, F. and Vieu, P. (2006). *Nonparameric Functional Data Analysis*. Springer, New York.

[118] Ferraty, F. and Vieu, P. (2009). Additive prediction and boosting for functional data. *Computational Statistics & Data Analysis,* **53**, 1400—1413.

[119] Friedman, N. A. (1970). *Introduction to Ergodic Theory.* Van Nostrand Reinhold mathematical studies.

[120] Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and Cp in multivariate linear regression. *Biometrika,* **84**, 707—716.

[121] Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics,* **164**, (1): 142—157.

[122] Giacofci, M., Lambert-Lacroix, S., Marot, G. and Picard, F. (2013). Wavelet-based clustering for mixed-effects functional models in high dimensions. *Biometrics,* **69**, (1): 31–40.

[123] Giraitis, L., Hidalgo, J. and Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with uknown pole. *The Annals of Statistics,* **29**, 987–1023.

[124] Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models *Journal of Econometrics,* **179**, 46—65.

[125] Giraitis, L. and Leipus, R. (1995). A generalized fractionally differencing approach in long memory modeling. *Lietuvos Matematikos Rinkinys,* **35**, 65–81.

[126] 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**, (1): 66—82.

[127] Giraldo, R., Delicado, P. and Mateu, J. (2012). Hierarchical clustering of spatially correlated functional data. *Statistica Neerlandica,* **66**, 403-–421.

[128] Goicoa, T., Ugarte, M. D., Etxeberria, J. and Militno, A. F. (2012). Comparing CAR and P-spline models in spatial disease mapping. *Environmental and Ecological Statistics,* **19**, (4): 573–599.

[129] Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products.* Elsevier.

[130] Granger, W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis,* **1**, 15–29.

[131] Gray, H. L., Zhang, N. F. and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis,* **10**, 233–257.

[132] Guillas, S. and Lai, M. J. (2010). Bivariate splines for spatial functional regression models. *Journal of Nonparametric Statistics,* **22**, (4): 477–497.

[133] Guo, W. (2002). Inference in smoothing spline analysis of variance. *Journals of the Royal Statistical Society. Series B,* **64**, 887–898.

[134] Guo, H., Lim, C. Y. and Meerschaert, M. M. (2009). Local Whittle estimator for anisotropic random field. *Journal of Multivariate Analysis,* **100**, 993–1028.

[135] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika,* **69**, 95–105.

[136] Guyon, X. (1995). *Random fields on a network.* Springer-Verlag.

[137] Hall, P., Poskitt, D. and Presnell, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics,* **43**, 1–9.

[138] Harten, A. (1993). Discrete multiresolution analysis and generalized wavelets. *Applied Numerical Mathematics,* **12**, 153–193.

[139] Hassler, U. (1994). (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis,* **15**, 19–30.

[140] Hautsch, N. and Klotz, S. (1999). *Estimating the neighborhood influence on decision makers: theory and an application on the analysis of innovation decisions.* University of Konstanz.

[141] He, X., Tong, X., Sun, J. and Cook, R. J. (2008). Regression analysis of multivariate panel count data *Biostatistics,* **9**, (2): 234—248.

[142] Heyde, C. C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications,* **45**, 169–182.

[143] Heyman, D., Deloof, M. and Ooghe, H. (2008). The financial structure of private held Belgian firms. *Small Business Economics,* **30**, 301–313.

[144] Honoré, B. E. (2002). Nonlinear models with panel data. *Portuguese Economic Journal,* **1**, (2): 163–179.

[145] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time varying coefficient models with longitudinal data. *Biometrika,* **85**, 809–822.

[146] Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics,* **38**, 1845–1884.

[147] Hörmann, S. and Kokoszka, P. (2011). *Functional Time Series.* Handbook of Statistics. **30**, 157—186.

[148] Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society,* **75**, 103–122.

[149] Horváth, L. and Reeder, R. (2012). Detecting changes in functional linear models. *Journal of Multivariate Analysis,* **111**, 310-–334.

[150] Hosking, J. R. M. (1981). Fractional differencing. *Biometrika,* **68**, 165–176.

[151] Hsu, N. J. and Tsai, H. (2009). Semiparametric estimation for seasonal long-memory time series using generalized exponential models. *Journal of Statistical Planning and Inference,* **139**, 1992–2009.

[152] Hu, X. J., Sun, J. and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Journal of Statistics,* **30**, 25—43.

[153] Huang, C. Y., Wang, M. C. and Zhang, Y. (2006) Analysing panel count data with informative observation times. *Biometrika,* **93**, (4): 763-–775.

[154] Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational. *Practice,* **75**, 181–189.

[155] Ivanov, A. V. (1997). *Asymptotic Theory of Nonlinear Regression.* Kluwer Academic Publishers, Dordrecht.

[156] Ivanov, A. V. and Leonenko, N. N. (2004). Asymptotic theory for non-linear regression with long-range dependence. *Mathematical Methods of Statistics,* **13**, 153–178.

[157] Ivanov, A. V., Leonenko, N. N., Ruiz-Medina, M. D. and Savich, I. N. (2013). Limit theorems for weighted non-linear transformations of Gaussian stationary processes with singular spectra. *Annals of Probability,* **4**, (2): 1088–1114.

[158] Jaffard, S. (1999). The multifractal nature of Lévy processes. *Probability Theory and Related Fields,* **114**, 207—227.

[159] Jain, A. K. (1981). Advances in mathematical models for image processing. *Proceedings of the IEEE,* **69**, 502–528.

[160] James, G. M. (2002). Generalized linear models with functional predictors. *Journals of the Royal Statistical Society. Series B,* **64**, 411–432.

[161] James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregular sampled curves. *Journals of the Royal Statistical Society. Series B,* **63**, 533–550.

[162] James, G. M. and Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association,* **98**, 397–408.

[163] James, G. M., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable. *Annals of Statistics,* **37**, 2083–2108.

[164] Kadri, H., Duflos, E., Preux, P., Canu, S. and Davy, M. (2010). *Nonlinear functional regression: a functional RKHS approach,* in AISTATSÍ0, JMLR: W&CP 9, Chia Laguna, Sardinia, Italy, 111–125.

[165] Kalbfleisch, J. D. and Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association,* **80**, (392): 863–871.

[166] Kato, T. (1995). *Perturbation Theory of Linear Operators.* Springer, Berlin.

[167] Kawano, S. and Konishi, S. (2009). Nonlinear logistic discrimination via regularized Gaussian basis expansions. *Communications in Statistics - Simulation and Computation,* **38**, (7): 1414—1425.

[168] Kelbert, M., Leonenko, N. N. and Ruiz-Medina, M. D. (2005). Fractional random fields associated with stochastic fractional heat equations. *Advances in Applied Probability,* **37**, 108–133.

[169] Kikuchi, K. and Negoro, A. (1995). Pseudo differential operators with variable order of differentiation. *Reports of the Faculty of Liberal Arts, Shizuoka University, Natural Sciences,* **31**, 19–27.

[170] Kumar, B. V. R. and Mehra, M. (2005). Wavelet based preconditioners for sparse linear systems. *Applied Mathematics and Computation,* **171**, 203–224.

[171] Laksaci, A., Rachdi, M. and Rahmani, S. (2013). Spatial modelization: Local linear estimation of the conditional distribution for functional data. *Spatial Statistics,* **6**, 1-–23.

[172] Laukaitis, A. (2007). An empirical study for the estimation of autoregressive hilbertian processes by wavelet packet method. *Nonlinear Analysis: Modelling and Control,* **12**, (1): 65—75.

[173] Lee, Y. J. (2014). Testing a linear dynamic panel data model against nonlinear alternatives. *Journal of Econometrics,* **178**, 146—166.

[174] Lee, W. S. and Kassim, A. A. (2007). Signal and image approximation using interval wavelet transform. *IEEE Transactions on Image Processing,* **16**, 46–56.

[175] Lee, L. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics,* **154**, 165–185.

[176] Leonenko, N. N. (1999). *Limit Theorems for Random Fields with Singular Spectrum.* Mathematics and its Applications 465. Kluwer Academic Publishers, Dordrecht, Boston, London.

[177] Leonenko, N. N. and Ruiz-Medina, M. D. (2006). Scaling laws for the multidimensional Burgers equation with quadratic external potential. *Journal of Statistical Physicscs,* **124**, 191–205.

[178] Leonenko, N. N., Ruiz-Medina, M. D. and Taqqu, M. (2011). Fractional elliptic, hyperbolic and parabolic random fields. *Electronic Journal of Probability,* **16**, 1134–1172.

[179] Leonenko, N. N. and Sakhno, L. M. (2006). On the Whittle estimators for some classes of continuous-parameter random processes and fields. *Statistics & Probability Letters,* **76**, 781–795.

[180] Leopold, H. G. (1989). On Besov spaces of variable order of differentiation. *Zeitschrift für Analysis und ihre Anwendungen,* **8**, 69–82.

[181] Leopold, H. G. (1991). On function spaces of variable order of differentiation. *Forum Mathematicum,* **3**, 1–21.

[182] Leopold, H. G. (1999). Embedding of function spaces of variable order of differentation in function spaces of variable order of integration. *Czechoslovak Mathematical Journal,* **49**, 633–644.

[183] Leng, X. and Müller, H. G. (2006). Time ordering of gene coexpression. *Biostatistics,* **7**, 569–584.

[184] Li, D., Chen, J. and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal,* **14**, 387—408.

[185] Li, P. L. and Chiou, J. M. (2011). Identifying cluster number for subspace projected functional data clustering original. *Computational Statistics & Data Analysis,* **55**, 2090—2103.

[186] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics,* **38**, (6): 3321-–3351.

[187] Li, W. K. and McLeod, A. I. (1986). Fractional time series modelling. *Miscellanea,* **73**, 217–221.

[188] Li, B. and Yu, Q. (2008). Classification of functional data: a segmentation approach. *Computational Statistics & Data Analysis,* **52**, 4790—4800.

[189] Lian, H. (2012). Empirical likelihood confidence intervals for nonparametric functional data analysis. *Journal of Statistical Planning and Inference,* **142**, 1669—1677.

[190] Lin, Z., Li. Q. and Sun, Y. (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics*, **178**, 167—179.

[191] Liu, X. L. and Müller, H. G. (2003). Modes and clustering for time-warped gene expression profile data. *Bioinformatics,* **19**, 1937–1944.

[192] López-Pintado, S. and Romo, J. (2006). *Depth-based classification for functional data.* In: R. Liu, R. Serfling, D. L. Souvaine, (Eds.). *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications.* DIMACS Series **72**, 103—121.

[193] Ma, P., Zhong, W., Feng, Y. and Liu, J. S. (2008). Bayesian functional data clustering for TemporalMicroarray data. *Hindawi Publishing Corporation International Journal of Plant Genomics.*

[194] Mahía, R. (2000). Análisis de estacionariedad con datos de panel: una ilustración para los tipos de cambio, precios y mantenimiento de la ppa en latinoamérica. *Mimeo,* Instituto L. R. Klein.

[195] Mandelbrot, B. and Van Ness, J. W. (1968). Fractional Brownian motion, fractional noises and applications. *Society for Industrial and Applied Mathematics,* **10**, 422—437.

[196] Maravilla, D., Mendozaa, B. and Jáureguib, E. (2008). Solar signals in the minimum extreme temperature records in the southern region of the Gulf of Mexico. *Advances in Space Research,* **42**, 1593–1600.

[197] Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis,* **98**, 1231–1261.

[198] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications,* **115**, 155–177.

[199] Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika,* **66**, 209–217.

[200] Martin, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statististics-Theory & Methods,* **19**, 55–81.

[201] Martin, R. J. (1996). Some results on unilateral $ARMA$ laticce processes. *Journal of Statistical Planning And Inference,* **50**, 395–411.

[202] Mayda, A. M. (2008). International migration: A panel data analysis of the determinants of bilateral flows. *CReAM Discussion Paper Series 0707, Centre for Research and Analysis of Migration (CReAM),* Department of Economics, University College London.

[203] McElroy, T. S. and Holan, S. H. (2012). On the computation of autocovariances for generalized Gegenbauer processes. *Statistica Sinica,* **22**, 1661–1687.

[204] McKitrick, R., McIntyre, S. and Herman, C. (2010). Panel and multivariate methods for tests of trend equivalence in climate data series. *Atmospheric Science Letters,* **11**, (4): 270-–277.

[205] Meintanis, S. G. (2011). Testing for normality with panel data. *Journal of Statistical Computation and Simulation*, **81**, (11): 1745–1752.

[206] Mendelson, S. (2002). Learnability in Hilbert spaces with reproducing kernels. *Journal of Complexity,* **18**, 152–170.

[207] Meyer, Y. (1991). Ondelettes sur l'intervalle. *Revista Matemática Iberoamericana,* **7**, 115—133.

[208] Michaelas, N., Chittenden, F. and Poutzioris, P. (1999). Financial policy and capital structure choice in U.K. SMEs: empirical evidence from company panel data. *Small Business Economics,* **12**, 113–130.

[209] Monestiez, P. and Nerini, D. (2008). A cokriging method for spatial functional data with applications in oceanology. *Functional and Operational Statistics. Contributions to Statistics,* Chapter **36**, 237–242.

[210] Mourid, T. (2002). Estimation and prediction of functional autoregressive processes. *Statistics,* **36**, (2): 125-–138.

[211] Müller, H. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association,* **82**, (397): 231–238.

[212] Müller, H. G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics,* **33**, 774–805.

[213] Nakai, M. and Ke, W. (2011). Review of methods for handling missing data in longitudinal data analysis. *International Journal of Mathematical Analysis,* **5**, 1–13.

[214] Nayaran, P. K. and Narayan, S. (2010). Carbon dioxide emissions and economic growth: Panel data evidence from developing countries. *Energy Policy*, **38**, 661—666.

[215] Nerini, D., Monestiez, P. and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis,* **101**, 409–418.

[216] Nerini, D. and Ghattas, B. (2007). Classifying densities using functional regression trees: applications in oceanology. *Computational Statistics & Data Analysis,* **51**, 4984-–4993.

[217] Nguyen, P., Brown, P. E. and Stafford, J. (2012). Mapping cancer risk in southwestern Ontario with changing census boundaries. *Biometrics,* **68**, (4): 1228–1237.

[218] Nilsson, J. (2001). Spatial reorganization of SST anomalies by stationary atmospheric waves. *Dynamics of Atmospheres and Oceans,* **34**, 1–21.

[219] Nualart, D. and Sanz-Solé, M. (1979). A Markov property for two-parameter Gaussian processes. *Stochastica,* **3**, (1): 1—16.

[220] Olenko, A. (2013). Limit theorems for weighted functionals of cyclical long-range dependent random fields. *Stochastic Analysis and Applications,* **31**, (2): 199–213.

[221] Olokoyo, F. O. (2013). Capital structure and corporate performance of nigerian quoted firms: a panel data approach. *African Development Review-Revue Africaine De Developpement,* **25**, (3): 358–369.

[222] Papari, G. and Petkov, N. (2009). Reduced Inverse Distance Weighting Interpolation for Painterly Rendering. in: *International Conference on Computer Analysis of Images and Pat-*

terns – Proceedings CAIP 2009. Lecture Notes in Computer Science 5702, 509–516, Springer, Berlin.

[223] Parent, O. and LeSage, J. P. (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B,* **44**, 633-–645.

[224] Parent, O. and LeSage, J. P. (2011). A space-time filter for panel data models containing random effects. *Computational Statistics & Data Analysis,* **55**, 475–490.

[225] Preda, C. and Saporta, G. (2005a). Clusterwise PLS regression on a stochastic process. *Computational Statistics & Data Analysis,* **49**, 991–2008.

[226] Preda, C. and Saporta, G. (2005b). PLS regression on a stochastic process. *Computational Statistics & Data Analysis,* **48**, 149–158.

[227] Psillaki, M. and Daskalakis, N. (2009). Are the determinants of capital structure country or firm specific?. *Small Business Economics,* **33**, 319–333.

[228] Qian, J. and Wang, L. (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics,* **167**, 483-–493.

[229] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *Journal of Statistical Planning and Inference,* **137**, 2784–2801.

[230] Ramm, A. G. (2005). *Random Fields Estimation.* World Scientific Publishing Co., New York.

[231] Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis.* Springer, New York.

[232] Reisen, V., Rodrigues, A. L. and Palma, W. (2006). Estimation of seasonal fractionally integrated processes. *Computational Statistics & Data Analysis,* **50**, 568–582.

[233] Rincón, M. and Ruiz-Medina, M. D. (2012a). Local wavelet-vaguelette-based functional classification of gene expression data. *Biometrical Journal,* **54**, 75–93.

[234] Rincón, M. and Ruiz-Medina, M. D. (2012b). Wavelet-RKHS-based functional statistical classification. *Advances in Data Analysis and Classification,* **6**, (3): 201—217.

[235] Robinson, P. M. and Sanz, J. V. (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis,* **97**, 1090–1120.

[236] Rocca, M., Rocca, T. and Cariola, A. (2010). The influence of local institutional differences on the capital structure of SMEs: evidence from Italy. *International Small Business Journal,* **28**, 234–257.

[237] Romano, E., Balzanella, A. and Verde, R. (2010). *Clustering Spatio-functional data: a model-based approach.* Studies in Classification, Data Analysis, and Knowledge Organization. Springer, New York, 167–175.

[238] Romano, E., Balzanella, A. and Verde, R. (2013). *A regionalization method for spatial functional data based on variogram models: an application on environmental data.* In: Advances in Theoretical and Applied Statistics. Springer-Verlag, Berlin Heidelberg, to appear.

[239] Romano, E., Giraldo, R. and Mateu, J. (2011). *Clustering spatially correlated functional data.* In: F. Ferraty (Ed.). *Recent Advances in Functional Data Analysis and Related Topics.* Springer-Verlag Berlin Heidelberg, Chapter **43**, 277–282.

[240] Romano, E. and Verde, R. (2011). *Clustering geostatistical functional data.* In: A. Di Ciaccio, M. Coli, J.M. Angulo (Eds.). *Advanced Statistical Methods for the Analysis of Large Data-Sets, Series Studies in Theoretical and Applied Statistics.* Springer, Berlin, 23–31.

[241] Rosenblatt, M. (1985). *Stationary sequences and random fields.* Birkaüser.

[242] Rossi, F. and Villa, N. (2006). Support vector machine for functional data classification. *Neurocomputing,* **69**, 730—742.

[243] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journals of the Royal Statistical Society. Series B,* **71**, (Part 2): 319–392.

[244] Ruiz-Medina, M. D. (2011). Spatial autoregressive and moving average Hilbertian processes. *Journal of Multivariate Analysis,* **102**, (2): 292–305.

[245] Ruiz-Medina, M. D. (2012a). Spatial functional prediction from spatial autoregressive Hilbertian. *Environmetrics,* **23**, (1): 119–128.

[246] Ruiz-Medina, M. D. (2012b). New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics,* **1**, 82-–91.

[247] Ruiz-Medina, M. D., Angulo, J. M. and Anh, V. V. (2003). Fractional-order regularization and wavelet approximation to the inverse estimation problem for random fields. *Journal of Multivariate Analysis,* **85**, 192—216.

[248] Ruiz-Medina, M. D., Anh, V. V. and Angulo, J. M. (2004). Fractional generalized random fields of variable order. *tochastic Analysis and Applications,* **22**, 775–800.

[249] Ruiz-Medina, M. D., Anh, V. V., Espejo, R. M. and Frías, M. P. (2013). Heterogeneous spatial dynamical regression in a Hilbert-valued context. *Stochastic Analysis and Applications,* **31**, 509–527.

[250] Ruiz-Medina, M. D., Anh, V. V., Espejo, R. M., Angulo, J. M. and Frías, M. P. (2013). Least-squares estimation of multifractional random fields in a Hilbert-valued context. *Journal of Optimization. Theory and Applications,* DOI: 10.1007/s10957-013-0423-4.

[251] Ruiz-Medina, M. D. and Espejo, R. M. (2011). Incorporating Spatial Interaction between Large Dimensional Temperature Series in Atmosphere -Ocean Modelling of Global Climate Change. *Procedia Environmental Sciences,* **7**, 2–7.

[252] Ruiz-Medina, M. D. and Espejo, R. M. (2012). Spatial autoregressive functional plug-in prediction of ocean surface temperature. *Stochastic Environmental Research and Risk Assessment,* **26**, 335–344. DOI: 10.1007/s00477-012-0559-z.

[253] Ruiz-Medina, M. D. and Espejo, R. M. (2013a). Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming. *International Journal of Applied Earth Observation and Geoinformation,* **22**, (1): 27–39. DOI: 10.1016/j.jag.2012.01.021.

[254] Ruiz-Medina, M. D. and Espejo, R. M. (2013b). Maximum-Likelihood Asymptotic Inference for Autoregressive Hilbertian Processes. *Methodology and Computing Applied Probability.* DOI: 10.1007/s11009-013-9329-8.

[255] Ruiz-Medina, M. D., Espejo, R. M. and Romano, E. (2014). Spatial functional normal mixed effect approach for curve classification. *Advances in Data Analysis and Classification.* DOI: 10.1007/s11634-014-0174-6.

[256] Ruiz-Medina, M. D., Espejo, R. M., Ugarte, M. D. and Militino, A.F. (2013). Functional time series analysis of spatio-temporal epidemiological data. *Stochastic Environmental Research and Risk Assessment.* DOI: 10.1007/s00477-013-0794-y.

[257] Ruiz-Medina, M. D. and Fernández-Pascual, R. (2010). Spatiotemporal filtering from fractal spatial functional data sequences. *Stochastic Environmental Research and Risk Assessment,* **24**, 527–538.

[258] Ruiz-Medina, M. D. and Salmerón, R. (2010). Functional maximum-likelihood estimation of ARH(p) models. *Stochastic Environmental Research and Risk Assessment,* **24**, 131–146.

[259] Ruiz-Medina, M. D. and Salmerón, R. (2011). *Asymptotic properties of functional maximum-likelihood ARH parameter estimators.* Revue des Nouvelles Technologies de l'Information (Special issue of JSFdS Bordeaux 2009), 33–58.

[260] Ruiz-Medina, M. D., Salmerón, R. and Angulo, J. M. (2007). Kalman filtering from POP-based diagonalization of ARH(1). *Computational Statistics & Data Analysis,* **51**, 4994–5008.

[261] Salamh, M. (2011). Determinants of lodgepole pine growth: Static and dynamic panel data models. *Canadian journal of statistics-revue canadienne de statistique,* **39**, (2): 190–192.

[262] Salmerón, R. and Ruiz-Medina, M. D. (2009). Multispectral decomposition of functional autoregressive models. *Stochastic Environmental Research and Risk Assessment,* **23**, (3): 289–297.

[263] Salmerón, S. and Ruiz-Medina, M. D. (2011). Functional statistical time series analysis of the dividend policy of Spanish companies. *AESTIMATIO, the IEB international journal of finance,* **3**, 2–17.

[264] Schölkopf, B. and Smola, A. (2002). *Learning with Kernels.* MIT Press, Cambridge.

[265] Schweinberger, M. (2012). Statistical modelling of network panel data: Goodness of fit. *British Journal of Mathematical and Statistical Psychology,* **65**, 263-–281.

[266] Secchi, P., Vantini, S. and Vitelli, V. (2011). A clustering algorithm for spatially dependent functional data. *Procedia Environmental Sciences,* **7**, 176—181.

[267] Sentürk, D. and Müller, H. G. (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association,* **105**, 1256–1264.

[268] Shang, H. L. (2013). Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density. *Computational Statistics & Data Analysis,* **67**, 185—198.

[269] Sogorb-Mira, F. (2005). How SME uniqueness affects capital structure: Evidence from a 1994 – 1998 Spanish data panel. *Small Business Economics,* **25**, 447–457.

[270] Soltani, A. R. and Hashemi, M. (2011). Periodically correlated autoregressive Hilbertian processes. *Statistical Inference for Stochastic Processes,* **14**, (1): 177–188.

[271] Spitzner, D. J., Marron, J. S. and Essick, G. K. (2003). Mixed-model functional ANOVA for studying human tactile perception. *Journal of the American Statistical Association,* **98**, 263–272.

[272] Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

[273] Stein, M. L. (2009). Spatial interpolation of high-frequency monitoring data. *Annals of Applied Statistics,* **3**, 272–291.

[274] Su, L. and Lu, X. (2013). Nonparametric dynamic panel data models: Kernel estimation and specification testing. *Journal of Econometrics,* **176**, 112-–133.

[275] Sun, B. and Sutradhar, B. C. (2013). GQL estimation in linear dynamic models for panel data. *Journal of Statistical Computation and Simulation,* **83**, (3): 568–580.

[276] Tandeo, P., Ailliot, P. and Autret, E. (2011). Linear Gaussian state-space model with irregular sampling: application to sea surface temperature. *Stochastic Environmental Research and Risk Assessment,* **25**, (6): 793–804.

[277] Taniguchi, M. (1987). Minimum Contrast Estimation for Spectral Densities of Stationary Processes. *Journal of the Royal Statistical Society. Series B (Methodological),* **49**, (3): 315–325.

[278] Tarpey, T. and Kinateder, K. K. J. (2003). Clustering functional data. *Journal of Classification,* **20**, 93—114.

[279] Tjostheim, D. (1978). Statistical spatial series modeling. *Advances in Applied Probability,* **10**, 130–154.

[280] Tjostheim, D. (1981). Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Transactions on Geoscience and Remote Sensing,* **19**, 15–24.

[281] Tjostheim, D. (1983). Statistical spatial series modeling II: Some further results on unilateral processes. *Advances in Applied Probability,* **15**, 562–584.

[282] Triebel, H. (1978). *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland Publishing Co, Amsterdam.

[283] Utrero-González, N. (2007). Banking regulation, institutional framework and capital structure: International evidence from industry data. *The Quarterly Review of Economics and Finance,* **4**, 481–506.

[284] Ugarte, M. D., Goicoa, T. and Militino, A. F. (2009a). Empirical Bayes and fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics & Data Analysis,* **53**, (8): 2938–2949.

[285] Ugarte, M. D., Goicoa, T. and Militino, A. F. (2010). Spatio-temporal modelling of mortality risks using penalized splines. *Environmetrics,* **21**, 270–289.

[286] Ugarte, M. D., Goicoa, T., Etxeberria, J. and Militino, A. F. (2012). A P-spline ANOVA type model in space-time disease mapping. *Stochastic Environmental Research and Risk Assessment,* **26**, (6): 835–845.

[287] Ugarte, M. D., Goicoa, T., Ibáñez, B. and Militino, A. F. (2009b). Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics,* **20**, (6): 647–665.

[288] Van Der Wijst, N. and Thurik, R. (1993). Thurik, Determinants of small firm debt ratios: an analysis of retail panel data. *Small Business Economics,* **5**, 55–65.

[289] Vakhania, N. N., Tarieladze, V. I. and Chebonyan, S. A. (1987). *Probability Distributions in Banach Spaces.* D. Reidel Publishing Company, Dordrecht.

[290] Verardi, V. and Wagner, J. (2010). Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. *Jahrbücher für Nationalökonomie und Statistik,* **231**, (4): 546–557.

[291] Vidal-Sanz, J. M. (2009). Automatic spectral density estimation for random fields on a lattice via bootstrap. *Test,* **18**, (1): 96–114.

[292] Walk, H. (1977). An invariance principle for the Robbins-Monro process in a Hilbert space. *Z. Wahrscheinlichkeitstheorie und verw, Gebiete,* **39**, 135–150.

[293] Wang, S. S., Zhou, D. Q., Zhou, P. and Wang, Q. W. (2011). CO_2 emissions, energy consumption and economic growth in China: A panel data analysis. *Energy Policy,* **39**, 4870—4875.

[294] WeiLin, X., WeiGuo, Z. and XiLi, Z. (2012). Minimum contrast estimator for fractional Ornstein-Uhlenbeck processes. *Science China Mathematics,* **55**, (7): 1497–1511.

[295] Wikle, C. K. and Cressie, N. (1999). A dimension-reducted approach to space-time Kalman filtering. *Biometrika,* **86**, (4): 815–829.

[296] Woodward, W. A., Cheng, Q. C. and Gray, H. L. (1998). A k-factor $GARMA$ long-memory model. *Journal of Time Series Analysis,* **19**, 485–504.

[297] Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Hoover, Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association,* **93**, 1388–1402.

[298] Wu, S. W., Deng, F. R., Niu, J., Huang, Q. S., Liu, Y. C. and Guo, X. B. (2011). The relationship between traffic-related air pollutants and cardiac autonomic function in a panel of healthy adults: a further analysis with existing data. *Inhalation Toxicology,* **23**, (5): 289–303.

[299] Wu, Y., Fan, J. and Müller, H. G. (2010). Varying-coefficient functional linear regression. *Bernoulli,* **16**, 730–758.

[300] Wu, J. H. and Li, G. D. (2014). Moment-based tests for individual and time effects in panel data models. *Journal of econometrics,* **178**, 569–581.

[301] Wu, J. and Su, W. (2010). Estimation of moments for linear panel data models with potential existence of time effects. *Statistics & Probability Letters,* **80**, 1933—1939.

[302] Wu, H. and Zhang, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis.* John Wiley & Sons, New York.

[303] Wu, J. H. and Zhu, L. X. (2012). Estimation of and testing for random effects in dynamic panel data models. *TEST,* **21**, (3): 477–497.

[304] Yang, J. Y., Peng, Z. L., Yu, Z., Zhang, R. J., Anh, V. and Wang, D. (2009a). Prediction of protein structural classes by recurrence quantification analysis based on chaos game representation. *Journal of Theoretical Biology,* **257**, 618–626.

[305] Yang, J. Y., Yu, Z. and Anh, V. (2009b). Clustering structures of large proteins using multifractal analyses based on a 6-letter model and hydrophobicity scale of amino acids. *Chaos, Solitons and Fractals,* **40**, 607–620.

[306] Yao, Q. and Brockwell, P. J. (2006). Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli,* **12**, (3): 403–429.

[307] You, J. and Zhou, X. (2006). Statistical inference in a panel data semiparametric regression model with serially correlated errors. *Journal of Multivariate Analysis,* **97**, 844—873.

[308] You, J., Zhou, X. and Zhou, Y. (2010). Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors. *Journal of Multivariate Analysis,* **101**, 1079–1101.

[309] Yu, H. L. and Christakos, G. (2011). Modeling and estimation of heterogeneous spatiotemporal attributes under conditions of uncertainty. *IEEE Transactions on Geoscience and Remote Sensing,* **49**, (1): 366–376.

[310] Yu, J., Jong, R. and Lee, L. (2012). Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics,* **167**, 16—37.

[311] Zhang, J., Feng, S., Li, G. and Lian, H. (2011a). Empirical likelihood inference for partially linear panel data models with fixed effects. *Economics Letters,* **113**, 165—167.

[312] Zhang, Z. and Müller, H. G. (2011b). Functional density synchronization. *Computational Statistics & Data Analysis,* **55**, 2234—2249.

[313] Zhao, X. and Tong, X. (2011). Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics & Data Analysis,* **55**, 291–300.

[314] Zoglat, A. (2008). Functional analysis of variance. *Applied Mathematical Sciences,* **2**, 1115–1129.

Appendix A9*

Papers submitted.
On a class of minimum contrast estimators for Gegenbauer random fields

Espejo, R. M., Leonenko, N, Olenko, A. and Ruiz-Medina, M. D. (2014).

On a class of minimum contrast estimators for Gegenbauer random fields. *TEST,* (Submitted).

Abstract

The article introduces spatial long-range dependent models based on the fractional difference operators associated with the Gegenbauer polynomials. The results on consistency and asymptotic normality of a class of minimum contrast estimators of long-memory parameters of the models are obtained. A methodology to verify assumptions for consistency and asymptotic normality of minimum contrast estimators is developed. Numerical results are presented to confirm the theoretical findings.

A9*.1 Introduction

Among the extensive literature on long-range dependence, relatively few publications are devoted to cyclical long-memory processes or long-range dependent random fields. However, models with singularities at non-zero frequencies are of great importance in applications. For example, many time series show cyclical/seasonal evolutions. Singularities at non-zero frequencies produces peaks in the spectral density whose locations define periods of the cycles. A survey of some recent asymptotic results for cyclical long-range dependent random processes and fields can be found in Ivanov *et al.* (2013) and Olenko (2013).

In image analysis popular isotropic spatial processes with singularities of the spectral density at non-zero frequencies are wave, J-Bessel, and Gegenbauer models. Espejo, Leonenko, and Ruiz-Medina (2014) investigated probabilistic properties of spatial Gegenbauer models. A realization of the Gegenbauer random field on 100×100 grid is shown in Figure A9*.1.

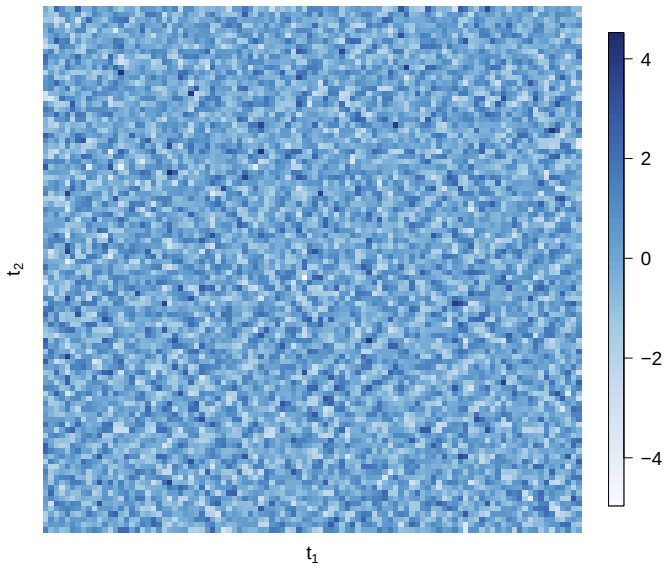


Figure A9*.1: Simulated realization of the Gegenbauer random field.

This article studies minimum contrast estimators (MCEs) of parameters of the Gegenbauer random fields. The MCE methodology has been widely applied in different statistical frameworks (see, for example, Anh, Leonenko, and Sakhno, 2004, 2007; WeiLin, WeiGuo, and XiLi, 2012). One of the first works which used a MCE methodology for the parameter estimation of spectral densities of stationary processes was the paper by Taniguchi (1987). Guyon (1995) introduced a class of MCEs for random fields. Anh, Leonenko, and Sakhno (2004) derived consistency and asymptotic normality of a class of MCEs for stationary processes within the class of fractional Riesz-Bessel motion (see Anh, Angulo, and Ruiz-Medina, 1999). Results based on the second and third-order cumulant spectra were given by Anh, Leonenko, and Sakhno (2007). They also provided asymptotic properties of second and third-order sample spectral functionals. These properties are of independent interest, since they can be applied to study the limiting properties of nonparametric estimators of processes with short or long-range dependence. WeiLin, WeiGuo, and XiLi (2012) applied the minimum contrast parameter estimation to approximate the drift parameter of the Ornstein-Uhlenbeck process, when the corresponding stochastic differential equation is driven by the fractional Brownian motion with a specific Hurst index.

A burgeoning literature on spatio-temporal estimation has emerged in recent decades (see Beran, Ghosh, and Schell 2009; Chan and Tsai 2012; Giraitis, Hidalgo, and Robinson, 2001; Guo, Lim, and Meerschaert 2009; Li and Mcleod, 1986; Reisen, Rodrigues, and Palma, 2006, among others). One of the most popular estimation tools applied was the maximum likelihood estimation method (MLE). Reisen, Rodrigues, and Palma (2006) addressed the problem of parameter estimation of fractionally integrated processes with seasonal components. In order to estimate the fractional parameters, they propose several log-periodogram regression estimators with different bandwidths selected around and/or between the seasonal frequencies. The same methodology was used by Li and McLeod (1986) for fractionally differenced autoregressive-moving average processes in the stationary time series context. Several contributions have also

been made for MLE of long memory spatial processes (see, for example, Anh and Lunney, 1995). For two-dimensional spatial data the paper by Basu and Reinsel (1993) introduced a spatial unilateral first-order autoregressive moving average (ARMA) model. To implement MLE they provided a proper treatment to border cell values with a substantial effect in estimation of parameters. Beran, Ghosh and Schell (2009) addressed the problem of the least-squares estimation of autoregressive fractionally integrated moving-average (FARIMA) processes with long-memory. Cohen and Francos (2002) investigated asymptotic properties of least-squares estimators in regression models for two-dimensional random fields. Maximization of the Whittle likelihood has been also considered in the recent literature on the MCE (see for example, Chan and Tsai, 2012, Boissy *et al.*, 2005, Leonenko and Sakhno 2006). Leonenko and Sakhno (2006) gave a continuous version of the Whittle contrast functional supplied with a specific weight function for the estimation of continuous-parameter stochastic processes, deriving the consistency and asymptotic normality of such estimators. Guo, Lim, and Meerschaert (2009) demonstrated that the Whittle maximum likelihood estimator is consistent and asymptotically normal for stationary seasonal autoregressive fractionally integrated moving-average (SARFIMA) processes.

Parameter estimation of stationary Gegenbauer random processes was considered by numerous authors, see, for example, Gray, Zhang and Woodward, (1989); Chung (1996a,b); Woodward, Cheng, and Gray (1998); Collet and Fadili (2006); McElroy and Holan (2012). Gray, Zhang and Woodward (1989) used the generating function of the Gegenbauer polynomials to develop long memory Gegenbauer autoregressive moving-average (GARMA) models that generalize the FARIMA process. GARMA models were estimated by applying the MLE methodology. Chung (1996a) also applied this methodology with slight modifications based on the conditional sum of squares method. Chung (1996b) extended these results to the two-parameter context within the GARMA process class. Woodward, Cheng and Gray (1998) introduced a k-factor extension of the GARMA model that allowed to associate the long-memory behavior with each one of the

k Gegenbauer frequencies involved.

In this article we restrict our consideration to the estimation of long memory parameters. It is motivated in part by cyclic processes, for which pole locations are known. Also in some applications the spectral density singularity location can be estimated in advance. Various methods, including semiparametric, wavelet, and pseudo-maximum likelihood techniques, of the estimation of the spectral density singularity location were discussed by, for example, Arteche and Robinson (2000), Giraitis, Hidalgo, and Robinson (2001), and Ferrara and Guegan (2001).

This paper introduces and studies the MCE of parameters of spatial Gegenbauer processes. Specifically, analogous of continuous-space results by Anh, Leonenko, and Sakhno (2004) are formulated for random fields defined on integer grids. The consistency and asymptotic normality of the MCE are obtained using a spatial discrete version of the Ibragimov contrast function. The article develops a methodology to practically verify general theoretical assumptions for consistency and asymptotic normality of MCEs for specific models. The results provide a rigorous platform to conduct model selection and statistical inference.

The outline of the article is the following. In section A9*.2, we start by introducing the main notations of the paper. Some fundamental definitions and the main results of this article, Theorem 15 and 16, are given in A9*.3. Section A9*.4 consists of the proofs of the main results. Section A9*.5 presents simulation studies which support the theoretical findings. The Appendix provides auxiliary materials that specify for our case the conditions that ensure the consistency and asymptotic normality of the MCE based on the Ibragimov contrast function formulated in Anh, Leonenko, and Sakhno (2004).

In what follows we use the symbol C to denote constants which are not important for our discussion. Moreover, the same symbol C may be used for different constants appearing in the same proof.

All calculations in the article were performed using the software *R* version 3.0.2 and *Maple*

16, Maplesoft.

A9*.2 Gegenbauer random fields

This section introduces some of the main definitions of the Gegenbauer random fields given in Espejo, Leonenko, and Ruiz-Medina, (2014) (see also, Chung, 1996a,b; Gray, Zhang, and Woodward, 1989; and Woodward, Cheng, and Gray, 1998, for the temporal case).

Let Y_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$, be a random field defined on the grid lattice $\mathbb{Z}^2$. Consider the fractional difference operator ∇_u^d defined by

$$\nabla_u^d = (I - 2uB + B^2)^d = (1 - 2\cos\nu B + B^2)^d = [(1 - e^{i\nu}B)(1 - e^{-i\nu}B)]^d, \tag{A9*}$$

where B is the backward-shift operator, $u = \cos\nu$, i.e. $\nu = \arccos(u)$, $|u| \leq 1$, and $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Assume that Y satisfies the following state equation

$$\nabla_{u_1}^{d_1} \circ \nabla_{u_2}^{d_2} Y_{t_1,t_2} = \left(I - 2u_1B_1 + B_1^2\right)^{d_1} \circ \left(I - 2u_2B_2 + B_2^2\right)^{d_2} Y_{t_1,t_2} = \varepsilon_{t_1,t_2}, \quad (t_1,t_2) \in \mathbb{Z}^2, \tag{A9*}$$

where $\nabla_{u_i}^{d_i}$, $i = 1,2$, is given by equation (A9*), with B_i, $i = 1,2$, denoting the backward-shift operator for each spatial coordinate, i.e. $B_1 Y_{t_1,t_2} = Y_{t_1-1,t_2}$, and $B_2 Y_{t_1,t_2} = Y_{t_1,t_2-1}$. Here, ε_{t_1,t_2}, $(t_1,t_2) \in \mathbb{Z}^2$, is a zero-mean white noise field with the common variance $E[\varepsilon_{t_1,t_2}^2] = \sigma_\varepsilon^2$. The random field Y is called a spatial Gegenbauer white noise in Espejo, Leonenko, and Ruiz-Medina, (2014).

By equation (A9*) the Gegenbauer random field Y can be defined in terms of the inverse of the operator $\nabla_{u_1}^{d_1} \circ \nabla_{u_2}^{d_2}$ expanded in a Gegenbauer polynomial series as follows

$$\begin{aligned} Y_{t_1,t_2} &= \nabla_{u_2}^{-d_2} \circ \nabla_{u_1}^{-d_1} \varepsilon_{t_1,t_2} = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)\, C_{n_2}^{(d_2)}(u_2) B_1^{n_1} B_2^{n_2} \varepsilon_{t_1,t_2} \\ &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1)\, C_{n_2}^{(d_2)}(u_2) \varepsilon_{t_1-n_1,t_2-n_2}, \end{aligned} \tag{A9*}$$

where $d_i \neq 0$, $i = 1, 2$, and $C_n^{(d)}(u)$ is the Gegenbauer polynomial given by

$$C_n^{(d)}(u) = \sum_{k=0}^{[n/2]} (-1)^k \frac{(2u)^{n-2k}\Gamma(d-k+n)}{k!(n-2k)!\Gamma(d)}.$$

The Gegenbauer random field has the following property, see Espejo, Leonenko, and Ruiz-Medina, (2014).

Proposition 7. *If* $0 < d_i < \frac{1}{2}$ *and* $|u_i| < 1$, $i = 1, 2$, *then* Y *is a stationary invertible long range dependent random field.*

The spectral density of a stationary Gegenbauer random field is given by Chung, (1996a,b), and Hsu and Tsai, (2009), for the one-parameter case, and Espejo, Leonenko, and Ruiz-Medina, (2014), for the two-parameter case:

$$\begin{aligned} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \frac{\sigma_\varepsilon^2}{(2\pi)^2} \left|1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}\right|^{-2d_1} \left|1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}\right|^{-2d_2} \\ &= \frac{\sigma_\varepsilon^2}{(2\pi)^2} \left\{|2\cos\lambda_1 - 2u_1|\right\}^{-2d_1} \left\{|2\cos\lambda_2 - 2u_2|\right\}^{-2d_2}, \end{aligned} \tag{A9*}$$

where $\boldsymbol{\theta} = (\boldsymbol{u}, \boldsymbol{d}) = (u_1, u_2, d_1, d_2) \in \Theta = (-1, 1)^2 \times (0, 1/2)^2$, $u_i = \cos\nu_i$, and $-\pi \leq \lambda_i \leq \pi$, $i = 1, 2$. Using the spectral density function (A9*), one can compute the auto-covariance function of Y as follows:

$$\gamma(j_1, j_2, \boldsymbol{\theta}) = \frac{\sigma_\varepsilon^2}{4\pi} \prod_{i=1}^{2} \Gamma(1-2d_i)[2\sin(\nu_i)]^{\frac{1}{2}-2d_1} \left[P_{j_i-\frac{1}{2}}^{2d_i-\frac{1}{2}}(u_i) + (-1)^{j_i} P_{j_i-\frac{1}{2}}^{2d_i-\frac{1}{2}}(-u_i)\right],$$

where $P_a^b(z)$ is the associated Legendre function of the first kind, consult §8 in Abramowitz and Stegun, 1972.

From Chung (1996a,b), Gray, Zhang, and Woodward (1989) and Gradshteyn and Ryzhik (1980), the following asymptotic approximation of the autocovariance function can be obtained

$$\gamma(j_1, j_2, \boldsymbol{\theta}) = \prod_{i=1}^{2} \frac{2^{1-d_i}\sigma_\epsilon^2}{\pi} \sin^{-2d_i}(\nu_i) \sin(d_i\pi)\Gamma(1-2d_i)\cos(j_i\nu_i) \frac{\Gamma(j_i+2d_i)}{\Gamma(j_i+1)}[1 + \mathcal{O}(j_i^{-1})].$$

The random field Y is long range dependent as its auto-covariance function satisfies the condition $\sum_{(j_1,j_2)\in\mathbb{Z}^2} |\gamma(j_1, j_2, \boldsymbol{\theta})| = +\infty$.

Figure A9*.2 gives an example of the spectral density and the auto-covariance function of the Gegenbauer random field for the values of the parameters $u_1 = 0.4$, $u_2 = 0.3$, $d_1 = 0.2$, and $d_2 = 0.3$.

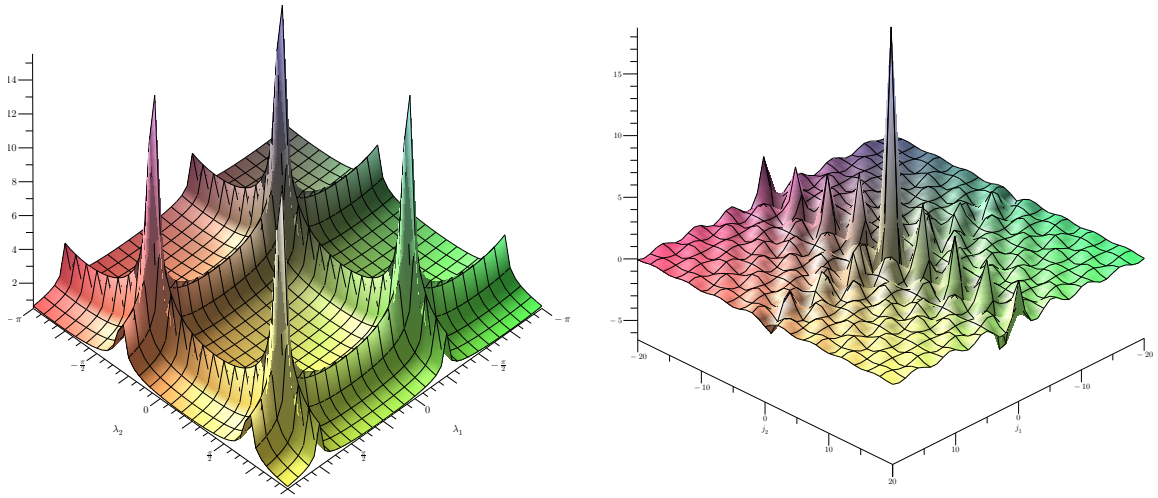


Figure A9*.2: Spectral density and auto-covariance function for $u_1 = 0.4$, $u_2 = 0.3$, $d_1 = 0.2$, and $d_2 = 0.3$

A9*.3 Asymptotic properties of MCEs

Suppose that the conditions imposed in Proposition 7 to ensure stationarity, invertibility and long-range dependence hold. Assume the value of the parameter $\boldsymbol{u} = (u_1, u_2)$ is known a priori or was estimated based on previous runs of the application. Then $\boldsymbol{\theta} = (\theta_1, \theta_2) = (d_1, d_2) \in \Theta = (0, 1/2)^2$ is the vector of parameters to estimate of the Gegenbauer random field defined by equation (A9*) (see also equation (A9*) for the corresponding spectral density).

Let Y_{t_1,t_2}, $t_1, t_2 = 0, ..., T$, be a part of a realization of the Gegenbauer random field.

Let $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, be a nonnegative function. Suppose the condition **A3** in the Appendix holds. We define

$$\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} \tag{A9*}$$

and consider the factorization

$$f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta}) \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}). \tag{A9*}$$

For all $\boldsymbol{\theta} \in \Theta$ the function $\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ has the following property

$$\int_{[-\pi,\pi]^2} \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = 1. \tag{A9*}$$

Let $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ be a non-random real-valued function, usually referred as the *contrast function*, given by

$$K(\boldsymbol{\theta_0}, \boldsymbol{\theta}) := \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \frac{\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta_0})}{\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})} \, d\boldsymbol{\lambda},$$

and let the *contrast field* be

$$U(\boldsymbol{\theta}) := - \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda},$$

where $\boldsymbol{\theta_0} = (d_{10}, d_{20})$ is the true parameter value. In what follows, P_0 denotes the probability distribution with the density function $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)$.

The empirical version $\hat{U}_T(\boldsymbol{\theta})$, $T \in \mathbb{Z}$, $\boldsymbol{\theta} \in \Theta$, is defined by

$$\hat{U}_T(\boldsymbol{\theta}) := - \int_{[-\pi,\pi]^2} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda}, \tag{A9*}$$

where $I_T(\boldsymbol{\lambda})$ is the periodogram of the observations Y_{t_1,t_2}, $t_1, t_2 = 0, ..., T$, of the Gegenbauer random field, that is,

$$I_T(\boldsymbol{\lambda}) := \frac{1}{(2\pi T)^2} \left| \sum_{t_1=0}^{T} \sum_{t_2=0}^{T} e^{-i(t_1\lambda_1 + t_2\lambda_2)} Y_{t_1,t_2} \right|^2.$$

The MCE is defined by the empirical contrast field $\hat{U}_T(\boldsymbol{\theta})$ and the contrast function $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ being $K(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \geq 0$, and having a unique minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

In particular, we choose

$$w(\boldsymbol{\lambda}) = |\lambda_1^2 - \nu_1^2|^{a_1}|\lambda_2^2 - \nu_2^2|^{a_2} w_0(\boldsymbol{\lambda}), \quad \boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [-\pi, \pi]^2, \tag{A9*}$$

where $a_i > 1$, $i = 1, 2$, $w_0(\boldsymbol{\lambda})$ is a non-negative function, which is not identically equal to zero and bounded on $[-\pi, \pi]^2$. The developed methodology is readily adjustable to other classes of weight functions.

Theorems 15 and 16 give consistency and asymptotic normality results for the MCE.

Theorem 15. *Let Y_{t_1,t_2}, $(t_1, t_2) \in \mathbb{Z}^2$, be a stationary Gegenbauer random field which spectral density satisfies equation* (A9*). *If $\hat{U}_T(\boldsymbol{\theta})$ is the empirical contrast field defined by equation* (A9*), *then*

- *Y_{t_1,t_2} satisfies the conditions* **A1-A6** *in the Appendix;*
- *the minimum contrast estimator $\hat{\boldsymbol{\theta}}_T = (\hat{d}_1, \hat{d}_2) = arg\min_{\boldsymbol{\theta}\in\Theta} \hat{U}_T(\boldsymbol{\theta}) \in \Theta$ is a consistent estimator of the parameter vector $\boldsymbol{\theta}$. That is, there is a convergence in P_0 probability:*

$$\hat{\boldsymbol{\theta}}_T \xrightarrow{P_0} \boldsymbol{\theta_0}, \quad T \longrightarrow \infty;$$

- *$\hat{\sigma}_T^2 \xrightarrow{P_0} \sigma^2(\boldsymbol{\theta_0})$, $T \longrightarrow \infty$, where the variance estimator $\hat{\sigma}_T^2$ is given by*

$$\hat{\sigma}_T^2 = \int_{[-\pi,\pi]^2} I_T(\boldsymbol{\lambda}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}.$$

To formulate Theorem 16 we introduce the following notations. The unbiased estimator of the correlation function $\gamma(t_1, t_2, \boldsymbol{\theta})$, $\boldsymbol{t} = (t_1, t_2) \in \mathbb{Z}^2$, of the Gegenbauer random field Y_{t_1,t_2} is

$$\hat{\gamma}_T(\boldsymbol{t}) = \frac{1}{(T - |t_1|)(T - |t_2|)} \sum_{k=0}^{T-|t_1|} \sum_{l=0}^{T-|t_2|} Y_{k,l} Y_{|t_1|+k,|t_2|+l}.$$

Note, that all indices of the random field in the sum above are within the set $\{(t_1, t_2) : t_1, t_2 = 0, ..., T\}$, where the observations are available.

The unbiased periodogram is given by

$$I_T^*(\lambda_1,\lambda_2)=\frac{1}{(2\pi)^2}\sum_{t_1=1-T}^{T-1}\sum_{t_2=1-T}^{T-1}e^{-i(\lambda_1 t_1+\lambda_2 t_2)}\hat{\gamma}_T(\boldsymbol{t}),$$

and the corresponding empirical *contrast field* is

$$\hat{U}_T^*(\boldsymbol{\theta})=-\int_{[-\pi,\pi]^2}I_T^*(\boldsymbol{\lambda})w(\boldsymbol{\lambda})\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\;d\boldsymbol{\lambda}.$$

We also define $\hat{\sigma}_T^{2*}=\int_{[-\pi,\pi]^2}I_T^*(\boldsymbol{\lambda})w(\boldsymbol{\lambda})\;d\boldsymbol{\lambda}$ and the associated adjusted MCE

$$\hat{\boldsymbol{\theta}}_T^*=(\hat{d}_1^*,\hat{d}_2^*)=\arg\min_{\boldsymbol{\theta}\in\Theta}\hat{U}_T^*(\boldsymbol{\theta}). \tag{A9*}$$

Theorem 16. *If $Y_{t_1,t_2},\ (t_1,t_2)\in\mathbb{Z}^2$, is a stationary Gegenbauer random field which spectral density satisfies equation* (A9*) *with $(d_1,d_2)\in(0,1/4)^2$, then*

- *Y_{t_1,t_2} satisfies the conditions* **A1-A9** *in the Appendix;*
- *the adjusted MCE defined by* (A9*) *is asymptotically normal. That is,*

$$T(\hat{\boldsymbol{\theta}}_T^*-\boldsymbol{\theta}_0)\xrightarrow{D}\mathcal{N}_2(0,\mathbf{S}^{-1}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)),\quad T\longrightarrow\infty,$$

where the entries of the matrices $\mathbf{S}(\boldsymbol{\theta})=(s_{ij}(\boldsymbol{\theta}))$ and $\mathbf{A}(\boldsymbol{\theta})=(a_{ij}(\boldsymbol{\theta}))$ are

$$\begin{aligned}s_{ij}(\boldsymbol{\theta}) &= \int_{[-\pi,\pi]^2}f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda})\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\;d\boldsymbol{\lambda}\\ &= \sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}w(\boldsymbol{\lambda})\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})-\frac{1}{\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})}\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda},\end{aligned} \tag{A9*}$$

$$\begin{aligned}a_{ij}(\boldsymbol{\theta}) &= 8\pi^2\int_{[-\pi,\pi]^2}f^2(\boldsymbol{\lambda},\boldsymbol{\theta})w^2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\log\left(\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\frac{\partial}{\partial\theta_j}\log\left(\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\;d\boldsymbol{\lambda}\\ &= 8\pi^2\sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}w^2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\;d\boldsymbol{\lambda}.\end{aligned} \tag{A9*}$$

To avoid the edge effect Anh, Leonenko, and Sakhno (2004) employed the modified periodogram approach suggested by Guyon (1982). We use their assumptions in Theorem 16. Note

that some authors pointed few problems in using I_T^*, see Vidal-Sanz (2009), Yao and Brockwell (2006), and references therein. Various other modifications, for example, typed variograms, smoothed variograms, kernel estimators, to reduce the edge effect have been proposed. It would be interesting to prove analogous of the results by Anh, Leonenko, and Sakhno (2004) and Theorem 16 for these modifications too. However, it is beyond the scope of this paper. Moreover, it remains as an open problem whether the edge-effect modification is essential for the asymptotic normality or not, see Yao and Brockwell (2006).

A9*.4 Proofs

To prove the theorems we will use the following results.

Lemma 5. [Dorogovtsev (1989)] *Let $(X, \mathcal{F}, \mu)$ be a measurable space, $A \in \mathcal{F}$, and $\Theta \subset \mathbb{R}$ be an open set. Suppose the function $F : X \times \Theta \to \mathbb{R}$ satisfies the following conditions:*

1. *For all $\theta \in \Theta$: $F(\cdot, \theta) \in L_1(A)$;*
2. *For almost all $x \in A$ the derivative $\frac{\partial F(x,\cdot)}{\partial \theta}$ exists for all $\theta \in \Theta$;*
3. *There is an integrable function $g : X \to \mathbb{R}$ such that $\left|\frac{\partial F(x,\theta)}{\partial \theta}\right| \leq g(x)$ for almost all $x \in A$.*

Then there exists

$$\frac{\partial}{\partial \theta} \int_A F(x,\theta) d\mu(x) = \int_A \frac{\partial F(x,\theta)}{\partial \theta} d\mu(x).$$

Lemma 6. *The function $\sigma^2(\boldsymbol{\theta})$ is bounded and separated from zero on Θ. Moreover, its first and second order derivatives are bounded on Θ and can be computed by*

$$\frac{\partial}{\partial \theta_i}\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} = -2 \int_{[-\pi,\pi]^2} \log|2\cos\lambda_i - 2u_i|\; w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda}, \quad (\text{A9}^*)$$

$$\begin{aligned} \frac{\partial^2}{\partial \theta_j \partial \theta_i}\sigma^2(\boldsymbol{\theta}) &= \int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_j \partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda} \\ &= 4 \int_{[-\pi,\pi]^2} \log|2\cos\lambda_i - 2u_i|\; \log|2\cos\lambda_j - 2u_j|\; w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}, \boldsymbol{\theta})\, d\boldsymbol{\lambda}, \quad (\text{A9}^*) \end{aligned}$$

where $i, j = 1, 2$.

Proof By the choice of the weight function we obtain

$$\sup_{[-\pi,\pi]^2\times\Theta} f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda}) < +\infty \tag{A9*}$$

and

$$\sup_{\theta\in\Theta}\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda})d\boldsymbol{\lambda} \leq 4\pi^2 \sup_{[-\pi,\pi]^2\times\Theta} f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda}) < +\infty.$$

Hence, $\sigma^2(\boldsymbol{\theta})$ is bounded.

Note, that $\sup_{[-\pi,\pi]^2\times\Theta}\left\{|2\cos\lambda_1 - 2u_1|\right\}^{2d_1}\left\{|2\cos\lambda_2 - 2u_2|\right\}^{2d_2} < +\infty$. Also, by the choice of the weight function, there exists $\delta > 0$ and a set $A_0 \subset [-\pi,\pi]^2$ of non-zero Lebesgue measure, i.e. $\boldsymbol{\lambda}(A_0) > 0$, such that $w(\boldsymbol{\lambda}) > \delta$ for all $\boldsymbol{\lambda} \in A_0$. Therefore,

$$\inf_{\theta\in\Theta}\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda})d\boldsymbol{\lambda} \geq \frac{\delta\,\boldsymbol{\lambda}(A_0)}{\sup_{[-\pi,\pi]^2\times\Theta}\left\{|2\cos\lambda_1 - 2u_1|\right\}^{2d_1}\left\{|2\cos\lambda_2 - 2u_2|\right\}^{2d_2}} > 0,$$

which means that $\sigma^2(\boldsymbol{\theta})$ is separated from zero on Θ.

Now, to study $\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})$ we compute $\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})$, $i = 1, 2$:

$$\begin{aligned}\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta}) &= \frac{\sigma_\varepsilon^2}{(2\pi)^2}\left\{|2\cos\lambda_j - 2u_j|\right\}^{-2d_j}\frac{\partial}{\partial d_i}\left\{|2\cos\lambda_i - 2u_i|\right\}^{-2d_i} \\ &= \log\left(2\cos\lambda_i - 2u_i\right)^{-2}\frac{\sigma_\varepsilon^2}{(2\pi)^2}\left[2(\cos\lambda_i - u_i)\right]^{-2d_i}\left[2(\cos\lambda_j - u_j)\right]^{-2d_j} \\ &= -2\log|2\cos\lambda_i - 2u_i|\, f(\boldsymbol{\lambda},\boldsymbol{\theta}). \end{aligned} \tag{A9*}$$

Using (A9*) and (A9*) we conclude that

$$\sup_{[-\pi,\pi]^2\times\Theta}\left|w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right| < +\infty. \tag{A9*}$$

Thus, by (A9*) and Lemma 5 there exists

$$\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\; d\boldsymbol{\lambda}$$

and

$$\sup_{\Theta}\left|\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right| \leq 4\pi^2 \sup_{[-\pi,\pi]^2\times\Theta}\left|w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right| < +\infty.$$

It is not difficult to find $\frac{\partial^2}{\partial\theta_i\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})$. By (A9*) the second derivatives of f are given by

$$\frac{\partial^2}{\partial\theta_i^2}f(\boldsymbol{\lambda},\boldsymbol{\theta}) = -2\log\left(2\cos\lambda_i - 2u_i\right)\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta}) = 4\log\left|2\cos\lambda_i - 2u_i\right|^2 f(\boldsymbol{\lambda},\boldsymbol{\theta}), \qquad \text{(A9*)}$$

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta}) = 4\log\left|2\cos\lambda_i - 2u_i\right|\cdot\log\left|2\cos\lambda_j - 2u_j\right| f(\boldsymbol{\lambda},\boldsymbol{\theta}), \quad i,j=1,2, \quad i\neq j. \qquad \text{(A9*)}$$

It follows from (A9*), (A9*), and (A9*) that

$$\sup_{[-\pi,\pi]^2\times\Theta}\left|w(\boldsymbol{\lambda})\frac{\partial^2}{\partial\theta_j\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right| < +\infty, \quad i,j=1,2. \qquad \text{(A9*)}$$

Finally, by (A9*) and Lemma 5 there exists

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}\sigma^2(\boldsymbol{\theta}) = \int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda})\frac{\partial^2}{\partial\theta_j\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\; d\boldsymbol{\lambda}$$

and

$$\sup_{\Theta}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right| \leq 4\pi^2 \sup_{[-\pi,\pi]^2\times\Theta}\left|w(\boldsymbol{\lambda})\frac{\partial^2}{\partial\theta_j\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right| < +\infty. \qquad \text{(A9*)}$$

■

Proof of Theorem 15. We will prove that the conditions **A1-A6** in the Appendix are satisfied. Therefore, we will be able to apply Theorem **??** by Anh, Leonenko and Sakhno (2004) and obtain the statement of Theorem 15.

The condition **A1** holds, since $\boldsymbol{\theta}_0$ belongs to the parameter space $\Theta = \left(0,\frac{1}{2}\right)^2$ which is an interior of the compact set $\left[0,\frac{1}{2}\right]^2$.

It follows from representation (A9*) of the spectral density that $f(\boldsymbol{\lambda},\boldsymbol{\theta}_1) \neq f(\boldsymbol{\lambda},\boldsymbol{\theta}_2)$, for $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Thus, the condition **A2** is satisfied.

The class of non-negative weight functions $w(\boldsymbol{\lambda})$ defined by (A9*) consists of symmetric functions. Note that

$$\left|\cos(\lambda_i) - \cos(\nu_i)\right| = 2\left|\sin\left(\frac{\lambda_i+\nu_i}{2}\right)\sin\left(\frac{\lambda_i-\nu_i}{2}\right)\right| \sim C\left|\lambda_i^2 - \nu_i^2\right|,$$

when $\lambda_i \to \pm\nu_i$. Thus, by (A9*) and representation (A9*) of the spectral density we get $w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta}) \in L_1([-\pi,\pi]^2)$ for all $\boldsymbol{\theta}$.

To verify **A4**, that is, to prove

$$\nabla_{\boldsymbol{\theta}} \int_{[-\pi,\pi]^2} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} = \int_{[-\pi,\pi]^2} \nabla_{\boldsymbol{\theta}} \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} = 0,$$

we find

$$w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) = \frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda},\boldsymbol{\theta})\right] - \frac{w(\boldsymbol{\lambda})}{\sigma^4(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right] f(\boldsymbol{\lambda},\boldsymbol{\theta}) =: S_1(\boldsymbol{\lambda},\boldsymbol{\theta}) - S_2(\boldsymbol{\lambda},\boldsymbol{\theta}) \tag{A9*}$$

and apply Lemma 1.

By (A9*) and the choice of the weight function we obtain

$$\sup_{[-\pi,\pi]^2\times\Theta} \big|\log|2\cos\lambda_i - 2u_i|\,\big|\, f(\boldsymbol{\lambda},\boldsymbol{\theta}) w(\boldsymbol{\lambda}) < +\infty. \tag{A9*}$$

Therefore, by equation (A9*) and Lemma 6:

$$|S_1(\boldsymbol{\lambda},\boldsymbol{\theta})| \le \frac{C}{\sigma^2(\boldsymbol{\theta})} \le \frac{C}{\min_\Theta \sigma^2(\boldsymbol{\theta})} \in L_1([-\pi,\pi]^2). \tag{A9*}$$

Now, it follows from (A9*), (A9*) and Lemma 6 that we can estimate $S_2(\boldsymbol{\lambda},\boldsymbol{\theta})$ as

$$|S_2(\boldsymbol{\lambda},\boldsymbol{\theta})| = \left| w(\boldsymbol{\lambda}) \frac{\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right] f(\boldsymbol{\lambda},\boldsymbol{\theta})}{\sigma^4(\boldsymbol{\theta})} \right| \le C\, \frac{w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda},\boldsymbol{\theta})}{\min_\Theta \sigma^4(\boldsymbol{\theta})} \le C. \tag{A9*}$$

Finally, **A4** follows from (A9*), (A9*), and Lemma 5 with $g(x) = C$.

Note that $L_1([-\pi,\pi]^2) \cap L_2([-\pi,\pi]^2) = L_2([-\pi,\pi]^2)$. To verify the condition **A5** for the weight function $w(\cdot)$ we have to show that $f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) \in L_2([-\pi,\pi]^2)$, for all $\boldsymbol{\theta} \in \Theta$. By (A9*) and (A9*) the product $f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})$ is bounded for all $\boldsymbol{\lambda}$ except $\{\boldsymbol{\lambda} : \lambda_i = \pm\nu_i,\ i = 1,2\}$. Let $\tilde{d}_i = \max(d_i, d_{i0})$. Then, for $\lambda_i \to \pm\nu_i$, $i = 1,2$:

$$[f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda},\boldsymbol{\theta})]^2 \le C \prod_{i=1}^{2} |\lambda_i \pm \nu_i|^{2a_i - 4\tilde{d}_i} \log|\lambda_i \pm \nu_i| \in L_1([-\pi,\pi]^2).$$

Therefore, combining the above results we conclude that **A5** holds.

To verify the condition **A6** we use the following function $v(\boldsymbol{\lambda}) = |\lambda_1^2 - \nu_1^2|^\beta |\lambda_2^2 - \nu_2^2|^\beta$, $\beta \in (0, 1/2)$. Note, that

$$|\lambda_1^2 - \nu_1^2|^\beta |\lambda_2^2 - \nu_2^2|^\beta \log f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \sim C|\lambda_1^2 - \nu_1^2|^\beta |\lambda_2^2 - \nu_2^2|^\beta \left(d_1 \log|\lambda_1^2 - \nu_1^2| + d_2 \log|\lambda_1^2 - \nu_1^2|\right) \to 0,$$

when $\lambda_i \to \pm\nu_i$. Thus, by the choice of $v(\cdot)$ and properties of $\sigma(\boldsymbol{\theta})$ the function

$$h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = v(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = |\lambda_1^2 - \nu_1^2|^\beta |\lambda_2^2 - \nu_2^2|^\beta \left(\log f(\boldsymbol{\lambda}, \boldsymbol{\theta}) - 2 \log \sigma(\boldsymbol{\theta})\right)$$

is uniformly continuous on $[-\pi, \pi]^2 \times \Theta$.

Also, it holds

$$\left| f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) \frac{w(\boldsymbol{\lambda})}{v(\boldsymbol{\lambda})} \right| \leq C v^{-1}(\boldsymbol{\lambda}) \in L_2([-\pi, \pi]^2).$$

Since the conditions **A1**-**A6** are satisfied Theorem **??** follows from Theorem **??** in Anh, Leonenko, and Sakhno (2004). ■

Proof of Theorem 16 to prove the asymptotic normality of the MCE in Theorem 16 we will show that the conditions **A7**-**A9** of the Appendix hold.

We begin by proving the condition **A7**. First, to verify the twice differentiability of the function $\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ on Θ we formally compute the second-order derivatives of Ψ :

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_j \partial\theta_i} \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_j} \left[\frac{\left[\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\right] \sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i} \sigma^2(\boldsymbol{\theta})\right] f(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\sigma^4(\boldsymbol{\theta})} \right] \\
&= \frac{1}{\sigma^4(\boldsymbol{\theta})} \left\{ \left[\frac{\partial^2}{\partial\theta_j \partial\theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\right] \sigma^2(\boldsymbol{\theta}) + \left[\frac{\partial}{\partial\theta_j} \sigma^2(\boldsymbol{\theta})\right] \left[\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\right] \right\} \\
&- \frac{1}{\sigma^4(\boldsymbol{\theta})} \left\{ \left[\frac{\partial^2}{\partial\theta_j \partial\theta_i} \sigma^2(\boldsymbol{\theta})\right] f(\boldsymbol{\lambda}, \boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i} \sigma^2(\boldsymbol{\theta})\right] \left[\frac{\partial}{\partial\theta_j} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\right] \right\} \\
&- \frac{1}{\sigma^8(\boldsymbol{\theta})} \left\{ \frac{\partial}{\partial\theta_j} \sigma^4(\boldsymbol{\theta}) \left(\left[\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})\right] \sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i} \sigma^2(\boldsymbol{\theta})\right] f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right\} \quad \text{(A9*)}
\end{aligned}$$

Note that in Lemma 6 we proved that the derivatives $\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})$, $\frac{\partial^2}{\partial\theta_j\partial\theta_i}\sigma^2(\boldsymbol{\theta})$, $\frac{\partial}{\partial\theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, and $\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ exist. Hence, by the above computations and Lemma 6 the function Ψ is twice differentiable on Θ.

In addition, by estimates (A9*), (A9*), (A9*), Lemma 6, and representation (A9*) the product $w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})$ is bounded on $[-\pi,\pi]^2\times\Theta$. Hence,

$$w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\in L_1([-\pi,\pi]^2)\cap L_2([-\pi,\pi]^2),\quad i,j=1,2,\quad \boldsymbol{\theta}\in\Theta.$$

To prove part 2 of the condition **A7**, we first note that by (A9*) and (A9*):

$$\frac{\partial}{\partial\theta_i}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})=\frac{\partial}{\partial\theta_i}\log(f(\boldsymbol{\lambda},\boldsymbol{\theta}))-\frac{\partial}{\partial\theta_i}\log(\sigma^2(\boldsymbol{\theta}))=-2\log|2\cos\lambda_i-2u_i|-\frac{\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})}{\sigma^4(\boldsymbol{\theta})}.$$

By Lemma 6 the second term is bounded. Hence, it follows from (A9*) and (A9*) that the product $w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_i}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})$ is bounded on $[-\pi,\pi]^2\times\Theta$, that implies

$$w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)\frac{\partial}{\partial\theta_i}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\in L_k([-\pi,\pi]^2),\quad k\geq 1,\quad i=1,2,\quad \boldsymbol{\theta}\in\Theta.$$

To verify the condition **A8** we first check the positive definiteness of the matrices $S(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})$.

The entries of $S(\boldsymbol{\theta})$ can be rewritten as

$$s_{i,j}(\boldsymbol{\theta})=\int_{[-\pi,\pi]^2}f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda})\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}=\sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda},$$

where $\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})=f(\boldsymbol{\lambda},\boldsymbol{\theta})w(\boldsymbol{\lambda})/\sigma^2(\boldsymbol{\theta})$.

By (A9*), (A9*), and Lemma 6 the function $\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})$ is integrable, i.e. there is a constant C such that $\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})/C$ is a density. Hence, $S(\boldsymbol{\theta})=C\sigma^2(\boldsymbol{\theta})\mathcal{I}(\theta)$, where $\mathcal{I}$ is the Fisher information matrix of the random vector $\tilde{\boldsymbol{X}}$ with the density $\tilde{\Psi}_w(\boldsymbol{\lambda},\boldsymbol{\theta})=\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})/\int_{[-\pi,\pi]^2}\Psi_w(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}$. Therefore, $S(\boldsymbol{\theta})$ is non-negative definite. Note, that $\mathcal{I}(\theta)=-\left(E\left(\tilde{Q}_i\tilde{Q}_j\right)\right)_{i,j=1,2}$, where $\tilde{Q}_i=\frac{\partial}{\partial\theta_i}\tilde{\Psi}_w(\tilde{\boldsymbol{X}},\boldsymbol{\theta})$. The random variables $\tilde{Q}_1$ and $\tilde{Q}_2$ are not a.s. linearly related which implies positive definiteness of $S(\boldsymbol{\theta})$.

The entries of $A(\boldsymbol{\theta})$ can be rewritten as

$$a_{i,j}(\boldsymbol{\theta})=8\pi^2\sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}w^2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}=C\sigma^2(\boldsymbol{\theta})\,E\left(Q_iQ_j\right),$$

where $Q_i = \frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{X},\boldsymbol{\theta})$ and the random vector $\boldsymbol{X}$ has the density $w^2(\boldsymbol{\lambda})/\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})\,d\boldsymbol{\lambda}$.

As $\left(E\left(Q_iQ_j\right)\right)_{i,j=1,2}$ is a non-negative definite matrix, $A(\boldsymbol{\theta})$ is non-negative definite too. Moreover, it is positive definite, because the random variables Q_1 and Q_2 are not a.s. linearly related.

Now we compute elements of the matrix $S(\boldsymbol{\theta})$. By (A9*)

$$\begin{aligned}
s_{i,j}(\boldsymbol{\theta}) &= \sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2} w(\boldsymbol{\lambda})\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta}) - \frac{1}{\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})}\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda}\\
&= \sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}\left(w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_j}\left[\frac{\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})}{\sigma^4(\boldsymbol{\theta})}\right]\right.\\
&\quad -\frac{w(\boldsymbol{\lambda})}{f(\boldsymbol{\lambda},\boldsymbol{\theta})\sigma^6(\boldsymbol{\theta})}\left(\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\\
&\quad \left.\times\left(\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\right)d\boldsymbol{\lambda}\\
&= \int_{[-\pi,\pi]^2}\left(\frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})}\left(\left[\frac{\partial^2}{\partial\theta_j\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) + \left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\right)\right.\\
&\quad -\frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})}\left(\left[\frac{\partial^2}{\partial\theta_j\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\right)\\
&\quad -2\frac{w(\boldsymbol{\lambda})}{\sigma^6(\boldsymbol{\theta})}\left(\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\left(\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\right)\\
&\quad -\frac{w(\boldsymbol{\lambda})}{\sigma^4(\boldsymbol{\theta})f(\boldsymbol{\lambda},\boldsymbol{\theta})}\left(\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\\
&\quad \left.\times\left(\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)d\boldsymbol{\lambda}.\right.
\end{aligned}$$

By (A9*), (A9*), (A9*), and (A9*) we obtain

$$s_{i,j}(\boldsymbol{\theta}) = 3\int_{[-\pi,\pi]^2}\frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda} - \int_{[-\pi,\pi]^2}\frac{w(\boldsymbol{\lambda})}{f(\boldsymbol{\lambda},\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]d\boldsymbol{\lambda}$$

$$= \frac{3}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right] - 4\int_{[-\pi,\pi]^2}\log\left|2\cos\lambda_i - 2u_i\right|\log\left|2\cos\lambda_j - 2u_j\right|\,w(\boldsymbol{\lambda})f(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}.$$

By (A9*) the elements of the matrix $A(\boldsymbol{\theta})$ are

$$\begin{aligned} a_{i,j}(\boldsymbol{\theta}) &= 8\pi^2\sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda} \\ &= \frac{8\pi^2}{\sigma^2(\boldsymbol{\theta})}\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})\left(\left[\frac{\partial}{\partial\theta_i}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta})-\left[\frac{\partial}{\partial\theta_i}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right) \\ &\quad\times\left(\left[\frac{\partial}{\partial\theta_j}f(\boldsymbol{\lambda},\boldsymbol{\theta})\right]\sigma^2(\boldsymbol{\theta})-\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]f(\boldsymbol{\lambda},\boldsymbol{\theta})\right)\,d\boldsymbol{\lambda}. \end{aligned}$$

Hence, by (A9*) we get

$$a_{i,j}(\boldsymbol{\theta}) = S_1 - S_2(i,j) - S_2(j,i) + S_3,$$

where

$$\begin{aligned} S_1 &= 32\pi^2\sigma^2(\boldsymbol{\theta})\int_{[-\pi,\pi]^2}\log|2\cos\lambda_i-2u_i|\log|2\cos\lambda_j-2u_j|\;w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}, \\ S_2(i,j) &= 16\pi^2\sigma^2(\boldsymbol{\theta})\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\int_{[-\pi,\pi]^2}\log|2\cos\lambda_i-2u_i|\;w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}, \\ S_3 &= \frac{8\pi^2}{\sigma^2(\boldsymbol{\theta})}\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\left[\frac{\partial}{\partial\theta_j}\sigma^2(\boldsymbol{\theta})\right]\int_{[-\pi,\pi]^2} w^2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda},\boldsymbol{\theta})d\boldsymbol{\lambda}. \end{aligned}$$

The proof of the condition **A9** is based on the approach in Bentkus (1972). Notice, that by (A9*) there exists a factorization $w(\boldsymbol{\lambda}) = w_1(\boldsymbol{\lambda})\cdot w_2(\boldsymbol{\lambda})$ of $w(\boldsymbol{\lambda})$ such that both $\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0) = f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)w_1(\boldsymbol{\lambda})$ and $w_2(\boldsymbol{\lambda})$ are bounded functions of $\boldsymbol{\lambda}$. Let us denote $\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta}) = w_2(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})$. Then,

$$\begin{aligned} &T\int_{[-\pi,\pi]^2}(EI_T^*(\boldsymbol{\lambda})-f(\boldsymbol{\lambda},\boldsymbol{\theta}_0))w(\boldsymbol{\lambda})\frac{\partial}{\partial\theta_i}\log\Psi(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda} = T\int_{[-\pi,\pi]^2}(EI_T^*(\boldsymbol{\lambda})w_1(\boldsymbol{\lambda})-\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0))\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda} \\ &= T\int_{[-\pi,\pi]^2}(EI_T^*(\boldsymbol{\lambda})w_1(\boldsymbol{\lambda})-E\tilde{I}_T^*(\boldsymbol{\lambda}))\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda} + T\int_{[-\pi,\pi]^2}(E\tilde{I}_T^*(\boldsymbol{\lambda})-\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0))\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}, \end{aligned} \tag{A9*}$$

where $\tilde{I}_T^*(\boldsymbol{\lambda})$ denotes the unbiased periodogram of the Gegenbauer random field with the spectral density $\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)$.

Due to the boundness of $\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)$, analogously to Guyon (1982) and Bentkus (1972), the second term in (A9*) vanishes when $T\to\infty$. Therefore, to prove the condition **A9**, it is enough to show that the first term in (A9*) vanishes too.

Let $\tilde{\gamma}(\boldsymbol{t},\boldsymbol{\theta}_0)$ denote the auto-covariance function of the Gegenbauer random field with the spectral density $\tilde{f}(\boldsymbol{\lambda},\boldsymbol{\theta}_0)$. By multidimensional Parseval's theorem, see Brychkov et al. (1992), we get

$$\int_{[-\pi,\pi]^2}(EI_T^*(\boldsymbol{\lambda})w_1(\boldsymbol{\lambda})-E\tilde{I}_T^*(\boldsymbol{\lambda}))\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}=\frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2}\Big(w_1(\boldsymbol{\lambda})\sum_{t_1=1-T}^{T-1}\sum_{t_2=1-T}^{T-1}e^{-i(\lambda_1t_1+\lambda_2t_2)}\gamma(\boldsymbol{t},\boldsymbol{\theta}_0)$$

$$-\sum_{t_1=1-T}^{T-1}\sum_{t_2=1-T}^{T-1}e^{-i(\lambda_1t_1+\lambda_2t_2)}\tilde{\gamma}(\boldsymbol{t},\boldsymbol{\theta}_0)\Big)\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}=\frac{1}{(2\pi)^2}\int_{[-\pi,\pi]^2}\Big(w_1(\boldsymbol{\lambda})\sum_{(t_1,t_2)\in\mathbb{Z}^2}e^{-i(\lambda_1t_1+\lambda_2t_2)}\gamma(\boldsymbol{t},\boldsymbol{\theta}_0)\times$$

$$I_{[1-T,T-1]^2}(t_1,t_2)-\sum_{(t_1,t_2)\in\mathbb{Z}^2}e^{-i(\lambda_1t_1+\lambda_2t_2)}\tilde{\gamma}(\boldsymbol{t},\boldsymbol{\theta}_0)I_{[1-T,T-1]^2}(t_1,t_2)\Big)\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}=$$

$$\int_{[-\pi,\pi]^2}\Big(w_1(\boldsymbol{\lambda})\int_{[-\pi,\pi]^2}f(\boldsymbol{x},\boldsymbol{\theta}_0)\Phi_{T-1}(\boldsymbol{\lambda}-\boldsymbol{x})\,d\boldsymbol{x}-\int_{[-\pi,\pi]^2}\tilde{f}(\boldsymbol{x},\boldsymbol{\theta}_0)\Phi_{T-1}(\boldsymbol{\lambda}-\boldsymbol{x})\,d\boldsymbol{x}\Big)\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda}$$

$$\int_{[-\pi,\pi]^2}\Big(\int_{[-\pi,\pi]^2}f(\boldsymbol{x}+\boldsymbol{\lambda},\boldsymbol{\theta}_0)\Phi_{T-1}(\boldsymbol{x})(w_1(\boldsymbol{\lambda})-w_1(\boldsymbol{\lambda}+\boldsymbol{x}))\,d\boldsymbol{x}\Big)\tilde{w}_2(\boldsymbol{\lambda},\boldsymbol{\theta})\,d\boldsymbol{\lambda},\qquad\text{(A9*)}$$

where $\Phi_{T-1}(\cdot)$ is the Fejér kernel.

Let $f(\boldsymbol{\lambda},\boldsymbol{\theta}_0)\in L^2([-\pi,\pi]^2)$, i.e. $(d_1,d_2)\in(0,1/4)^2$. Then, repeating the proof of part 1) of Theorem 2.2 in Bentkus (1972) for the two-dimensional case with $p=2$ we obtain that the integral in (A9*) is bounded by $C\varepsilon_T/T$, where $\varepsilon_T\to 0$ when $T\to\infty$. It implies that first term in (A9*) vanishes and completes the proof of the condition **A9**. ■

A9*.5 Simulation studies

In this section we presented some numerical results to confirm the theoretical findings.

Figure A9*.3 demonstrates a series of box plots to characterize the sample distribution of MCEs of the parameters d_i, $i=1,2$, as a function of T. To compute it Monte Carlo simulations of the Gegenbauer field with 100 replications for each $T=10,20,30,40,50$ were performed. For the parameters $u_1=0.4$, $u_2=0.3$, $d_1=0.2$, and $d_2=0.3$ realizations of Y_{t_1,t_2} were simulated using the truncated sum $\sum_{n_1=0}^{40}\sum_{n_2=0}^{40}$ in (A9*). For example, a realization of the Gegenbauer random field on a 100×100 grid is shown in Figure A9*.1. We set the parameter values of

the weight function in (A9*) to $a_1 = a_2 = 2$ and $w_0(\boldsymbol{\lambda}) \equiv 1$. The periodogram $I_T(\boldsymbol{\lambda})$ was computed and the minimizing argument $\hat{\boldsymbol{\theta}}_T$ of the functional $\hat{U}_T(\boldsymbol{\theta})$ was found numerically for each simulation. Figure A9*.3 demonstrates that the sample distribution of $\hat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_0$ as T increases. The plot of the sample probabilities $P_0(|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$ in Figure A9*.4 also confirms convergence in probability of $\hat{\boldsymbol{\theta}}_T$ to $\boldsymbol{\theta}_0$.

For each generated realization we also computed the value of $\hat{\sigma}_T^2$ using $I_T(\boldsymbol{\lambda})$. Analogously to Figure A9*.3 and A9*.4, plots in Figure A9*.5 and A9*.6 support convergence in probability of $\hat{\sigma}_T^2$ to $\sigma^2(\boldsymbol{\theta_0})$ when T increases.

To verify the result of Theorem 16 we used sample values $\hat{\boldsymbol{\theta}}_{50}^*$ which minimized the functional $\hat{U}_T^*(\boldsymbol{\theta})$ for each simulation. To avoid possible negative values the modified periodogram $I_T^*(\boldsymbol{\lambda})$ was truncated at zero by the R program. Bearing in mind the edge effect and modified periodogram's correction, Figure A9*.9 demonstrates that the results are close to the expected ones even for the relatively small $T = 50$. The normal Q-Q plot of each component of $\hat{\boldsymbol{\theta}}_{50}^*$ in Figure A9*.9 matches with the theoretical normal distribution. To test the bivariate normality hypothesis about $\hat{\boldsymbol{\theta}}_{50}^*$ we used the Shapiro-Wilk, energy, and kurtosis tests of multivariate normality from the R packages MVNORMTEST, ENERGY, and ICS. In all the tests, p-values (0.9491, 0.4605, and 0.5314) confirmed that $\hat{\boldsymbol{\theta}}_T^*$ asymptotically follows a bivariate normal distribution. Simulations for other values of the parameters were run, with similar results.

Hence, we conclude that the MCEs are consistent estimators and the distributions of $\hat{\boldsymbol{\theta}}_T^*$ converge to the bivariate normal law. Note, that the simulation studies confirm the convergence not only for $(d_1, d_2) \in (0, 1/4)^2$, but for all possible values of (d_1, d_2) in $(0, 1/2)^2$.

A9*.6 Directions for future research

The estimation methodology based on the unbiased periodogram was introduced in Guyon (1982, 1995), see also Heyde and Gay (1993). Recently, the paper by Robinson and Sanz (2006)

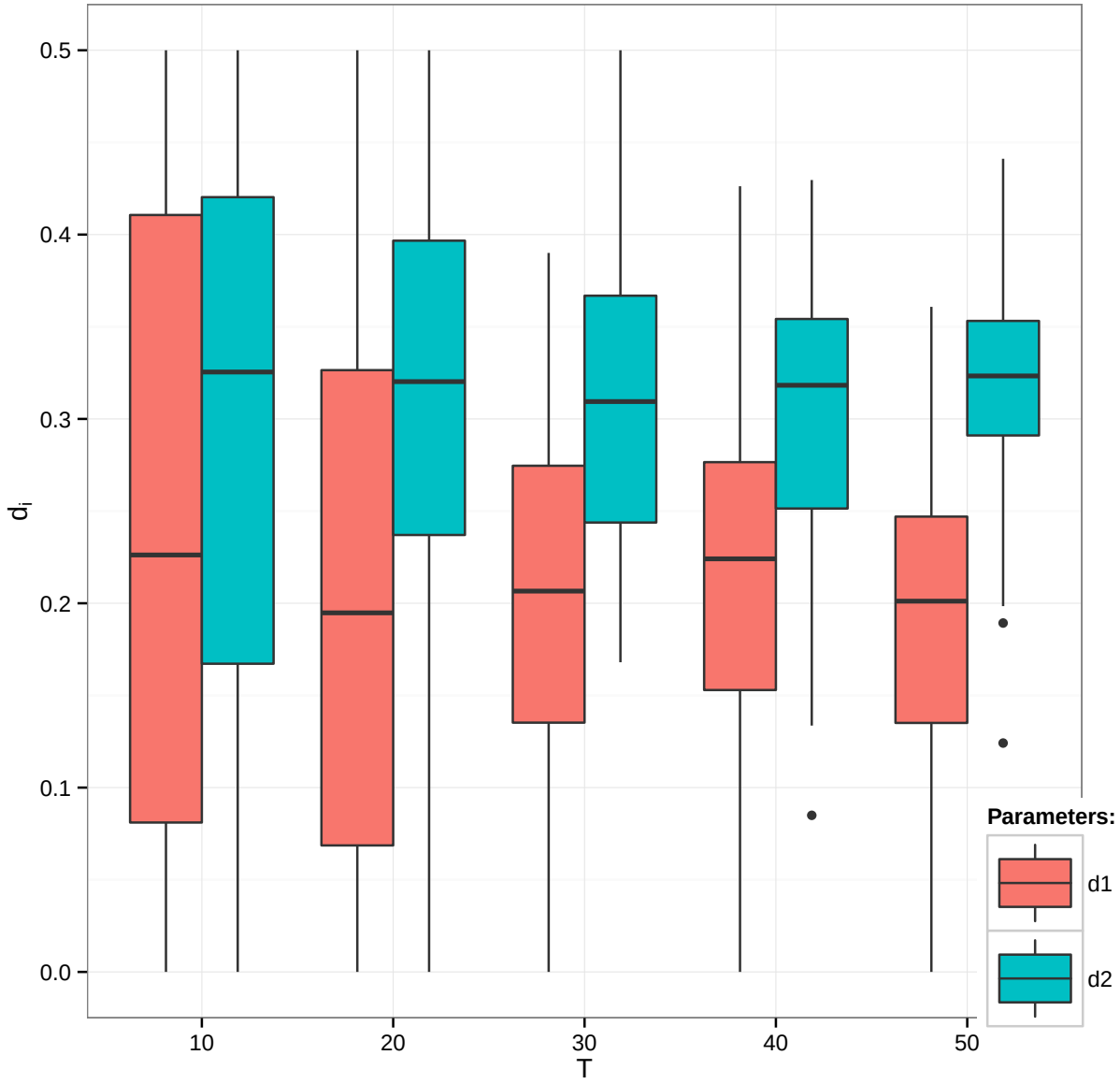


Figure A9*.3: Boxplots of sampled values of $\hat{\boldsymbol{\theta}}_T$.

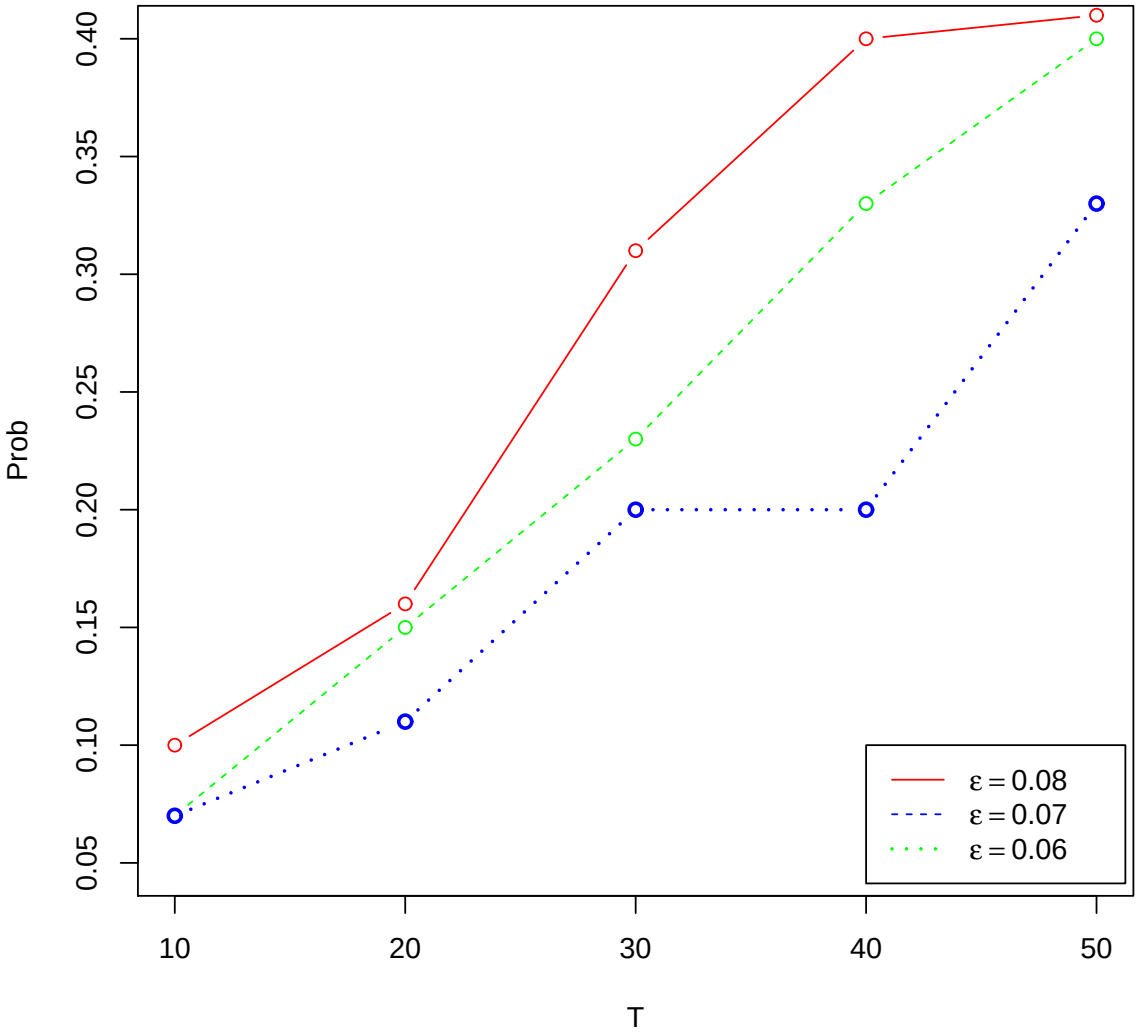


Figure A9*.4: Sample probabilities $P_0(|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$.

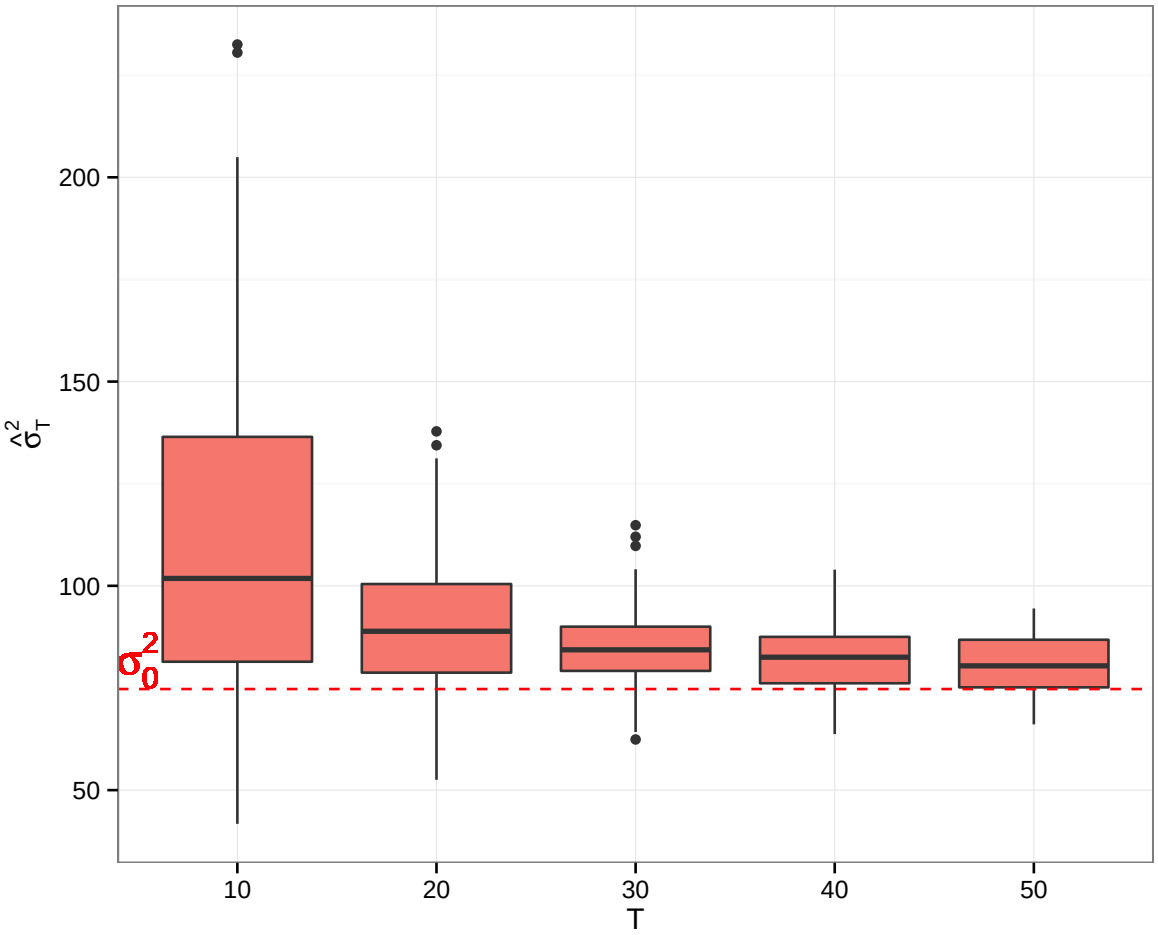


Figure A9*.5: Boxplots of sampled values of $\hat{\sigma}_T^2$.

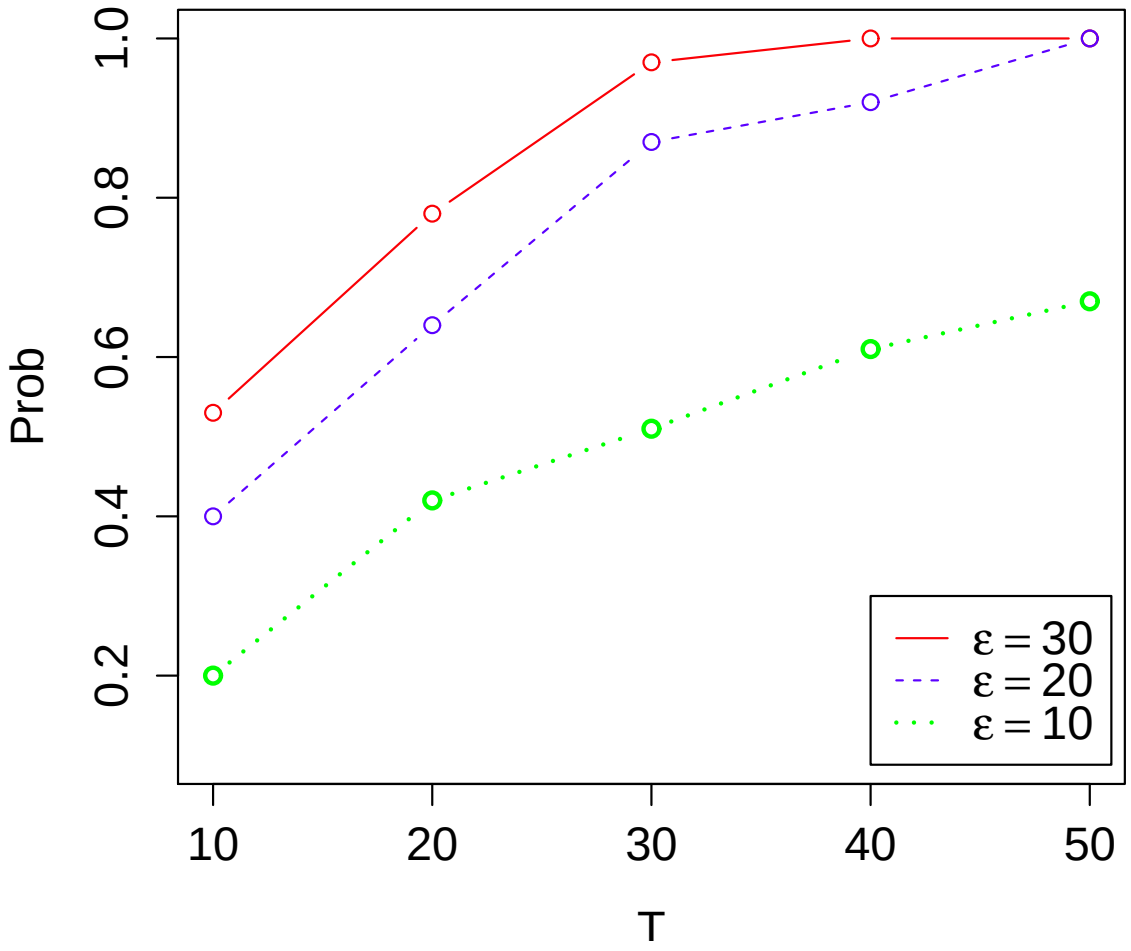


Figure A9*.6: Sample probabilities $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta_0})| < \varepsilon)$.

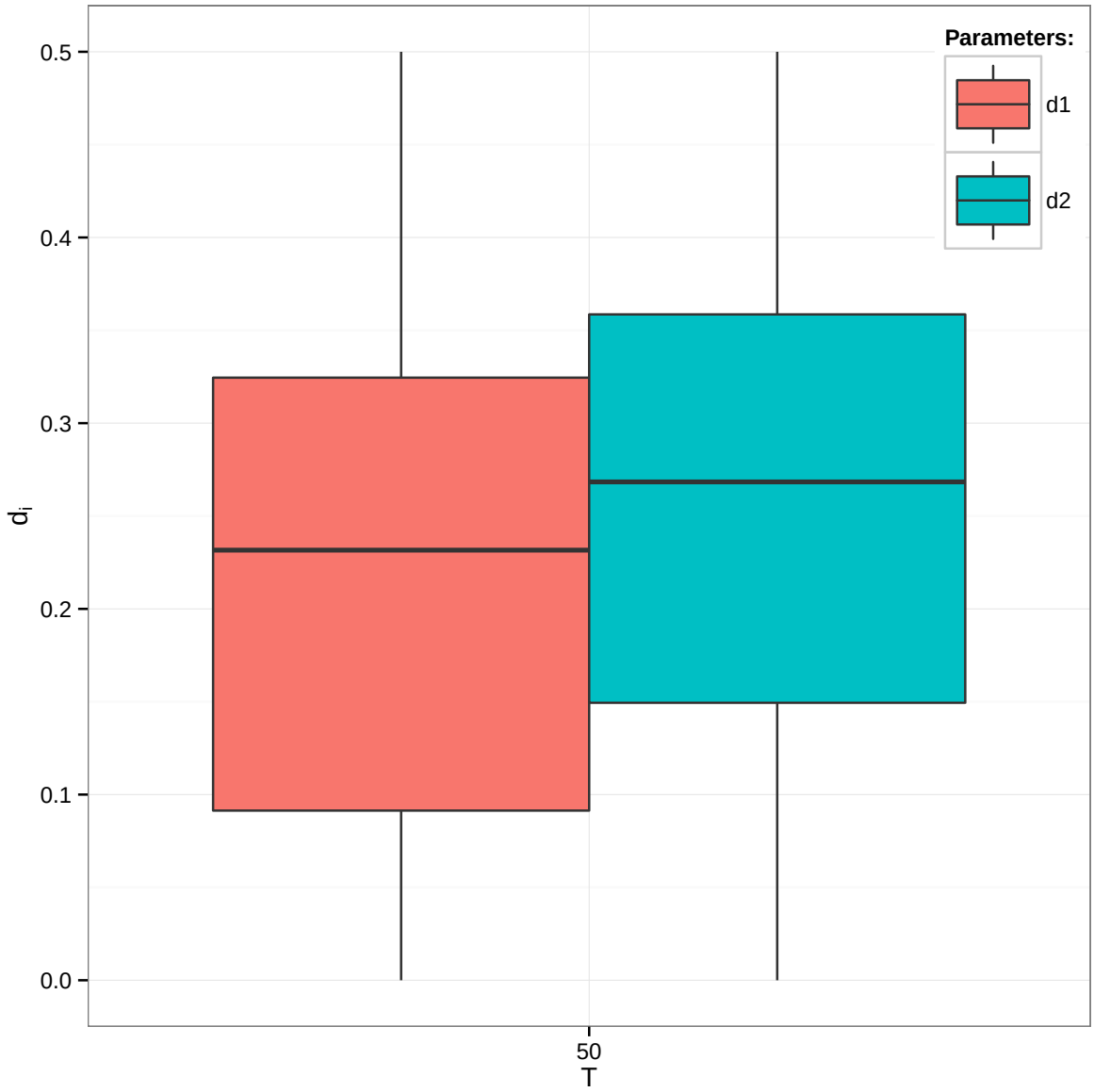


Figure A9*.7: Boxplots of sampled values of $\hat{\boldsymbol{\theta}}^*_{50}$.

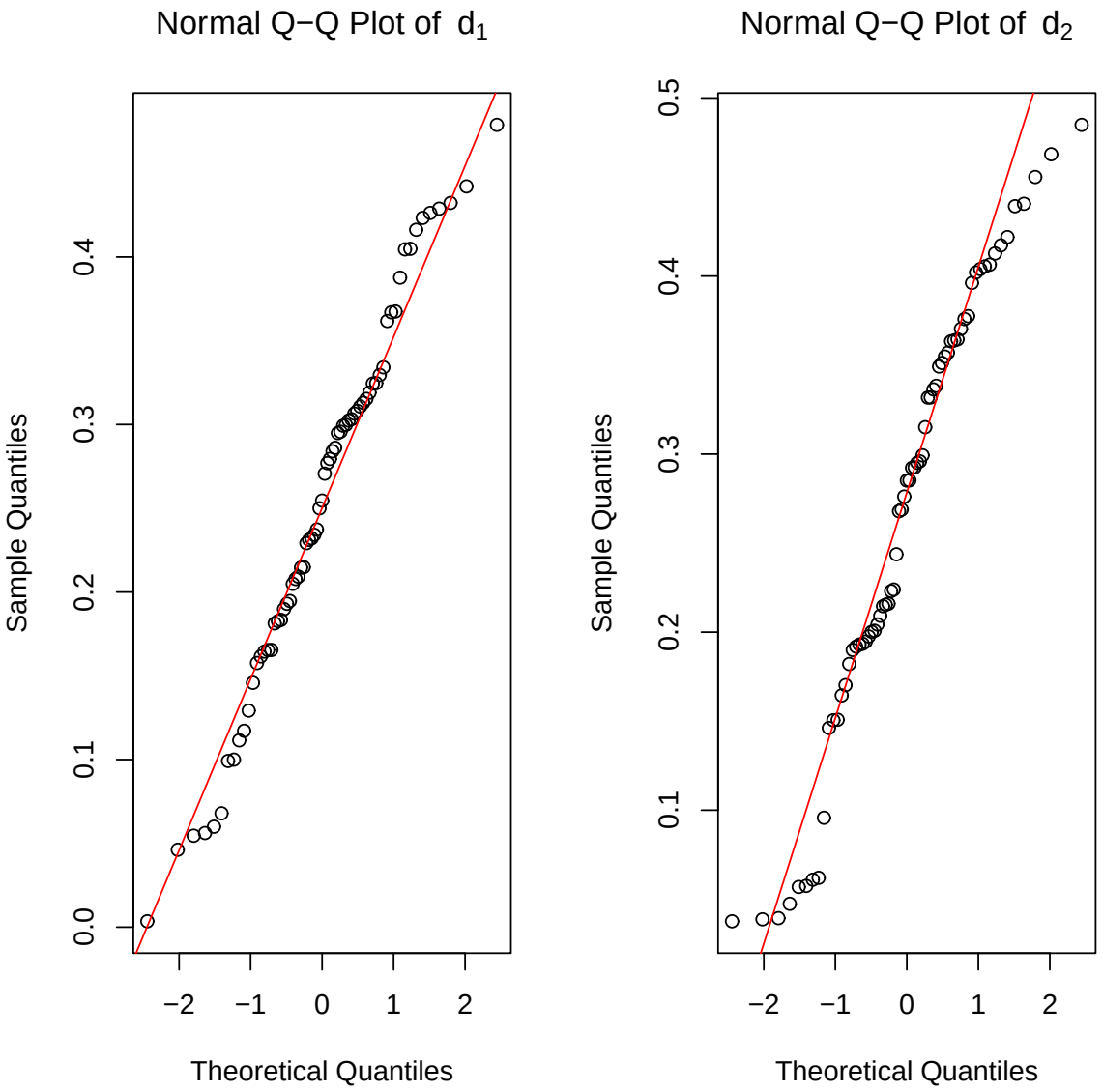


Figure A9*.8: Normal Q-Q plots for each component of $\hat{\boldsymbol{\theta}}^*_{50}$.

Figure A9*.9: Sample distributions of adjusted MCEs.

provided a detailed discussion on the topic, see also the references therein. It studied mainly the difficulties arising in the application of this methodology in high dimensions. In particular, there were investigated problems arising in relation to non-uniformly increasing domain asymptotics associated with different expansion rates of the studied domain in each spatial direction. In this paper, we considered the case $d = 2$ and restricted out attention to the case of uniformly increasing domain asymptotics. The case of non-uniformly increasing domain asymptotics is left for future investigations.

An extended version of the derived results can be obtained for more general formulations of the unbiased periodogram. In particular, different growing rates can be allowed for each spatial dimension in the definition of the sampling area. For example, one can consider the following generalized version of the two-dimensional unbiased periodogram (see, for example, Robinson and Sanz, 2006)

$$I_g(\lambda_1, \lambda_2) = \frac{1}{(2\pi)^2} \sum_{t_1=1-g_1(T)}^{g_1(T)-1} \sum_{t_2=1-g_2(T)}^{g_2(T)-1} e^{-i(\lambda_1 t_1 + \lambda_2 t_2)} \hat{\gamma}_T(\boldsymbol{t}),$$

where the functions $g_i(T)$, $i = 1, 2$, satisfy some suitable conditions (for example, $g_i(T) \to \infty$, $T \to \infty$, and $g_i(T) \leq CT$, $C < 1$, for $i = 1, 2$, and sufficiently large T).

It would be interesting to extend the methodology by Bentkus (1972) to prove the condition **A9** for all (d_1, d_2) in $(0, 1/2)^2$.

Note that our simulation results show that the proposed minimum contrast estimation methodology works in the case of uniformly increasing domain asymptotics.

Appendix

The conditions for consistency and asymptotic normality of the MCE for parameters of stationary fractional Riesz-Bessel type random fields given in Anh, Leonenko, and Sakhno (2004) are specified below for random fields on $\mathbb{Z}^2$.

A1. Let $Y_{t_1,t_2},\ \boldsymbol{t} = (t_1, t_2) \in \mathbb{Z}^2$, be a real-valued measurable stationary Gaussian random field with zero mean and a spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [-\pi, \pi]^2$, $\boldsymbol{\theta} \in \Theta$, and Θ is a compact set. Assume that $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, where $\boldsymbol{\theta}_0$ is the true value of the parameter vector $\boldsymbol{\theta}$.

A2. If $\boldsymbol{\theta_1} \neq \boldsymbol{\theta_2}$ then $f(\boldsymbol{\lambda}, \boldsymbol{\theta_1}) \neq f(\boldsymbol{\lambda}, \boldsymbol{\theta_2})$ for almost all $\boldsymbol{\lambda} \in [-\pi, \pi]^2$ with respect to the Lebesgue measure.

A3. There exists a nonnegative function $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, such that

1. $w(\boldsymbol{\lambda})$ is symmetric about $(0, 0)$, i.e. $w(\boldsymbol{\lambda}) = w(-\boldsymbol{\lambda})$;
2. $w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1\left([-\pi, \pi]^2\right)$ for all $\boldsymbol{\theta} \in \Theta$.

A4. The derivatives $\nabla_{\boldsymbol{\theta}} \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ exist and it is legitimate to differentiate under the integral sign in equation (A9*), i.e.

$$\nabla_{\boldsymbol{\theta}} \int_{[-\pi,\pi]^2} \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda})\ d\boldsymbol{\lambda} = \int_{[-\pi,\pi]^2} \nabla_{\boldsymbol{\theta}} \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda})\ d\boldsymbol{\lambda} = 0.$$

A5. For all $\boldsymbol{\theta} \in \Theta$ the function $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, satisfies

$$f(\boldsymbol{\lambda}, \boldsymbol{\theta_0}) w(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1\left([-\pi, \pi]^2\right) \cap L_2\left([-\pi, \pi]^2\right).$$

A6. There exists a function $v(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in [-\pi, \pi]^2$, such that

1. the function $h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = v(\boldsymbol{\lambda}) \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is uniformly continuous on $[-\pi, \pi]^2 \times \Theta$;
2. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) / v(\boldsymbol{\lambda}) \in L_1([-\pi, \pi]^2) \cap L_2([-\pi, \pi]^2)$.

A7. The function $\Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ is twice differentiable on Θ and

1. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1([-\pi, \pi]^2) \bigcap L_2([-\pi, \pi]^2)$, for all i, j, and $\boldsymbol{\theta} \in \Theta$;
2. $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_k([-\pi, \pi]^2)$, for all i, $\boldsymbol{\theta} \in \Theta$, and $k \geq 1$.

A8. The matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))$ and $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))$ with the elements defined by (A9*) and (A9*) are positive definite.

A9. The spectral density $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, the weight function $w(\boldsymbol{\lambda})$, and the function $\frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ are such that for all i and $\boldsymbol{\theta} \in \Theta$:

$$T \int_{[-\pi,\pi]^2} (EI_T^*(\boldsymbol{\lambda}) - f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \Psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \, d\boldsymbol{\lambda} \longrightarrow 0, \quad \text{as} \quad T \longrightarrow \infty.$$