

Wiley Series in Probability and Statistics

# **GEOSTATISTICAL FUNCTIONAL DATA ANALYSIS**

**JORGE MATEU | RAMÓN GIRALDO**

**WILEY**



## **Geostatistical Functional Data Analysis**

## WILEY SERIES IN PROBABILITY AND STATISTICS

Established by *Walter A. Shewhart and Samuel S. Wilks*

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at <http://www.wiley.com/go/wsp>



# Geostatistical Functional Data Analysis

*Edited by*

*Jorge Mateu*

University Jaume I of Castellon  
Castellon, Spain

*Ramón Giraldo*

National University of Colombia  
Bogota, Colombia

**WILEY**

This edition first published 2022  
© 2022 John Wiley & Sons Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Jorge Mateu and Ramón Giraldo to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

#### *Registered Offices*

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

#### *Editorial Office*

9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at [www.wiley.com](http://www.wiley.com).

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

#### *Limit of Liability/Disclaimer of Warranty*

The contents of this work are intended to further general scientific research, understanding, and discussion only and are not intended and should not be relied upon as recommending or promoting scientific method, diagnosis, or treatment by physicians for any particular patient. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of medicines, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each medicine, equipment, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

#### *Library of Congress Cataloging-in-Publication Data*

Names: Mateu, Jorge, editor. | Giraldo, Ramón, editor.

Title: Geostatistical functional data analysis / edited by Jorge Mateu, Ramón Giraldo.

Description: Hoboken, NJ : Wiley, 2022. | Series: Wiley series in probability and statistics | Includes bibliographical references and index.

Identifiers: LCCN 2021015788 (print) | LCCN 2021015789 (ebook) | ISBN 9781119387848 (hardback) | ISBN 9781119387909 (adobe pdf) | ISBN 9781119387886 (epub)

Subjects: LCSH: Geology—Statistical methods. | Kriging. | Spatial analysis (Statistics) | Functional analysis.

Classification: LCC QE33.2.S82 G434 2022 (print) | LCC QE33.2.S82 (ebook) | DDC 551.072/7—dc23

LC record available at <https://lcn.loc.gov/2021015788>

LC ebook record available at <https://lcn.loc.gov/2021015789>

Cover Design: Wiley

Cover Image: © Googee/Shutterstock

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

## Contents

**List of Contributors** *xiii*

**Foreword** *xvi*

### **1 Introduction to Geostatistical Functional Data Analysis** *1*

*Jorge Mateu and Ramón Giraldo*

- 1.1 Spatial Statistics *1*
- 1.2 Spatial Geostatistics *7*
  - 1.2.1 Regionalized Variables *7*
  - 1.2.2 Random Functions *7*
  - 1.2.3 Stationarity and Intrinsic Hypothesis *9*
- 1.3 Spatiotemporal Geostatistics *12*
  - 1.3.1 Relevant Spatiotemporal Concepts *12*
  - 1.3.2 Spatiotemporal Kriging *16*
  - 1.3.3 Spatiotemporal Covariance Models *17*
- 1.4 Functional Data Analysis in Brief *18*
- References *22*

### **Part I Mathematical and Statistical Foundations** *27*

### **2 Mathematical Foundations of Functional Kriging in Hilbert Spaces and Riemannian Manifolds** *29*

*Alessandra Menafoglio, Davide Pigoli, and Piercesare Secchi*

- 2.1 Introduction *29*
- 2.2 Definitions and Assumptions *30*
- 2.3 Kriging Prediction in Hilbert Space: A Trace Approach *33*
  - 2.3.1 Ordinary and Universal Kriging in Hilbert Spaces *33*
  - 2.3.2 Estimating the Drift *36*
  - 2.3.3 An Example: Trace-Variogram in Sobolev Spaces *37*

|       |                                                                     |    |
|-------|---------------------------------------------------------------------|----|
| 2.3.4 | An Application to Nonstationary Prediction of Temperatures Profiles | 39 |
| 2.4   | An Operatorial Viewpoint to Kriging                                 | 42 |
| 2.5   | Kriging for Manifold-Valued Random Fields                           | 45 |
| 2.5.1 | Residual Kriging                                                    | 45 |
| 2.5.2 | An Application to Positive Definite Matrices                        | 47 |
| 2.5.3 | Validity of the Local Tangent Space Approximation                   | 49 |
| 2.6   | Conclusion and Further Research                                     | 53 |
|       | References                                                          | 53 |

### **3 Universal, Residual, and External Drift Functional Kriging**

*Maria Franco-Villoria and Rosaria Ignaccolo*

|       |                                                       |    |
|-------|-------------------------------------------------------|----|
| 3.1   | Introduction                                          | 56 |
| 3.2   | Universal Kriging for Functional Data (UKFD)          | 56 |
| 3.3   | Residual Kriging for Functional Data (ResKFD)         | 58 |
| 3.4   | Functional Kriging with External Drift (FKED)         | 60 |
| 3.5   | Accounting for Spatial Dependence in Drift Estimation | 61 |
| 3.5.1 | Drift Selection                                       | 62 |
| 3.6   | Uncertainty Evaluation                                | 62 |
| 3.7   | Implementation Details in R                           | 64 |
| 3.7.1 | <i>Example: Air Pollution Data</i>                    | 64 |
| 3.8   | Conclusions                                           | 69 |
|       | References                                            | 71 |

### **4 Extending Functional Kriging When Data Are Multivariate Curves: Some Technical Considerations and Operational Solutions**

*David Nerini, Claude Manté, and Pascal Monestiez*

|       |                                                |    |
|-------|------------------------------------------------|----|
| 4.1   | Introduction                                   | 73 |
| 4.2   | Principal Component Analysis for Curves        | 74 |
| 4.2.1 | Karhunen–Loève Decomposition                   | 74 |
| 4.2.2 | Dealing with a Sample                          | 76 |
| 4.3   | Functional Kriging in a Nutshell               | 78 |
| 4.3.1 | Solution Based on Basis Functions              | 79 |
| 4.3.2 | Estimation of Spatial Covariances              | 81 |
| 4.4   | An Example with the Precipitation Observations | 82 |
| 4.4.1 | Fitting Variogram Model                        | 83 |
| 4.4.2 | Making Prediction                              | 83 |
| 4.5   | Functional Principal Component Kriging         | 85 |
| 4.6   | Multivariate Kriging with Functional Data      | 88 |

|          |                                                                                                                                           |            |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 4.6.1    | Multivariate FPCA                                                                                                                         | 91         |
| 4.6.2    | MFPCA Displays                                                                                                                            | 93         |
| 4.6.3    | Multivariate Functional Principal Component Kriging                                                                                       | 94         |
| 4.6.4    | Mixing Temperature and Precipitation Curves                                                                                               | 96         |
| 4.7      | Discussion                                                                                                                                | 98         |
| 4.A      | Appendices                                                                                                                                | 100        |
| 4.A.1    | Computation of the Kriging Variance                                                                                                       | 100        |
|          | References                                                                                                                                | 102        |
| <b>5</b> | <b>Geostatistical Analysis in Bayes Spaces: Probability Densities and Compositional Data</b>                                              | <b>104</b> |
|          | <i>Alessandra Menafoglio, Piercesare Secchi, and Alberto Guadagnini</i>                                                                   |            |
| 5.1      | Introduction and Motivations                                                                                                              | 104        |
| 5.2      | Bayes Hilbert Spaces: Natural Spaces for Functional Compositions                                                                          | 105        |
| 5.3      | A Motivating Case Study: Particle-Size Data in Heterogeneous Aquifers – Data Description                                                  | 108        |
| 5.4      | Kriging Stationary Functional Compositions                                                                                                | 110        |
| 5.4.1    | Model Description                                                                                                                         | 110        |
| 5.4.2    | Data Preprocessing                                                                                                                        | 112        |
| 5.4.3    | An Example of Application                                                                                                                 | 113        |
| 5.4.4    | Uncertainty Assessment                                                                                                                    | 116        |
| 5.5      | Analyzing Nonstationary Fields of FCs                                                                                                     | 119        |
| 5.6      | Conclusions and Perspectives                                                                                                              | 123        |
|          | References                                                                                                                                | 124        |
| <b>6</b> | <b>Spatial Functional Data Analysis for Probability Density Functions: Compositional Functional Data vs. Distributional Data Approach</b> | <b>128</b> |
|          | <i>Elvira Romano, Antonio Irpino, and Jorge Mateu</i>                                                                                     |            |
| 6.1      | FDA and SDA When Data Are Densities                                                                                                       | 130        |
| 6.1.1    | Features of Density Functions as Compositional Functional Data                                                                            | 131        |
| 6.1.2    | Features of Density Functions as Distributional Data                                                                                      | 135        |
| 6.2      | Measures of Spatial Association for Georeferenced Density Functions                                                                       | 138        |
| 6.2.1    | Identification of Spatial Clusters by Spatial Association Measures for Density Functions                                                  | 139        |
| 6.3      | Real Data Analysis                                                                                                                        | 141        |
| 6.3.1    | The SDA Distributional Approach                                                                                                           | 143        |
| 6.3.2    | The Compositional–Functional Approach                                                                                                     | 145        |
| 6.3.3    | Discussion                                                                                                                                | 147        |

|     |                 |     |
|-----|-----------------|-----|
| 6.4 | Conclusion      | 149 |
|     | Acknowledgments | 151 |
|     | References      | 151 |

## **Part II Statistical Techniques for Spatially Correlated Functional Data 155**

|          |                                                                                    |            |
|----------|------------------------------------------------------------------------------------|------------|
| <b>7</b> | <b>Clustering Spatial Functional Data</b>                                          | <b>157</b> |
|          | <i>Vincent Vandewalle, Cristian Preda, and Sophie Dabo-Niang</i>                   |            |
| 7.1      | Introduction                                                                       | 157        |
| 7.2      | Model-Based Clustering for Spatial Functional Data                                 | 158        |
| 7.2.1    | The Expectation–Maximization (EM) Algorithm                                        | 160        |
| 7.2.1.1  | E Step                                                                             | 161        |
| 7.2.1.2  | M Step                                                                             | 161        |
| 7.2.2    | Model Selection                                                                    | 161        |
| 7.3      | Descendant Hierarchical Classification (HC) Based on Centrality Methods            | 162        |
| 7.3.1    | Methodology                                                                        | 164        |
| 7.4      | Application                                                                        | 165        |
| 7.4.1    | Model-Based Clustering                                                             | 167        |
| 7.4.2    | Hierarchical Classification                                                        | 169        |
| 7.5      | Conclusion                                                                         | 171        |
|          | References                                                                         | 172        |
| <b>8</b> | <b>Nonparametric Statistical Analysis of Spatially Distributed Functional Data</b> | <b>175</b> |
|          | <i>Sophie Dabo-Niang, Camille Ternynck, Baba Thiam, and Anne-Françoise Yao</i>     |            |
| 8.1      | Introduction                                                                       | 175        |
| 8.2      | Large Sample Properties                                                            | 178        |
| 8.2.1    | Uniform Almost Complete Convergence                                                | 180        |
| 8.3      | Prediction                                                                         | 181        |
| 8.4      | Numerical Results                                                                  | 184        |
| 8.4.1    | Bandwidth Selection Procedure                                                      | 184        |
| 8.4.2    | Simulation Study                                                                   | 185        |
| 8.5      | Conclusion                                                                         | 193        |
| 8.A      | Appendix                                                                           | 194        |
| 8.A.1    | Some Preliminary Results for the Proofs                                            | 194        |
| 8.A.2    | Proofs                                                                             | 196        |
| 8.A.2.1  | Proof of Theorem 8.1                                                               | 196        |
| 8.A.2.2  | Proof of Lemma A.3                                                                 | 196        |

|         |                      |     |
|---------|----------------------|-----|
| 8.A.2.3 | Proof of Lemma A.4   | 196 |
| 8.A.2.4 | Proof of Lemma A.5   | 201 |
| 8.A.2.5 | Proof of Lemma A.6   | 201 |
| 8.A.2.6 | Proof of Theorem 8.2 | 202 |
|         | References           | 207 |

## 9      **A Nonparametric Algorithm for Spatially Dependent Functional Data: Bagging Voronoi for Clustering, Dimensional Reduction, and Regression**    211

*Valeria Vitelli, Federica Passamonti, Simone Vantini, and Piercesare Secchi*

|         |                                               |     |
|---------|-----------------------------------------------|-----|
| 9.1     | Introduction                                  | 211 |
| 9.2     | The Motivating Application                    | 212 |
| 9.2.1   | Data Preprocessing                            | 214 |
| 9.3     | The Bagging Voronoi Strategy                  | 216 |
| 9.4     | Bagging Voronoi Clustering (BVClu)            | 218 |
| 9.4.1   | BVClu of the Telecom Data                     | 221 |
| 9.4.1.1 | Setting the BVClu Parameters                  | 221 |
| 9.4.1.2 | Results                                       | 223 |
| 9.5     | Bagging Voronoi Dimensional Reduction (BVDim) | 223 |
| 9.5.1   | BVDim of the Telecom Data                     | 225 |
| 9.5.1.1 | Setting the BVDim Parameters                  | 225 |
| 9.5.1.2 | Results                                       | 227 |
| 9.6     | Bagging Voronoi Regression (BVReg)            | 231 |
| 9.6.1   | Covariate Information: The DUSAF Data         | 232 |
| 9.6.2   | BVReg of the Telecom Data                     | 234 |
| 9.6.2.1 | Setting the BVReg Parameters                  | 234 |
| 9.6.2.2 | Results                                       | 235 |
| 9.7     | Conclusions and Discussion                    | 236 |
|         | References                                    | 239 |

## 10     **Nonparametric Inference for Spatiotemporal Data Based on Local Null Hypothesis Testing for Functional Data**    242

*Alessia Pini and Simone Vantini*

|          |                                               |     |
|----------|-----------------------------------------------|-----|
| 10.1     | Introduction                                  | 242 |
| 10.2     | Methodology                                   | 244 |
| 10.2.1   | Comparing Means of Two Functional Populations | 244 |
| 10.2.2   | Extensions                                    | 248 |
| 10.2.2.1 | Multiway FANOVA                               | 249 |
| 10.3     | Data Analysis                                 | 250 |
| 10.4     | Conclusion and Future Works                   | 256 |
|          | References                                    | 258 |

|           |                                                                                                            |            |
|-----------|------------------------------------------------------------------------------------------------------------|------------|
| <b>11</b> | <b>Modeling Spatially Dependent Functional Data by Spatial Regression with Differential Regularization</b> | <b>260</b> |
|           | <i>Mara S. Bernardi and Laura M. Sangalli</i>                                                              |            |
| 11.1      | Introduction                                                                                               | 260        |
| 11.2      | Spatial Regression with Differential Regularization for Geostatistical Functional Data                     | 264        |
| 11.2.1    | A Separable Spatiotemporal Basis System                                                                    | 265        |
| 11.2.2    | Discretization of the Penalized Sum-of-Square Error Functional                                             | 268        |
| 11.2.3    | Properties of the Estimators                                                                               | 271        |
| 11.2.4    | Model Without Covariates                                                                                   | 273        |
| 11.2.5    | An Alternative Formulation of the Model                                                                    | 274        |
| 11.3      | Simulation Studies                                                                                         | 274        |
| 11.4      | An Illustrative Example: Study of the Waste Production in Venice Province                                  | 278        |
| 11.4.1    | The Venice Waste Dataset                                                                                   | 278        |
| 11.4.2    | Analysis of Venice Waste Data by Spatial Regression with Differential Regularization                       | 279        |
| 11.5      | Model Extensions                                                                                           | 282        |
|           | References                                                                                                 | 283        |
| <b>12</b> | <b>Quasi-maximum Likelihood Estimators for Functional Linear Spatial Autoregressive Models</b>             | <b>286</b> |
|           | <i>Mohamed-Salem Ahmed, Laurence Broze, Sophie Dabo-Niang, and Zied Gharbi</i>                             |            |
| 12.1      | Introduction                                                                                               | 286        |
| 12.2      | Model                                                                                                      | 288        |
| 12.2.1    | Truncated Conditional Likelihood Method                                                                    | 291        |
| 12.3      | Results and Assumptions                                                                                    | 293        |
| 12.4      | Numerical Experiments                                                                                      | 298        |
| 12.4.1    | Monte Carlo Simulations                                                                                    | 298        |
| 12.4.2    | Real Data Application                                                                                      | 305        |
| 12.5      | Conclusion                                                                                                 | 312        |
| 12.A      | Appendix                                                                                                   | 313        |
|           | Proof of Proposition 12.A.1                                                                                | 313        |
|           | Proof of Theorem 12.1                                                                                      | 314        |
|           | Proof of Theorem 12.2                                                                                      | 317        |
|           | Proof of Theorem 12.3                                                                                      | 319        |
|           | Proof of Lemma 12.A.2                                                                                      | 322        |
|           | Proof of Lemma 12.A.3                                                                                      | 322        |
|           | Proof of Lemma 12.A.5                                                                                      | 323        |
|           | References                                                                                                 | 325        |



|           |                                                                                          |            |
|-----------|------------------------------------------------------------------------------------------|------------|
| <b>13</b> | <b>Spatial Prediction and Optimal Sampling for Multivariate Functional Random Fields</b> | <b>329</b> |
|           | <i>Martha Bohorquez, Ramón Giraldo, and Jorge Mateu</i>                                  |            |
| 13.1      | Background                                                                               | 329        |
| 13.1.1    | Multivariate Spatial Functional Random Fields                                            | 329        |
| 13.1.2    | Functional Principal Components                                                          | 330        |
| 13.1.3    | The Spatial Random Field of Scores                                                       | 331        |
| 13.2      | Functional Kriging                                                                       | 332        |
| 13.2.1    | Ordinary Functional Kriging (OFK)                                                        | 332        |
| 13.2.2    | Functional Kriging Using Scalar Simple Kriging of the Scores ( $FK_{SK}$ )               | 333        |
| 13.2.3    | Functional Kriging Using Scalar Simple Cokriging of the Scores ( $FK_{CK}$ )             | 333        |
| 13.3      | Functional Cokriging                                                                     | 336        |
| 13.3.1    | Cokriging with Two Functional Random Fields                                              | 336        |
| 13.3.2    | Cokriging with $P$ Functional Random Fields                                              | 338        |
| 13.4      | Optimal Sampling Designs for Spatial Prediction of Functional Data                       | 340        |
| 13.4.1    | Optimal Spatial Sampling for OFK                                                         | 341        |
| 13.4.2    | Optimal Spatial Sampling for $FK_{SK}$                                                   | 341        |
| 13.4.3    | Optimal Spatial Sampling for $FK_{CK}$                                                   | 342        |
| 13.4.4    | Optimal Spatial Sampling for Functional Cokriging                                        | 343        |
| 13.5      | Real Data Analysis                                                                       | 344        |
| 13.6      | Discussion and Conclusions                                                               | 348        |
|           | References                                                                               | 348        |

### **Part III Spatio–Temporal Functional Data 351**

|           |                                                                        |            |
|-----------|------------------------------------------------------------------------|------------|
| <b>14</b> | <b>Spatio–temporal Functional Data Analysis</b>                        | <b>353</b> |
|           | <i>Gregory Bopp, John Ensley, Piotr Kokoszka, and Matthew Reimherr</i> |            |
| 14.1      | Introduction                                                           | 353        |
| 14.2      | Randomness Test                                                        | 355        |
| 14.3      | Change-Point Test                                                      | 359        |
| 14.4      | Separability Tests                                                     | 362        |
| 14.5      | Trend Tests                                                            | 365        |
| 14.6      | Spatio–Temporal Extremes                                               | 369        |
|           | References                                                             | 373        |

|           |                                                                                             |            |
|-----------|---------------------------------------------------------------------------------------------|------------|
| <b>15</b> | <b>A Comparison of Spatiotemporal and Functional Kriging Approaches</b>                     | <b>375</b> |
|           | <i>Johan Strandberg, Sara Sjöstedt de Luna, and Jorge Mateu</i>                             |            |
| 15.1      | Introduction                                                                                | 375        |
| 15.2      | Preliminaries                                                                               | 376        |
| 15.3      | Kriging                                                                                     | 378        |
| 15.3.1    | Functional Kriging                                                                          | 378        |
| 15.3.1.1  | Ordinary Kriging for Functional Data                                                        | 378        |
| 15.3.1.2  | Pointwise Functional Kriging                                                                | 380        |
| 15.3.1.3  | Functional Kriging Total Model                                                              | 381        |
| 15.3.2    | Spatiotemporal Kriging                                                                      | 382        |
| 15.3.3    | Evaluation of Kriging Methods                                                               | 384        |
| 15.4      | A Simulation Study                                                                          | 385        |
| 15.4.1    | Separable                                                                                   | 385        |
| 15.4.2    | Non-separable                                                                               | 390        |
| 15.4.3    | Nonstationary                                                                               | 391        |
| 15.5      | Application: Spatial Prediction of Temperature Curves in the Maritime Provinces of Canada   | 394        |
| 15.6      | Concluding Remarks                                                                          | 400        |
|           | References                                                                                  | 400        |
| <b>16</b> | <b>From Spatiotemporal Smoothing to Functional Spatial Regression: a Penalized Approach</b> | <b>403</b> |
|           | <i>Maria Durban, Dae-Jin Lee, María del Carmen Aguilera Morillo, and Ana M. Aguilera</i>    |            |
| 16.1      | Introduction                                                                                | 403        |
| 16.2      | Smoothing Spatial Data via Penalized Regression                                             | 404        |
| 16.3      | Penalized Smooth Mixed Models                                                               | 407        |
| 16.4      | P-spline Smooth ANOVA Models for Spatial and Spatiotemporal data                            | 409        |
| 16.4.1    | Simulation Study                                                                            | 411        |
| 16.5      | P-spline Functional Spatial Regression                                                      | 413        |
| 16.6      | Application to Air Pollution Data                                                           | 415        |
| 16.6.1    | Spatiotemporal Smoothing                                                                    | 416        |
| 16.6.2    | Spatial Functional Regression                                                               | 416        |
|           | Acknowledgments                                                                             | 421        |
|           | References                                                                                  | 421        |
|           | <b>Index</b>                                                                                | <b>424</b> |

## List of Contributors

**Ana M. Aguilera**

University of Granada  
Department of Statistics and  
Operational Research  
Spain

**Mohamed-Salem Ahmed**

University of Lille  
France

**Mara S. Bernardi**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

**Gregory Bopp**

Pennsylvania State University  
Department of Statistics  
USA

**Martha Bohorquez**

National University of Colombia  
Department of Statistics  
Colombia

**Laurence Broze**

University of Lille  
France

**María del Carmen Aguilera Morillo**

Universitat Politècnica de València  
Department of Statistics and  
Operational Research and Quality  
Spain

**Sophie Dabo-Niang**

University of Lille  
France

**Maria Durban**

Universidad Carlos III  
Department of Statistics  
Spain

**John Ensley**

Pennsylvania State University  
Department of Statistics  
USA

**Maria Franco-Villoria**

Università di Modena e Reggio Emilia  
Department of Economics  
“Marco Biagi”  
Italy

**Zied Gharbi**

University of Lille  
France

**Ramón Giraldo**

National University of Colombia  
Department of Statistics  
Colombia

**Alberto Guadagnini**

Politecnico di Milano  
Department of Civil and  
Environmental Engineering  
Italy

and

The University of Arizona  
Department of Hydrology and  
Atmospheric Sciences  
USA

**Rosaria Ignaccolo**

Università degli Studi di Torino  
Dipartimento di Economia e Statistica  
“Cognetti de Martiis”  
Italy

**Antonio Irpino**

University of Campania “Luigi  
Vanvitelli”  
Department of Mathematics and  
Physics  
Italy

**Piotr Kokoszka**

Colorado State University  
Department of Statistics  
USA

**Sara Sjöstedt de Luna**

Umeå University  
Department of Mathematics and  
Mathematical Statistics  
Sweden

**Dae-Jin Lee**

BCAM–Basque Center for Applied  
Mathematics  
Spain

**Claude Manté**

Université du Sud Toulon-Var  
CNRS/INSU, IRD, MIO, Aix-Marseille  
Université  
France

**Jorge Mateu**

University Jaume I of Castellon  
Department of Mathematics  
Spain

**Alessandra Menafoglio**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

**Pascal Monestiez**

INRAE - Unité BioSP  
France

**David Nerini**

Université du Sud Toulon-Var  
CNRS/INSU, IRD, MIO, Aix-Marseille  
Université  
France

**Federica Passamonti**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

**Davide Pigoli**

King's College London  
UK

**Alessia Pini**

Università Cattolica del Sacro Cuore  
Department of Statistical Sciences  
Italy

**Cristian Preda**

Institute of Statistics and Applied  
Mathematics of the Romanian  
Academy  
Romania

**Matthew Reimherr**

Pennsylvania State University  
Department of Statistics  
USA

**Elvira Romano**

University of Campania “Luigi  
Vanvitelli”  
Department of Mathematics and  
Physics  
Italy

**Laura M. Sangalli**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

**Piercesare Secchi**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

and

CADS - Center for Analysis Decisions  
and Society  
Human Technopole  
Italy

**Johan Strandberg**

Umeå University  
Department of Statistics  
Sweden

**Camille Ternynck**

University of Lille  
France

**Baba Thiam**

University of Lille  
France

**Vincent Vandewalle**

University of Lille  
France

**Simone Vantini**

Politecnico di Milano  
MOX - Department of Mathematics  
Italy

**Valeria Vitelli**

University of Oslo  
Oslo Center for Biostatistics and  
Epidemiology  
Department of Biostatistics  
Norway

**Anne-Françoise Yao**

Université Clermont-Auvergne  
France

## Foreword

Functional data analysis (FDA) is a branch of statistics that analyses data providing information about curves, surfaces, or anything else varying over a continuum. In its most general form, under an FDA framework each sample element is a function. The continuum over which these functions are defined is often time, but may also be spatial location, wavelength, probability, etc. In the 20 years since the first books and papers on this topic, this field of statistics has received the attention and encouragement of researchers in statistics and many applied disciplines and has become an important and dynamic area of modern statistics. Topics that have been covered include descriptive techniques, statistical inference, multivariate and non-parametric methods, regression, generalized linear models, time series, and spatial statistics.

Modern technology has made it possible to obtain large spatial and spatiotemporal data sets, and poses the challenge of statistical modeling of such data. The combination of spatial statistics with FDA has emerged as a key approach. This book presents new theories and methods to define, describe, characterize, and model functional data indexed in spatial or spatio-temporal domains. The main focus is on functional data obtained under a geostatistical framework, where the domain is fixed and continuous. Specific topics considered include kriging, clustering, regression, and optimal sampling, moving on in the last part of the book to spatiotemporal data. Some chapters also consider the treatment of functional data on lattices.

When we wrote our original book on the subject in the 1990s, James Ramsay and I hoped that we would encourage FDA as a way of thinking, not simply a collection of techniques. It has therefore been very pleasing to see the development of the field since then, and the abundance of research activity in the area has confirmed our hopes. I would urge readers and researchers to raise their sights above any specific methods, obviously important that they are, to ask how considering data as functions changes and broadens our statistical horizons. Particularly in the new era of data science, this concerns both what data can be collected and how they can be analyzed. I am sure this book will make a valuable contribution in helping them to do so.

November 2020

*Sir Bernard Silverman*  
University of Oxford  
University of Nottingham  
Oxford and Nottingham

## 1

## Introduction to Geostatistical Functional Data Analysis

*Jorge Mateu<sup>1</sup> and Ramón Giraldo<sup>2</sup>*

<sup>1</sup>*Department of Mathematics, University Jaume I of Castellón, Spain*

<sup>2</sup>*Department of Statistics, National University of Colombia, Bogotá, Colombia*

### 1.1 Spatial Statistics

Spatial statistics has developed rapidly during the last 30 years. We have seen an interesting progress both in theoretical developments and in practical studies. Some early applications were in mining, forestry, and hydrology. It seems to be honest to remark that the increasing availability of computer power and skillful computer software has stimulated the ability to solve increasingly complex problems. Clearly, these problems have some common elements: they were all of a spatial nature. Some theory was available, for example the random function theory as developed by Yaglom and others in the 1960s. But that was largely insufficient to find generic solutions for the whole class of problems, and hence, the applications required a new theory. Thereupon some far-reaching theories have been developed: image reconstruction, Markov random fields, point process statistics, geostatistics, and random sets, to mention just a few. As a next stage, these theories were applied successfully to new disciplinary problems leading to modifications and extensions of mathematical and statistical procedures. We therefore notice a general scientific process that has occurred in the field of spatial statistics: well-defined problems with a common character were suddenly on the agenda, and data availability and intensive discussion with practical and disciplinary researchers resulted in new theoretical developments. Often, it is difficult to say which was first, and what followed, but we see different theoretical models developed for different applications.

Spatial statistics has hence emerged as an important new field of science. One of the peculiarities is its power for visualization. A common cold-water fear of many statisticians and mathematicians to analyze images, to communicate their

results by maps, and to have to trust information in pictures was overcome. It has led to interesting theories and better and more objective procedures for dealing with spatial variation. Following Wittgenstein, we could state that we needed some geniuses to tackle the obvious. Now, many results of a spatial statistical analysis could be communicated smoothly toward the nonstatistical audience, like a disciplinary scientist, a policy-maker, or an interested student. They, in turn, were able to judge whether a problem was solved, whether a policy measure was relevant or was inspired by the beautiful pictures expressing deep thoughts on relevant issues.

The role in policy-making may be once more stressed. It is known that many policy-makers are inclined to make a decision on the basis of a well developed, well organized, and well understandable figure. They find it (rightly so!) rather boring to use long lists of statistical data. But as political decisions affect us all, it puts another responsibility on the back of statisticians: to make statistically sound maps. It is often hard to say what that should be, but at the very least, we should be able to generate pictures, maps, and graphs that rely on good data and that show important aspects for decision-making.

In this way, spatial statistics has become a refreshing wind in statistics. We do not need to do well much longer on difficult equations, long lists of data, and tables with simulated controlled scenarios. But, to be clear on the back of all these nice pictures a sound science with sometimes difficult and tedious derivations and deep thoughts are still required to make serious progress.

Spatial statistics recognizes and exploits the spatial locations of data when designing for, collecting, managing, analyzing, and displaying such data. Spatial data are typically dependent, for which there are classes of spatial models available that allow process prediction and parameter estimation. Spatially arranged measurements and spatial patterns occur in a surprisingly wide variety of scientific disciplines. The origins of human life link studies of the evolution of galaxies, the structure of biological cells, and settlement patterns in archaeology. Ecologists study the interactions among plants and animals. Foresters and agriculturalists need to investigate plant competition and account for soil variations in their experiments. The estimation of rainfall and of ore and petroleum reserves is of prime economic importance. Rocks, metals, and tissue and blood cells are all studied at a microscopic level. Geology, soil science, image processing, epidemiology, crop science, ecology, forestry, astronomy, atmospheric science, or simply any discipline that works with data collected from different spatial locations, need to develop models that indicate when there is dependence between measurements at different locations. Spatiotemporal variability is a relatively new area within Spatial Statistics, which explains the scarcity of space-time statistical tools 20 years ago. There has been a growing realization in the last decade that knowing where data were observed could help enormously in answering the substantive questions that precipitated their collection. One of the most powerful



tools for spatial data analysis is the map. For example, in military applications, the battlespace is mapped for command and control. The sensors are both *in situ* and remote, and they generate spatially distributed data of many different kinds. Producing a statistically optimal map, together with measures of map uncertainty, which is always up to date, is a complicated task. Once these types of statistical problems are solved, a geographic information system, or GIS, is well suited to forming the decision-making maps.

Spatial statistics can be considered a natural generalization of signal processing to higher dimensions. In traditional signal processing, one has a signal dependent on a scalar variable  $t$ , which may belong to a discrete set or which may be continuous. Spatial statistics is concerned with cases in which  $t$  is a multidimensional index of dimension  $d > 1$ . In most practical examples  $d = 2$ , though much of the basic theory and methodology is the same whatever the dimension. Although the models and methods of spatial statistics have not developed as rapidly as those for one-dimensional signal processing, there have nevertheless been substantial new developments in recent years. Standard and modern references on spatial statistics include the books of [1–4] among others.

Following Cressie [5], spatial data can be thought of as resulting from observations on the stochastic process  $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ , where  $D$  is possibly a random set in  $\mathbb{R}^d$ . If we believe that the roots of statistical science are in data, we can classify spatial areas according to the type of observations encountered. Thus, (i) if  $D$  is a fixed subset of  $\mathbb{R}^d$  and  $Z(\mathbf{s})$  is a random vector at location  $\mathbf{s} \in D$ , we are dealing with *geostatistical data*; (ii) if  $D$  is a fixed (regular or irregular) collection of countably many points of  $\mathbb{R}^d$  and  $Z(\mathbf{s})$  is a random vector at location  $\mathbf{s} \in D$ , we are dealing with *lattice data*; (iii) if  $D$  is a point process in  $\mathbb{R}^d$  and  $Z(\mathbf{s})$  is a random vector at location  $\mathbf{s} \in D$ , we are dealing with *point patterns*; (iv) if  $D$  is a point process in  $\mathbb{R}^d$  and  $Z(\mathbf{s})$  is itself a random set, we are dealing with *spatial objects*. Geostatistical-type problems are distinguished most clearly from lattice- and point-pattern-type problems by the ability of the spatial index  $\mathbf{s}$  to vary continuously over a subset of  $\mathbb{R}^d$ . A space-time process can be denoted by  $\{Z(\mathbf{s}, t) : \mathbf{s} \in D(t), t \in T\}$ , where each of  $Z$ ,  $D$ , and  $T$  is possibly random.

Spatial statistics is one of the major methodologies of *environmental statistics*. Its applications include producing spatially smoothed or interpolated representations of air pollution fields, calculating regional average means or regional average trends based on data at a finite number of monitoring stations, and performing regression analyses with spatially correlated errors to assess the agreement between observed data and the predictions of some numerical model. The notion of proximity in space is implicitly or explicitly present in the environmental sciences. Proximity is a relative notion, relative to the spatial scale of the scientific investigation. When a spatial dimension is present in an environmental study, the statistician's job is to create a statistical framework within which one carries out

defensible inferences on processes and parameters of interest. These modeling and inference strategies are not always easy to do, but are never impossible. If statistics is to continue to be the broker of variability, it must address difficult questions such as those found in the environmental sciences, otherwise, it will become marginalized as a discipline. Problems in the environmental sciences are inherently spatial (and temporal), observational in nature, and have experimental units that are highly variable.

In the last decade, spatial statistics has undergone enormous development in the area of statistical modeling. It started slowly, building from models that were purely descriptive of spatial dependence. Then, it became apparent that the process of interest was usually hidden by measurement error and that the principal goal should be inference on the hidden process from the noisy data. It has only been in the last few years that the full potential for hierarchical spatial statistical modeling has been glimpsed. There is an enormous amount of flexibility in hierarchical statistical models, such as the opportunity to account for nonlinearities. Their attractive feature is that at each level of the hierarchy, the model specification is simple, yet globally, the model can be quite complex. This approach could be summarized as a model locally, analyze globally.

Applications of spatial statistics cover many areas. Much of the original impetus for the subject was driven by geostatistics. It was in this context that the technique of kriging, optimal least squares interpolation over a random spatial field, was originally developed. In recent years, the applications of spatial statistics have increased enormously, with particularly fruitful applications in the environmental and ecological sciences. A typical problem is the sampling of a pollution field, such as ozone in the atmosphere or toxic chemicals in rivers and lakes. Another example is the use of meteorological measurements in studies of global climate change. In these fields, as in geostatistics, the objective may be to interpolate spatially between measurements, but there are also other objectives which may be quite different. Spatial statistics has also found applications in such diverse fields as sociology, for example social networks theory and financial economics.

The usual approach in geostatistics is based on an assumption that the spatial random field is stationary and isotropic. In the original geophysical applications which motivated the development of the field, this assumption was often justified by the fact that with sparse data, there was no reasonable alternative. A further point is that many geostatistical applications involved only one measurement at each site (or equivalently, only one replication of the random field) so there was no way of determining the complete spatial covariance function without some kind of stationarity assumption. In modern environmental applications, however, there are very often enough monitoring stations to go beyond such assumptions, and with multiple observations per site, it is also possible to estimate the covariance between any pair of sites without assuming stationarity across the field. Another

consideration is that very often, simple topography makes a stationary assumption implausible. Therefore, there are by now many reasons to go beyond a stationary model. In spite of this obvious need for nonstationary models; however, there is not, as yet, a wide variety of approaches to the problem.

Environmental issues have brought atmospheric science to the center of science and technology, where it now plays a key role in shaping national and international policy. Weather prediction plays a significant role in the planning of human affairs. Further, a broader appreciation of the role of weather and climate impacts on the environment of the planet has now led to nearly universal concern regarding potential climate change, its causes, impacts, and possible remedying. A large variety of statistical methods are used routinely in the atmospheric sciences. For example, techniques of multivariate time series are especially common. These include multivariate autoregressive, moving average models and Kalman filtering. Statistical methods for spatial data are also standard. A major tool in the analysis of space-time data is empirical orthogonal functions (EOF). Virtually, all atmospheric and oceanographic processes (e.g. wind, temperature, sea surface temperature, moisture) involve variability over space and time. One only needs examine the governing partial differential equations for wind processes, or their selected spatial-temporal averages, to see that mathematical and statistical descriptions of these dynamical processes depend on complicated temporal and spatial relationships. Furthermore, observations of geophysical processes typically include measurement errors and are often temporally and spatially incomplete, which may obscure the signal of interest.

In studies involving spatial data, it is seldom the case that data for only a single process are collected. Typically, there is a great expense associated with establishing spatial monitoring networks or other mechanisms of spatial data collection (e.g. satellites) and so measurements are usually made on two or more variables. Thus, statistical techniques for multivariate spatial data are critical for effective modeling of spatial processes.

Lately, there has been a rich and growing literature on space-time modeling. Fundamentally, it is clear that in the absence of a temporal component, second-order geostatistical models can be used to represent spatial variability. These are descriptive in the sense that, although they model spatial correlation, there is no causative interpretation associated with them. Thus, for space-time modeling, the geostatistical paradigm assumes a descriptive structure for both space and time (i.e. covariance structures are directly specified). For example, one can extend the geostatistical kriging methodology for spatial processes by assuming that time is just another spatial dimension. Alternatively, one can treat time slices of a spatial field as variables and apply a multivariate or cokriging approach. Although these approaches have been successful in many applications, there are fundamental differences between space and time, and it is not likely

that realistic covariance structures can be specified that accurately capture the complicated dynamical processes as found in geophysical applications.

In the absence of a spatial component, there is a large class of time series models that could be used to represent the temporal variability. These are dynamic in the sense that they exploit the fact that time flows in only one direction, and so the state of the process at the current time is related to what happened at previous times. Thus, one might consider the space–time process as a collection of spatially correlated time series in continuous space, or on a spatial lattice. Although these approaches include dynamical structures, without a descriptive spatial component one lacks the ability to perform spatial prediction at locations without observations. If both temporal and spatial components are present, it is natural to combine the temporally dynamic state-space approach and the spatially descriptive approach. These models are referred to as space–time dynamic models.

Spatial interpolation is an essential feature of many GIS. It is a procedure for estimating values of a variable at unsampled locations. A map with isolines is usually the visual output of such a process and plays a crucial role in decision-making. Based on Tobler’s law of geography, which stipulates that observations close together in space are more likely to be similar than those farther apart, the development of models attempting to represent the way close observations are related can sometimes be very problematic. The approaches can be divergent and may therefore lead to very different results. As a consequence, an understanding of the initial assumptions and methods used is the key to the spatial interpolation process.

Surprisingly, when spatial interpolation tools are integrated within GIS, they are often implemented in such a way that users have no real choice in selecting the best possible methods, and if they do have a choice, required input parameters are sometimes fixed, without any possible way of modifying them. One reason for the frequent blind use of spatial interpolation methods, and spatial statistics in general, probably has its origins in teaching. Despite the large variety of its applications, the discipline has been confined to those fields where it has seen its major developments. The progress made in spatial statistics is therefore usually presented only in journals dedicated to statistics, mining, and petroleum engineering. As a consequence, GIS users who have a different technical background often do not have an in-depth knowledge of such spatial interpolation techniques. Furthermore, since the conventional tests used in basic statistics usually generate some kind of categorical answer, the prerequisite experience and statistical knowledge necessary for the proper use of spatial interpolation techniques are often discouraging to this type of users. Nevertheless, during the last few years, the diversity of the applications of these methods has encouraged the publication of new books and new case studies and has stimulated a number of conferences on the subject.

## 1.2 Spatial Geostatistics

This section has been partially taken and summarized in parts from [6], intending to provide a brief overview to spatial geostatistics. The reader is referred to [6] for further and more complete details.

### 1.2.1 Regionalized Variables

Geostatistics can be defined as the study of regionalized phenomena, that is, phenomena that stretch across space and which have a certain spatial organization or structure. However, geostatistics is not applied to the regionalized phenomenon as such, which is a physical reality, but to a mathematical description of that reality, that is, a numerical function called *regionalized variable* or *regionalization*, defined in a geographical space, which is supposed to correctly represent and measure that phenomenon.

In order to delve deeper into the concept of regionalized variable, let us imagine we are interested in a feature of a given phenomenon that spans across space and that several measurements are taken in a domain  $D$  at a given moment in time. If the measurements are taken on objects or similar, the objects sampled can be considered a subset of a larger collection of objects, as many more measurements could have been taken, but were not for many possible reasons. If the observations were made at certain points in the domain, infinite measurements could be taken.

When  $\mathbf{s}$  spans across the domain under study,  $D$ , the set  $\{z(\mathbf{s}), \mathbf{s} \in D\}$ , is called a regionalized variable or regionalization, the set  $\{z(\mathbf{s}_i), i = 1, 2, 3, \dots\}$  being a collection of values of the regionalized variable, and each value of that collection being a regionalized value.

It is true that a deterministic approach can be employed to describe or model a regionalized phenomenon and obtain an accurate assessment of the values of the regionalization on the basis of a limited number of observations. However, this requires in-depth knowledge of the origin of the phenomenon and the physical or mathematical laws that govern the evolution of the regionalized variable. Furthermore, many of the regionalized phenomena that are usually studied are so complex that a deterministic approach can only partially portray them. That is why the deterministic approach is discarded and the probabilistic approach, which permits modeling both the knowledge of and also the uncertainty surrounding the regionalized random phenomenon, is adopted.

### 1.2.2 Random Functions

From a probabilistic perspective, the regionalized value can be seen as the result of a random mechanism, resulting in a *random variable* (r.v.). If the regionalized

values at all the points in the domain  $D$  are considered, it can be seen as a reality of an infinitely large set of r.v.s, one at each point in the domain, which is known as spatial *random function* (synonyms: stochastic process, random field).

When  $\mathbf{s}$  spans across the domain under study,  $D$ , we have a family of r.v.s,  $\{Z(\mathbf{s}), \mathbf{s} \in D\}$ , which constitutes a spatial random field (r.f.).

This methodological decision is one of the cornerstones of geostatistics: the regionalized variable is interpreted as a realization of a spatial r.f. At this point, we must state that the regionalized variable is often highly locally irregular (which makes it impossible to represent using a deterministic mathematical function) and has a certain spatial organization or structure. The probabilistic approach, or probabilistic geostatistics, which interprets the regionalized variable as a realization of a r.f., can take into account all the aspects of regionalization mentioned above, because, as stated in page 55 of [7]:

- i) At each location  $\mathbf{s}$ ,  $Z(\mathbf{s})$  is a r.v. (hence, the erratic aspect).
- ii) For any given set of points  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k$ , the r.v.s  $Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_k)$  are linked by a network of spatial correlations responsible for the similarity of the values they take (hence the structured aspect).

Let  $Z(\mathbf{s})$  be a r.f. and let us consider the set of points  $(\mathbf{s}_1, \dots, \mathbf{s}_k)$ . Then, the r.f.  $Z(\mathbf{s})$  is characterized by its  $k$ -dimensional distribution function. The set of  $k$ -dimensional distribution functions for all values of  $k$  and all possible choices of  $(\mathbf{s}_1, \dots, \mathbf{s}_k)$  in the domain is called the *spatial law of probability*.

For a given r.f.,  $Z(\mathbf{s})$ , the  $k$ -dimensional distribution function  $F(z(\mathbf{s}_1), \dots, (z(\mathbf{s}_k))$ :  $\mathbb{R}^d \rightarrow [0, 1]$  is defined as

$$F_{Z_1, \dots, Z_k}(z(\mathbf{s}_1), \dots, (z(\mathbf{s}_k)) = P[Z(\mathbf{s}_1) \leq z(\mathbf{s}_1), \dots, Z(\mathbf{s}_k) \leq z(\mathbf{s}_k)]. \quad (1.1)$$

In linear geostatistics, it is enough to know the first two moments of the distribution of  $Z(\mathbf{s})$ . What is more, in most practical applications, the available information does not allow to infer higher-order moments.

The expectation, expected value or first-order moment of a r.f. is defined as a nonrandom function of  $\mathbf{s}$  that coincides at each point with the expectation of the r.v. at that point  $\mu(\mathbf{s}) = E(Z(\mathbf{s}))$ , where  $\mu(\mathbf{s}_i) = E(Z(\mathbf{s}_i))$ ,  $\forall i \in \mathbb{N}$ . It is also called the drift of the r.f., especially when it varies with location.

The variance of a r.f. is defined as a nonrandom function of  $\mathbf{s}$  that coincides at each point with the variance of the r.v. at that point, i.e.  $V(\mathbf{s}) = V(Z(\mathbf{s}))$ , where  $V(\mathbf{s}_i) = V(Z(\mathbf{s}_i))$ ,  $\forall i \in \mathbb{N}$ .

The covariance function of a r.f. is defined as a nonrandom function of  $\mathbf{s}_i$  and  $\mathbf{s}_j$ , such that for any pair of values  $(\mathbf{s}_i, \mathbf{s}_j)$  coincides with the covariance between the r.v. at those two points

$$C(\mathbf{s}_i, \mathbf{s}_j) = C(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = E((Z(\mathbf{s}_i) - \mu(\mathbf{s}_i))(Z(\mathbf{s}_j) - \mu(\mathbf{s}_j))), \quad \forall \mathbf{s}_i, \mathbf{s}_j \in D. \quad (1.2)$$

- null hypothesis testing for functional data 256
- object oriented data analysis (OODA) 30
- object-oriented spatial statistics (O2S2) 29, 30, 123
- OKFD *see* ordinary kriging for functional data (OKFD)
- OLSs. *see* ordinary least squares (OLSs)
- OODA *see* object oriented data analysis (OODA)
- operatorial viewpoint to kriging 42–44
- ordinary functional kriging (OFK) 332–333, 375
- ordinary kriging for functional data (OKFD) 21, 59, 63, 378–380, 394, 400
- ordinary least squares (OLSs) 37
- O2S2. *see* object-oriented spatial statistics (O2S2)
- partial differential equation (PDE) 260
  - partial least square (PLS) 186, 404
  - particle-size curves (PSCs) 109, 110, 113, 116, 119
  - particle-size densities (PSDs) 112, 115, 121, 122, 124
  - Partitioning Heterogeneity Index (PHI) 164
  - PCA *see* principal component analysis (PCA)
  - PDFs *see* probability density functions (PDFs)
  - penalized functional spatial regression model (PFSRM) 416, 418–419
  - PFSRM. *see* penalized functional spatial regression model (PFSRM)
  - PHI *see* Partitioning Heterogeneity Index (PHI)
  - PLS *see* partial least square (PLS)
  - point patterns 3
  - pointwise consistent 247
  - pointwise error rate 247
  - pointwise functional kriging predictor (PWFK) 380–381
  - positive definitive matrices, Riemannian manifold-valued random fields in 47–49
  - precipitation observations 82–85
    - fitting variogram model 83
    - making 83–85
  - principal component analysis (PCA) 159, 160, 186, 355, 404
  - principal component analysis for curves
    - Karhunen–Loève decomposition 74–76
    - sample 76–78
  - probability density functions (PDFs) 105, 110–112, 123
    - spatial functional data analysis for 128–150
  - product-sum model 383
  - proximity 3
  - PSCs *see* particle-size curves (PSCs)
  - PSDs *see* particle-size densities (PSDs)
  - P-spline ANOVA models
    - definition 408
    - simulation study 411–413
    - smoothed spatiotemporal
      - at four selected location 416, 418, 420
      - space–time interaction 416–418
    - for spatial and spatiotemporal data
      - B-spline model matrix 410
      - main effects 409
      - multidimensional smooth model 409
      - two-dimensional basis 410
      - two-way interactions 410
  - quasi-maximum likelihood (QML) 288
  - quasi-maximum likelihood estimator (QMLE) 288
  - R, air pollution data 64–69
  - random functions 7–9
    - intrinsic 12
  - random variable (r.v.) 7–8
  - real data analysis 141–149
    - compositional–functional approach 145–148
    - SDA distributional approach 143–145
  - regionalization *see* regionalized variable
  - regionalized variable 7, 8
  - regression kriging 55
  - relative scale 106
  - REML *see* Restricted Maximum Likelihood (REML)
  - reproducing Kernel Hilbert spaces (RKHSs) 42
  - residual kriging for functional data (ResKFD) 58–59, 61
  - residual kriging 55
  - residual kriging, in Riemannian manifold-valued random fields 45–47
  - ResKFD *see* residual kriging for functional data (ResKFD)
  - Restricted Maximum Likelihood (REML) 61

- Riemannian manifold-valued random fields,
  - kriging for
  - application to positive definite matrices 47–49
  - local tangent space approximation, validity of 49–53
  - residual kriging 45–47
- RKHSs *see* reproducing Kernel Hilbert spaces (RKHSs)
- root mean square errors (RMSE) 276
- S-ANOVA model 412–413
- scale invariance 106
- SDA *see* symbolic data analysis (SDA)
- second-order isotropically stationary 377
- second-order stationarity 10–11
  - global 32–33, 35, 36
  - strongly 22
- second-order stationary 377
- second-order stationary variogram 14
- semivariogram 13, 14, 17
- separability tests
  - dependence structure in data 362
  - dimension reduction 364
  - “flip-flop” estimation procedure 362
  - functions, estimation of 362
  - Monte-Carlo algorithm 364
  - p*-values for norm-based 364–365
  - temporal covariance function
    - estimation 363
- separable covariance function 13
- separable model 383
- separable spatiotemporal basis system
  - construction on Venice domain 265, 266
  - cubic B-spline basis 268
  - linear basis function 267–268
  - triangulation of Venice province 265, 267
  - UTM coordinate system 267
- SFPCA *see* simplicial functional principal component analysis (SFPCA)
- SFS *see* spatial functional statistics (SFS)
- SHI *see* Subsampling Heterogeneity Index (SHI)
- signal processing 3
- simplicial functional principal component analysis (SFPCA) 142
- smoothing spatiotemporal data
  - air pollution dataset in Spain
  - P-spline ANOVA model 416–418
  - raw data set 415–416
  - spatial functional regression 416, 418–420
- P-spline functional spatial regression 413–415
- via penalized regression 404
  - B-spline basis 405–406
  - and smooth mixed models 407–408
  - splines or P-splines 405
  - unpenalized regression 405
- Sobolev spaces, trace-cariogram in 37–39
- spatial autocorrelation 139
- spatial autoregressive (SAR) model 313
  - QML estimators for 288
  - for real-valued data 287
  - structure of 287–288
- spatial covariance, estimation of 81–82
- spatial covariance function 31
- spatial dependence, in drift estimation
  - 61–62
  - drift selection 62
- spatial functional data analysis, for probability
  - density functions 128–150
  - density function as compositional
    - functional data, features of 131–135
  - density functions as distributional data,
    - features of 135–138
  - functional data analysis 130–138
  - georeferenced density functions, spatial
    - association measures for 138–141
    - spatial association measures for density
      - functions, spatial cluster
        - identification by 139–141
  - real data analysis 141–149
    - compositional–functional approach 145–148
    - SDA distributional approach 143–145
    - symbolic data analysis 130–138
- spatial functional random process
  - definition 376–377
  - second-order isotropically stationary 377
  - second-order stationary 377
- spatial functional statistics (SFS) 128
- spatial geostatistics
  - random functions 7–9
  - recognized variables 7
  - stationarity and intrinsic hypothesis 9–12
- spatial homogeneity 9
- spatial interpolation 6
- spatial law of probability 8
- spatial locations 375
- spatially dependent functional data modeling



- advantages 260
- extensions 282–283
- numerical analysis techniques 260
- simulation studies (*see* spatiotemporal regression with partial differential equation regularization (ST-PDE))
- spatial regression with differential regularization for geostatistical functional data 264–265
  - alternative formulation of model 274
  - discretization of spatial and temporal penalization terms 269–271
  - model without covariates 273–274
  - separable spatiotemporal basis system 265–268
  - spatiotemporal mean and covariance structures of estimator 271–273
- Venice waste data (*see* Venice waste dataset)
- spatially dependent functional data, nonparametric algorithm for 211–212, 236, 238
- Bagging Voronoi
  - clustering 221–223
  - dimensional reduction 223–231
  - regression 231–237
  - strategy 216–218
- case study 238
- Erlang, case study in city management
  - advantage of 212
  - data preprocessing 214–216
  - data set 214
  - measurements 212–213
  - reference signals identification 212–213
- Local Indicators of Spatial Association (LISA) 238
- spatially distributed functional data, nonparametric statistical analysis of 175–207
- large sample properties 178–181
  - uniform almost complete convergence 180–181
- numerical results
  - bandwidth selection procedure 184–185
  - simulation study 185–193
- prediction 181–184
- preliminary results 194–207
- spatially stationary covariance function 12–13
- spatial objects 3
- spatial statistics 1–6
- spatial stochastic process 20
- spatiotemporal covariance models 17–18
- spatio–temporal functional data analysis 353–354
  - change-point test 354, 359–361
  - computation of probabilities of heat waves 369–372
  - geophysical data 354
  - prototypical example 353–354
  - randomness test 354
    - Box–Ljung test 355
    - for daily temperature observations 356–357
  - Karhunen–Loève (KL) expansion 355
  - locations of 20 weather stations 356, 357
  - null hypothesis 355
  - 14 Russian weather stations near Moscow 358
  - temporal dimension reduction 355
  - using maximal lags 358
- separability tests
  - dependence structure in data 362
  - dimension reduction 364
  - “flip-flop” estimation procedure 362
  - functions, estimation of 362
  - Monte-Carlo algorithm 364
  - p*-values for norm-based 364–365
  - temporal covariance function estimation 363
- trend tests 365–369
- spatiotemporal geostatistics 12–18
- spatiotemporal covariance models 17–18
- spatiotemporal kriging 16–17
- spatiotemporal kriging
  - metric model 383–384
  - ordinary 17
  - product-sum model 383
  - separable model 383
  - simple 16–17
  - Sp.T. ordinary kriging predictor 382–383
  - Sp.T. universal kriging predictor 384
  - universal 17
- spatiotemporal regression with partial differential equation regularization (ST-PDE)
  - comparative advantage of 276
  - implementation of 276
  - simulation with and without covariates 276–278

- spatiotemporal regression with partial differential equation regularization (ST-PDE) (*contd.*)
  - test function 274–276
  - via extensive simulation studies 274
- spatiotemporal (Sp.T.) stochastic process 376
- spatiotemporal variability 2
- spatiotemporal variogram 13
- spectral density function 15
- spectral distribution function 15
- stationarity and intrinsic hypothesis 9–12
  - in the broad sense 10–11
  - in the strict sense 10
- stationary covariance function 13
  - spatially 12–13
  - temporarily 13
- statistical modeling 4
- ST-PDE. *see* spatiotemporal regression with partial differential equation regularization (ST-PDE)
- strictly stationary variogram 14
- strongly second-order stationarity 32
- Subsampling Heterogeneity Index (SHI) 164
- symbolic data analysis (SDA) 129–138, 141
  - density function as compositional functional data, features of 131–135
  - density functions as distributional data, features of 135–138
  - distributional approach 143–145
- symmetrical stationary covariance function 15
- temperatures profiles, nonstationary prediction of 39–42
- temporarily stationary covariance function 13
- Tobler's First Law of Geography 139
- Tobler's law of geography 6
- Total Average Variance (TAV) 226
- trace approach to kriging prediction, in Hilbert space 33–42
  - ordinary kriging 33–36
  - universal kriging 33–36
- trace-cariogram, in Sobolev spaces 37–39
- trace-covariogram 31–32, 35, 36, 111
- trace-semivariogram 379, 394–396
- trace-variogram 62, 65, 66, 111, 157, 341
- trend tests 365–369
- truncated conditional likelihood method 291–293
- two-stage least squares (2SLS) estimation method 287
- UKFD *see* universal kriging for functional data (UKFD)
- unadjusted *p*-value functions
  - definition of 245, 247
  - inferential properties 247
- uncertainty assessment 116–118
- uncertainty evaluation 62–63
- uniform almost complete convergence 180–181
- universal kriging for functional data (UKFD) 56–58, 394, 399
- universal transverse mercator (UTM) coordinate system 267
- US Census 141
- variance of random field 8
- variogram 85
  - covariogram 17
  - empirical 35
  - fitting model 83
  - intrinsically stationary 14
  - of random function 9
  - second-order stationary 14
  - semivariogram 13, 14, 17
  - spatiotemporal 13
  - strictly stationary 14
  - trace-variogram 111, 157
- Venice waste dataset
  - analysis by regression with differential regularization 279–282
  - annual per capita municipal waste 278
  - beds in accommodation facilities 279
  - spatial domain of 260, 261
  - temporal evolution of yearly per capita production 261–263
  - urbanized areas 279
- World Meteorological Organization (WMO) 356