The titles published in this series are listed at the end of this volume.
## Contents

Foreword .......................................................... ix
Organizing Committee and International Scientific Committee ............ xi
Contributors ................................................................ xiii

### Part I  Environment and Health

Geostatistical Analysis of Health Data: State-of-the-Art and Perspectives  
*P. Goovaerts* ........................................................ 3

Early Detection and Assessment of Epidemics by Particle Filtering  
*C. Jégat, F. Carrat, C. Lajaunie and H. Wackernagel* ...................... 23

Improvement of Forecast Noise Levels in Confined Spaces by Means of Geostatistical Methods  
*G. A. Degan, D. Lippiello, M. Pinzari and G. Raspa* ..................... 37

Geostatistical Modeling of Environmental Sound Propagation  
*O. Baume, H. Wackernagel, B. Gauvreaux, F. Junker, M. Bérengier and J.-P. Chilès* ..................................................... 45

Geostatistical Estimation of Electromagnetic Exposure  
*Y. O. Isselmou, H. Wackernagel, W. Tabbara and J. Wiart* ............. 59

How Spatial Analysis Can Help in Predicting the Level of Radioactive Contamination of Cereals  
*C. Mercat-Rommens, J.-M. Metivier, B. Briand and V. Durand* ........... 71

Stochastic Modelling Applied to Air Quality Space-Time Characterization  
*A. Russo, R. M. Trigo and A. Soares* ....................................... 83

Automatic Mapping Algorithm of Nitrogen Dioxide Levels from Monitoring Air Pollution Data Using Classical Geostatistical Approach: Application to the French Lille City  
*G. Cardenas and E. Perdrix* ............................................... 95
King Prawn Catch by Grade Category from an Economic and a Stock Management Perspective
*U. Mueller, L. Bloom, M. Kangas and N. Caputi* .......................... 103

Part II Hydrology

Machine Learning Methods for Inverse Modeling
*D. M. Tartakovsky, A. Guadagnini and B. E. Wohlberg* .............................. 117

Statistical Moments of Reaction Rates in Subsurface Reactive Solute Transport
*X. Sanchez-Vila, A. Guadagnini, M. Dentz and D. Fernández-Garcia* ............ 127

Including Conceptual Model Information when Kriging Hydraulic Heads
*M. Rivest, D. Marcotte and P. Pasquier* ................................. 141

Effect of Sorption Processes on Pump-and-Treat Remediation Practices Under Heterogeneous Conditions
*M. Riva, A. Guadagnini and X. Sanchez-Vila* ............................. 153

A Stochastic Approach to Estimate Block Dispersivities that Includes the Effect of Mass Transfer Between Grid Blocks
*D. Fernández-Garcia and J. J. Gómez-Hernández* ........................1 65

Fracture Analysis and Flow Simulations in the Roselend Fractured Granite
*D. Patriarche, E. Pili, P. M. Adler and J.-F. Thovert* ........................... 175

Assessment of Groundwater Salinisation Risk Using Multivariate Geostatistics
*A. Castrignanò, G. Buttafuoco and C. Giasi* ............................. 191

A MultiGaussian Kriging Application to the Environmental Impact Assessment of a New Industrial Site in Alcoy (Spain)
*J. R. Ilarri and J. J. Gómez-Hernández* .............................. 203

Hydrogeological Modeling of Radionuclide Transport in Heterogeneous Low-Permeability Media: A Comparison Between Boom Clay and Ieper Clay
*M. Huysmans and A. Dassargues* ................................. 211

Topological Kriging of Runoff
*J. O. Skøien and G. Blöschl* ................................. 221
Part III  Meteorology

Quantifying the Impact of the North Atlantic Oscillation on Western Iberia
R. M. Trigo ................................................................. 235

Monthly Average Temperature Modelling
M. Andrade-Bejarano .................................................... 247

Improving the Areal Estimation of Rainfall in Galicia (NW Spain)
Using Digital Elevation Information
J. M. M. Avalos and A. P. González .................................. 263

Identification of Inhomogeneities in Precipitation
Time Series Using Stochastic Simulation
A. C. M. Costa, J. Negreiros and A. Soares .................................. 275

Bayesian Classification of a Meteorological Risk Index
for Forest Fires: DSR
R.M. Durão, A. Soares, J.M.C. Pereira, J.A. Corte-Real
and M.F.E.S. Coelho .................................................. 283

Part IV  Remote Sensing

Super Resolution Mapping with Multiple Point Geostatistics
A. Boucher ................................................................. 297

Super-Resolution Mapping Using the Two-Point
Histogram and Multi-Source Imagery
P. M. Atkinson ............................................................... 307

Dating Fire Events on End of Season Maps of Burnt Scars
T. J. Calado and C. C. DaCamara ................................... 323

Influence of Climate Variability on Wheat Production in Portugal
C. Gouveia and R. M. Trigo ............................................... 335

Part V  Soil

Joint Simulation of Mine Spoil Uncertainty for Rehabilitation
Decision Making
R. Dimitrakopoulos and S. Mackie ...................................... 349

Temporal Geostatistical Analyses of N₂O Fluxes from Differently
Treated Soils
J. M. M. Avalos, A. Furon, C. Wagner-Riddle and A. P. González .......... 361
Zinc Baseline Level and its Relationship with Soil Texture in Flanders, Belgium
T. Meklit, M. V. Meirvenne, F. Tack, S. Verstraete, E. Gommeren and E. Sevens .......................... 373

Assessing the Quality of the Soil by Stochastic Simulation
A. Horta, J. Carvalho and A. Soares .................................. 385

Interpolation of Soil Moisture Content Aided by FDR Sensor Observations

Geostatistics for Contaminated Sites and Soils: Some Pending Questions
D. D’Or, H. Demougeot-Renard and M. Garcia .......................... 409

Evaluation of an Automatic Procedure Based on Geostatistical Methods for the Characterization of Contaminated Sediments
G. Raspa, C. Innocenti, F. Marconi, E. Mumelter and A. Salmeri ............... 421

Part VI    Methods

Nonlinear Spatial Prediction with Non-Gaussian Data: A Maximum Entropy Viewpoint
P. Bogaert and D. Fasbender ............................................ 445

Data Fusion in a Spatial Multivariate Framework: Trading off Hypotheses Against Information
D. Fasbender and P. Bogaert ............................................. 457

The Challenge of Real-Time Automatic Mapping for Environmental Monitoring Network Management
E. J. Pebesma, G. Dubois and D. Cornford ............................. 467

Geostatistical Applications of Spartan Spatial Random Fields
S. N. Elogne and D. T. Hristopulos ................................. 477

A New Parallelization Approach for Sequential Simulation
H.S. Vargas, H. Caetano and H. Mata-Lima ............................. 489

Clustering in Environmental Monitoring Networks: Dimensional Resolutions and Pattern Detection
D. Tuia, C. Kaiser and M. Kanevski .................................. 497
Foreword

geoENV: Ten Years Later

The environment has unquestionably become a key topic of focus and concern for today’s society, encompassing themes that include sustainable development, climate change, reduction of biological diversity, and carbon emissions along with the need for new energy paradigms. These themes are no longer the exclusive domain of academic and scientific exploration. They are now high-priority issues for governments and environmental agencies of all industrialized countries because of the tremendous effects on the industrialized and, to an even greater extent, developing world. Quantifying and predicting global environmental impacts and risks encompasses political, social, economic as well as technical dimensions, and is now an integral part of strategic planning for both governments and international organizations.

Geostatistics has become an important set of technical tools for environmental problem-solving, in particular spatial and temporal assessment of uncertainty of physical/environmental phenomena and related natural resources. Geostatistics has been applied to a variety of fields from the characterization of desertification, degradation of soil, air and water quality, to the evaluation of health and pollutant space-time relationships in the field of environmental epidemiology, and the assessment of climatic and meteorology for predicting the dynamic of natural phenomena.

Geostatistics for Environmental Applications (geoENV), a series of bi-annual conferences, was started in 1996 with the ambitious goal of bringing together the disparate geostatistical community to discuss ideas and methods regarding new and diverse applications in the environmental field. Thanks to everyone involved in the organization and scientific coordination of the conferences, first in Lisbon, then Valencia, Avignon, Barcelona, Neuchâtel and most recently in Rhodes, the geoENV international conferences and subsequent publication of selected papers have contributed to maintaining the high standards of scientific quality in approaching the diversity of new environmental modeling problems. Ten years after Lisbon, we are proud to see that geoENV has become a well-respected and well-supported scientific project.

This book marks the first decade of geoENV and reflects the status of the most up-to-date research in the field as presented in Rhodes. As in past years, scientists
who approach environmental problems with different methodological perspectives than those encountered in our field, were invited to present their methodologies at geoENV conferences, with the goal of enriching our own field through cross-fertilization with related fields and their approaches, concepts, tools and developments. Ricardo Trigo, the keynote speaker in Rhodes, introduced us to new methods for modeling climate change and assessing corresponding impacts on natural resources and human health. The additional two keynote papers from Pierre Goovaerts and Philippe Renard presented the state of the art of geostatistical applications in analyzing public health data and stochastic hydrology, respectively. The 42 papers of this book were presented in oral sessions of Methods, Environment and Health, Soil, Hydrology, Remote Sensing and Meteorology.

We would like to thank to all the authors and reviewers for their outstanding efforts and technical contributions to the present volume.

Amílcar Soares
Maria João Pereira
Roussos Dimitrakopoulos
Organizing Committee geoENV2006

Amilcar Soares, IST, Lisbon, Portugal
Denis Allard, INRA, Avignon, France
Hélène Demougeot-Renard, FSS International, Neuchâtel, Switzerland
Jaime Gómez-Hernández, UPV, Valencia, Spain
Maria João Pereira, IST, Lisbon, Portugal
Pascal Monestiez, INRA, Avignon, France
Phaedon Kyriakidis, UC Santa Barbara, Santa Barbara, USA
Philippe Renard, University of Neuchâtel, Switzerland
Pierre Goovaerts, BioMedware, Inc., USA
Roland Froidevaux, FSS International, Geneva, Switzerland
Roussos Dimitrakopoulos, BRC, U. Queensland
Xavier Sanches-Vila, UPC, Barcelona, Spain

International Scientific Committee geoENV2006

Aldo Fiori
Alberto Guadagnini
Alexander Boucher
Alfred Stein
Amilcar Soares
Carla Nunes
Carol Gotway Crawford
Chantal de Fouquet
Chris Loyd
Denis Allard
Denis Marcotte
Dick Brus
Dimitri D’Or
Edzer Pebesma
Geoffrey M. Jacquez
Hans Wackernagel
Harrie-Jan Hendricks
Helene Demougeot
Henrique G. Pereira
Hirota Saito
Jacques Rivoirard
Jaime Gomez-Hernandez
Jean Paul Chilés
Jennifer Mckinley
Jorge Sousa
José Almeida

Júlia Carvalho
Laura Guadagnini
Marc Van Merveinne
Margaret Oliver
Maria João Pereira
Maria-Theresa Schafmeister
Mohan Srivastava
Monica Riva
Muray Lark
Oy Leuangthong
Pascal Monestiez
Patrick Bogaert
Peter Atkinson
Peter Kettlewell
Phaedon Kiriakidis
Philippe Renard
Pierre Goovaerts
Ricardo Garcia-Herrera
Richard Webster
Roland Froidevaux
Roussos Dimitrakopoulos
Souheil Ezzedine
Vicente Serrano
Xavier Emery
Xavier Sanchez-Vila
Contributors

P.M. Adler
UPMC – Sisyphe, 4 place Jussieu, 75252 Paris cedex 05, France
padler@ccr.jussieu.fr

P.M. Atkinson
School of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ, UK
pma@soton.ac.uk

O. Baume
Department of Environmental Sciences, Wageningen University, The Netherlands
olivier.baume@wur.nl

M. Andrade-Bejarano
School of Biological Sciences, Statistics Section, Harry Pitt Building, The University of Reading, RG6 6FN, Reading, UK
snr02ma@reading.ac.uk

M. Bérengier
Section Acoustique Routière et Urbaine, Laboratoire Central des Ponts et Chaussées, 44341 Bouguenais, France
michel.berengier@lcpc.fr

L. Bloom
School of Engineering and Mathematics, Edith Cowan University, Perth, Western Australia
l.bloom@ecu.edu.au

G. Blöschl
Institute for Hydraulic and Water Resources Engineering, Vienna University of Technology, Karlsplatz 13/222, A-1040 Vienna, Austria
bloeschl@hydro.tuwien.ac.at

P. Bogaert
Department of Environmental Sciences and Land Use Planning, Université catholique de Louvain, Belgium
bogaert@enge.ucl.ac.be

A. Boucher
Dpt. of Geological and Environmental Sciences, Stanford University, CA, USA
aboucher@stanford.edu

B. Briand
Laboratory of Radioecological Studies for Marine and Terrestrial Ecosystems, Institute for Radioprotection and Nuclear Safety (IRSN), DEI/SESURE/LERCM, Cadarache, Bld 153, 13105 St-Paul-lez-Durance cedex, France

G. Buttafuoco
CNR – Institute for Agricultural and Forest Systems in the Mediterranean, Via Cavour, 87030 Rende (CS), Italy
buttafuoco@ieif.cs.cnr.it
Contributors

H. Caetano
CMRP – Centre for Modelling Petroleum Reservoirs, Mining and Georesources Department, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisboa, Portugal pcl42082@popsrv.ist.utl.pt

T.J. Calado
Centro Geofísico da Universidade de Lisboa, 1749-016 Lisboa, Portugal mtcalado@fc.ul.pt

N. Caputi
Marine and Fisheries Research Laboratories, Department of Fisheries, North Beach, WA 6020, Australia ncaputi@fish.wa.gov.au

G. Cardenas
Institut National de l’Environnement Industriel et des Risques, Parc Technologique ALATA, BP 2, 60550 Verneuil-en-Halatte, France giovanni.cardenas@ineris.fr

F. Carrat
INSERM UMR-S 707, Paris, France carrat@u707.jussieu.fr

J. Carvalho
Environmental Group of the Centre for Modelling Petroleum Reservoirs – CMRP, Instituto Superior Técnico, Lisboa, Portugal jcarvalho@ist.utl.pt

A. Castrignanò
CRA – Agronomic Research Institute, Bari, Italy annamaria.castrignano@enteca.it

J.-P. Chilès
Centre de Géosciences, École des Mines de Paris, 77305 Fontainebleau, France jean-paul.chiles@ensmp.fr

M.F.E.S. Coelho
Instituto de Meteorologia, Lisboa, Portugal Fatima.Coeelho@meteo.pt

D. Cornford
Neural Computing Research Group, Aston University, Birmingham B4 7ET UK d.cornford@aston.ac.uk

J.A. Corte-Real
Centro de Geofísica de Évora, Universidade de Évora, Portugal jmcr@uevora.pt

A.C.M. Costa
ISEGI, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal ccosta@isegi.unl.pt

C.C. DaCamara
Centro Geofísico da Universidade de Lisboa, 1749-016 Lisboa, Portugal

A. Dassargues
Applied Geology and Mineralogy, Department of Geology-Geography, Katholieke Universiteit Leuven, Belgium; Hydrogeology and Environmental Geology, University of Liège, Belgium alain.dassargues@ulg.ac.be

G.A. Degan
Dipartimento di Ingegneria Meccanica e Industriale; Università degli studi Roma Tre, Rome, Italy g.alfarodegan@uniroma3.it
C. Giasi  
Department of Environmental and Civil Engineering, Polytechnic of Bari, Italy  
c.giasi@poliba.it

J.J. Gómez Hernández  
Polytechnic University of Valencia, Department of Hydraulic Engineering and Environment  
Camino de Vera s/n., 46022 Valencia, Spain  
jaime@dihma.upv.es

E. Gommeren  
Openbare Afvalstoffenmaatschappij voor het Vlaamse Gewest (OVAM), Stationstraat 110, 2800 Mechelen, Belgium  
egommere@ovam.be

P. Goovaerts  
BioMedware, Inc. 516 North State Street, Ann Arbor, MI 48104 – 1236, USA  
goovaerts@biomedware.com

C. Gouveia  
CGUL, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Portugal; Escola Superior de Tecnologia, Instituto Politécnico de Setúbal, Portugal  
c.gouveia@est.ips.pt

A. Guadagnini  
Dipartimento di Ingegneria Idraulica, Ambientale, Infrastrutture Viarie, Rilevamento (DIIAR), Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy  
alberto.guadagnini@polimi.it

A. Horta  
Environmental Group of the Centre for Modelling Petroleum Reservoirs – CMRP, Instituto Superior Técnico, Lisboa, Portugal  
ahorta@ist.utl.pt

D.T. Hristopulos  
Department of Mineral Resources Engineering, Technical University of Crete, Chania 73100, Greece  
dionisi@mred.tuc.gr

M. Huysmans  
Applied Geology and Mineralogy, Department of Geology-Geography, Katholieke Universiteit Leuven, Belgium  
marijke.huysmans@geo.kuleuven.be

J.R. Ilarri  
Universidad Politécnica de Valencia, Instituto de Ingeniería del Agua y Medio Ambiente, Camino de Vera s/n., 46022 Valencia, Spain  
jrodrigo@upv.es

C. Innocenti  
Central Institute for Marine Research, Via di Casalotti, 300, 00166 Roma, Italy.  
c.innocenti@icram.org

Y.O. Isselmou  
France Télécom R & D, RESA/FACE, Issy-les-Moulineaux, France; Ecole des Mines, CG – Geostatistics, Fontainebleau, France

C. Jégat  
INSERM UMR-S 707, Paris, France; Ecole des Mines de Paris – Geostatistics group, Fontainebleau, France  
cyrille.jegat@mines-paris.org
J.A. Jiménez
IFAPA, Centro “Las Torres-Tomejil”,
Junta de Andalucía. Ctra.
Sevilla-Cazalla km 12.2, 41200 Alcalá del Río (Seville),
Spain

F. Junker
Département Analyse Mécanique et Acoustique, Electricité de France,
France
fabrice.junker@edf.fr

M. Kangas
Marine and Fisheries Research Laboratories, Department of Fisheries,
Northbeach,
Western Australia 6020
mkangas@fish.wa.gov.au

C. Kaiser
Institute of Geography, University of Lausanne, CH-1015 Switzerland

M. Kanevski
Institute of Geomatics and Analysis of Risk, University of Lausanne, CH-1015 Switzerland

C. Lajaunie
Ecole des Mines de Paris – Geostatistics group, Fontainbleau,
France
christian.lajaunie@ensmp.fr

D. Lippiello
Dipartimento di Ingegneria Meccanica e Industriale; Università degli studi Roma Tre, Rome
Italy
d.lippielo@uniroma3.it

S. Mackie
Xstrata, Brisbane, Qld, Australia

F. Marconi
Central Institute for Marine Research, via di Casalotti, 300 – 00166 Roma, Italy

D. Marcotte
École Polytechnique de Montréal,
Département des génies civil, géologique et des mines,
C.P. 6079, Succ.
Centre-ville, Montréal, Qc, Canada,
H3C 3A7
denis.marcotte@polymtl.ca

G. Martínez
IFAPA, Centro “Las Torres-Tomejil”,
Junta de Andalucía, Ctra.
Sevilla-Cazalla km 12.2, 41200 Alcalá del Río (Seville),
Spain

H. Mata-Lima
CMRP – Centre for Modelling Petroleum Reservoirs, Mining and Georesources Department, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisboa, Portugal
Departamento de Matemática e Engenharias, Universidade da Madeira, 9000-390 Funchal, Portugal
hlima@uma.pt

M. Van Meirvenne
Dept. Soil Management and Soil Care, Ghent University, Coupure 653,
9000 Gent, Belgium
marc.vanmeirvenne@ugent.be

T. Meklit
Dept. Soil Management and Soil Care, Ghent University, Coupure 653,
9000 Gent, Belgium
meklit.tariku@ugent.be
C. Mercat-Rommens
Laboratory of radioecological studies
for marine and terrestrial ecosystems,
Institute for Radioprotection and
Nuclear Safety (IRSN),
DEI/SESURE/LERCM, Cadarache,
Bld 153, 13105 St-Paul-lez-Durance
cedex, France
catherine.mercat-rommens@irsn.fr

J.-M. Metivier
Laboratory of radioecological studies
for marine and terrestrial ecosystems,
Institute for Radioprotection and
Nuclear Safety (IRSN),
DEI/SESURE/LERCM, Cadarache,
Bld 153, 13105 St-Paul-lez-Durance
cedex, France
jean-michel.metivier@irsn.fr

J.M.M. Avalos
Faculty of Sciences, University of
Coruña, A Zapateira 15071,
A Coruña, Spain
jmirasa@udc.es

U. Mueller
School of Engineering and
Mathematics, Edith Cowan University,
Perth, Western Australia
u.mueller@ecu.edu.au

E. Mumelter
Central Institute for Marine Research,
Via di Casalotti, 300,
00166 Roma, Italy
e.mumelter@icram.org

J.L. Muriel
IFAPA, Centro “Las Torres-Tomejil”,
Junta de Andalucía. Ctra.
Sevilla-Cazalla km 12.2, 41200 Alcalá
del Río (Seville)
Spain
josel.muriel@juntadeandalucia.es

J. Negreiros
Instituto Superior de Línguas e
Administração, Estrada
da Correira 53, 1500-210 Lisbon,
Portugal
c8057@isegi.unl.pt

P. Pasquier
Golder Associates, 9200, L’Acadie
blvd., Montréal, Québec,
Canada, H4N 2T2

D. Patriarche
Gaz de France
361, Avenue du President Wilson,
BP 33, 93211 Saint Denis La Plaine
Cedex, France
delphine.patriarche@gazdefrance.com

A.P. González
Faculty of Sciences, University of
Coruña, La Zapateira 15071,
La Coruña, Spain

E.J. Pebesma
Geosciences Faculty,
Utrecht University, The Netherlands
e.pebesma@geo.uu.nl

E. Perdrix
Ecole des Mines de Douai, France
perdrix@ensm-douai.fr

F. Perea
IFAPA, Centro “Las Torres-Tomejil”,
Junta de Andalucía. Ctra.
Sevilla-Cazalla km 12.2, 41200 Alcalá
del Río (Seville),
Spain

J.M.C. Pereira
Cartography Centre, Tropical Research
Institute, Lisboa, Portugal;
Department of Forestry, Instituto
Superior de Agronomia, Lisboa,
Portugal
jmocpereira@sapo.pt
E. Pili
Commissariat à l’Energie Atomique, Département Analyse Surveillance Environnement, BP 12, 91680 Bruyères-le-Châtel, France eric.pili@cea.fr

M. Pinzari
Dipartimento di Ingegneria Meccanica e Industriale; Università degli studi Roma Tre, Rome Italia pinzari@uniroma3.it

G. Raspa
Dipartimento di Ingegneria Chimica, dei Materiali delle Materie Prime e Metallurgia; La Sapienza Università di Roma, Dept. ICMMPM, via Eudossiana, 18 – 00184 Roma, Italy giuseppe.raspa@uniroma1.it

M. Riva
Dipartimento Ingegneria Idraulica, Ambientale, Infrastrutture Viarie, Rilevamento (DIAR), Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy monica.riva@polimi.it

M. Rivest
École Polytechnique de Montréal, Département des génies civil, géologique et des mines, C.P. 6079, Succ. Centre-ville, Montréal, Qc, Canada, H3C 3A7 martine.rivest@polymtl.ca

A. Russo
CMRP, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal arusso@ist.utl.pt

A. Salmeri
Central Institute for Marine Research, Via di Casalotti, 300, 00166 Roma, Italy a.salmeri@icram.org

X. Sanchez-Vila
Department of Geotechnical Engineering and Geosciences, Technical University of Catalonia, Gran Capità S/N, 08034 Barcelona, Spain xavier.sanchez-vila@upc.edu

E. Sevens
Openbare Afvalstoffenmaatschappij voor het Vlaamse Gewest (OVAM), Stationstraat 110, 2800 Mechelen, Belgium erwin.sevens@ovam.be

J.O. Skøien
Department of Physical Geography, Utrecht University, P.O. box 80115, 3508 TC Utrecht, The Netherlands jskoen@geo.uu.nl

A. Soares
Environmental Group of the Centre for Modelling Petroleum Reservoirs – CMRP, Instituto Superior Técnico, Lisbon, Portugal asoares@ist.utl.pt

W. Tabbara
Supélec, DRE-LSS, Gif-sur-Yvette, France walid.tabbara@lss.supelec.fr

F. Tack
Dept. Soil Management and Soil Care, Ghent University, Coupure 653, 9000 Gent, Belgium filip.tack@ugent.be
D.M. Tartakovsky
Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, CA 92093, USA; Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
dmt@ucsd.edu

J.-F. Thovert
Laboratoire de Combustion et Détonique, SP2MI, BP 179, 86960 Futuroscope Cedex, France
thovert@lcd.ensma.fr

R.M. Trigo
CGUL, Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Portugal; Departamento de Engenharias, Universidade Lusófona, Lisboa, Portugal
rmtrigo@fc.ul.pt

D. Tuia
Institute of Geomatics and Analysis of Risk, University of Lausanne, CH-1015 Switzerland
devis.tuia@unil.ch

K. Vanderlinden
IFAPA, Centro “Las Torres-Tomejil”, Junta de Andalucía. Ctra. Sevilla-Cazalla km 12.2, 41200 Alcalá del Río (Seville), Spain
karl.vanderlinden.ext@juntadeandalucia.es

H.S. Vargas
CMRP – Centre for Modelling Petroleum Reservoirs, Mining and Georesources Department, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001, Lisboa, Portugal
hugo.vargas@total.com

S. Verstraete
Dept. Soil Management and Soil Care, Ghent University, Coupure 653, 9000 Gent, Belgium
samuel.verstraete@ugent.be

H. Wackernagel
Centre de Géosciences, École des Mines de Paris, 77305 Fontainebleau, France
Ecole des Mines de Paris – Geostatistics group, Fontainebleau, France
hans.wackernagel@ensmp.fr

C. Wagner-Riddle
Dept. of Land Resource Science, University of Guelph, Guelph, Ontario, Canada
cwagnerr@uoguelph.ca

J. Wiart
France Télécom R & D, RESA/FACE, Issy-les-Moulineaux, France
joe.wiart@francetelecom.fr

B.E. Wohlberg
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
brendt@lanl.gov
Part I
Environment and Health
Geostatistical Analysis of Health Data: 
State-of-the-Art and Perspectives

P. Goovaerts

Abstract The analysis of health data and putative covariates, such as environmental, socio-economic, behavioral or demographic factors, is a promising application for geostatistics. It presents, however, several methodological challenges that arise from the fact that data are typically aggregated over irregular spatial supports and consist of a numerator and a denominator (i.e. population size). This paper presents an overview of recent developments in the field of health geostatistics, with an emphasis on three main steps in the analysis of aggregated health data: estimation of the underlying disease risk, detection of areas with significantly higher risk, and analysis of relationships with putative risk factors. The analysis is illustrated using age-adjusted cervix cancer mortality rates recorded over the 1970–1994 period for 118 counties of four states in the Western USA. Poisson kriging allows the filtering of noisy mortality rates computed from small population sizes, enhancing the correlation with two putative explanatory variables: percentage of habitants living below the federally defined poverty line, and percentage of Hispanic females. Area-to-point kriging formulation creates continuous maps of mortality risk, reducing the visual bias associated with the interpretation of choropleth maps. Stochastic simulation is used to generate realizations of cancer mortality maps, which allows one to quantify numerically how the uncertainty about the spatial distribution of health outcomes translates into uncertainty about the location of clusters of high values or the correlation with covariates. Last, geographically-weighted regression highlights the non-stationarity in the explanatory power of covariates: the higher mortality values along the coast are better explained by the two covariates than the lower risk recorded in Utah.

1 Introduction

Since its early development for the assessment of mineral deposits, geostatistics has been used in a growing number of disciplines dealing with the analysis of data
distributed in space and/or time. One field that has received little attention in the geostatistical literature is medical geography or spatial epidemiology, which is concerned with the study of spatial patterns of disease incidence and mortality and the identification of potential “causes” of disease, such as environmental exposure or socio-demographic factors (Waller and Gotway 2004). This lack of attention contrasts with the increasing need for methods to analyze health data following the emergence of new infectious diseases (e.g. West Nile Virus, bird flu), the higher occurrence of cancer mortality associated with longer life expectancy, and the burden of a widely polluted environment on human health.

Individual humans represent the basic unit of spatial analysis in health research. However, because of the need to protect patient privacy publicly available data are often aggregated to a sufficient extent to prevent the disclosure or reconstruction of patient identity. The information available for human health studies thus takes the form of disease rates, e.g. number of deceased or infected patients per 100,000 habitants, aggregated within areas that can span a wide range of scales, such as census units, counties or states. Associations can then be investigated between these areal data and environmental, socio-economic, behavioral or demographic covariates. Figure 1 shows an example of datasets that could support a study of the impact of demographic and socio-economic factors on cervix cancer mortality. The top map shows the spatial distribution of age-adjusted mortality rates recorded over the 1970-1994 period for 118 counties of four states in the Western USA. The corresponding population at risk is displayed in the middle map, either aggregated within counties or assigned to 25 km$^2$ cells. The bottom maps show two putative explanatory variables: percentage of habitants living below the federally defined poverty line, and percentage of Hispanic females. Indeed, Hispanic women tend to have elevated risk of cervix cancer, while poverty reduces access to health care and to early detection through the Pap smear test in particular (Friedell et al. 1992). These socio-demographic data are available at the census block level and were assigned to the nodes of a 5 km spacing grid for the purpose of this study (same resolution as the population map).

A visual inspection of the cancer mortality map conveys the impression that rates are much higher in the centre of the study area (Nye and Lincoln Counties), as well as in one Northern California county. This result must however be interpreted with caution since the population is not uniformly distributed across the study area and rates computed from sparsely populated counties tend to be less reliable, an effect known as “small number problem” and illustrated by the top scattergram in Fig. 1. The use of administrative units to report the results (i.e. counties in this case) can also bias the interpretation: had the two counties with high rates been much smaller in size, these high values likely would have been perceived as less problematic. Last, the mismatch of spatial supports for cancer rates and explanatory variables prevents their direct use in the correlation analysis.

Unlike datasets typically analyzed by geostatisticians, the attributes of interest are here measured exhaustively. Ordinary kriging, the backbone of any geostatistical analysis, thus seems of little use. Yet, I see at least three main applications of geostatistics for the analysis of such aggregated data:
1. Filtering of the noise caused by the small number problem using a variant of kriging with non-systematic measurement errors.

2. Modeling of the uncertainty attached to the map of filtered rates using stochastic simulation, and propagation of this uncertainty through subsequent analysis, such as the detection of aggregate of counties (clusters) with significantly higher or lower rates than neighboring counties.

3. Disaggregation of county-level data to map cancer mortality at a resolution compatible with the measurement support of explanatory variables.

Goovaerts (2005a, 2006a,b) introduced a geostatistical approach to address all three issues and compared its performances to empirical and Bayesian methods which have been traditionally used in health science. The filtering method is based on Poisson kriging and semivariogram estimators developed by Monestiez et al. (2006) for mapping the relative abundance of species in the presence of spatially heterogeneous observation efforts and sparse animal sightings. Poisson kriging was combined with p-field simulation to generate multiple realizations of the spatial distribution of cancer mortality risk. A limitation of all these studies is the assumption that the size and shape of geographical units, as well as the distribution of the population within those units, are uniform, which is clearly inappropriate in the example of Fig. 1.

The last issue of change of support was addressed recently in the geostatistical literature (Gotway and Young 2002, 2005; Kyriakidis 2004). In its general form kriging can accommodate different spatial supports for the data and the prediction, while ensuring the coherence of the predictions so that disaggregated estimates of count data are non-negative and their sum is equal to the original aggregated count. The coherence property needs however to be tailored to the current situation where aggregated rate data have various degree of reliability depending on the size of the population at risk (Goovaerts, 2006b).

This paper discusses how geostatistics can benefit three main steps of the analysis of aggregated health data: estimation of the underlying disease risk, detection of areas with significantly higher risk, and analysis of relationships with putative risk factors. An innovative procedure is proposed for the deconvolution of the semivariogram of aggregated rates and the disaggregation of these rates, accounting for heterogeneous population densities and the shape and size of administrative units. The different concepts are illustrated using the cervix cancer data of Fig. 1.

2 Estimating Mortality Risk from Observed Rates

For a given number $N$ of entities $v_\alpha$ (e.g. counties), denote the observed mortality rates as $z(v_\alpha) = d(v_\alpha)/n(v_\alpha)$, where $d(v_\alpha)$ is the number of recorded mortality cases and $n(v_\alpha)$ is the size of the population at risk. Let us assume for now that all entities $v_\alpha$ have similar shapes and sizes, with a uniform population density. These entities can thus be referenced geographically by their centroids with the vector of spatial
Fig. 1 Geographical distribution of cervix cancer mortality rates recorded for white females over the period 1970–1994, and the corresponding population at risk (aggregated within counties or assigned to 25 km$^2$ cells). Scatterplot illustrates the larger variance of rates computed from sparsely populated counties. Bottom maps show two putative risk factors: percentage of habitants living below the federally defined poverty line, and percentage of Hispanic females.

coordinate’s $\mathbf{u}_\alpha = (x_\alpha, y_\alpha)$. The disease count $d(\mathbf{u}_\alpha)$ is interpreted as a realization of a random variable $D(\mathbf{u}_\alpha)$ that follows a Poisson distribution with one parameter (expected number of counts) that is the product of the population size $n(\mathbf{u}_\alpha)$ by the local risk $R(\mathbf{u}_\alpha)$, see Goovaerts (2005a) for more details.
In Poisson kriging (PK), the risk over a given entity \( v_\alpha \) is estimated as a linear combination of the kernel rate \( z(u_i) \) and the rates observed in \((K-1)\) neighboring entities:

\[
\hat{r}_{PK}(u_\alpha) = \sum_{i=1}^{K} \lambda_i(u_\alpha)z(u_i)
\]

where \( \lambda_i(u_\alpha) \) is the weight assigned to the rate \( z(u_i) \) when estimating the risk at \( u_\alpha \). The \( K \) weights are the solution of the following system of linear equations:

\[
\sum_{j=1}^{K} \lambda_j(u_\alpha) \left[ C_R(u_i - u_j) + \delta_{ij} \frac{m^*}{n(u_i)} \right] + \mu(u_\alpha) = C_R(u_i - u_\alpha) \quad i = 1, \ldots, K
\]

where \( \delta_{ij}=1 \) if \( u_i=u_j \) and 0 otherwise, and \( m^* \) is the population-weighted mean of the \( N \) rates. The addition of an “error variance” term, \( m^*/n(u_i) \), for a zero distance accounts for variability arising from population size, leading to smaller weights for less reliable data (i.e. measured over smaller populations). The prediction variance associated with the estimate (1) is computed using the traditional formula for the ordinary kriging variance:

\[
\sigma^2_{PK}(u_\alpha) = C_R(0) - \sum_{i=1}^{K} \lambda_i(u_\alpha)C_R(u_i - u_\alpha) - \mu(u_\alpha)
\]

The computation of kriging weights and kriging variance (Equations (2) and (3)) requires knowledge of the covariance of the unknown risk, \( C_R(h) \), or equivalently its semivariogram \( \gamma_R(h)=C_R(0)-C_R(h) \). Following Monestiez et al. (2006) the semivariogram of the risk is estimated as:

\[
\hat{\gamma}_R(h) = \frac{1}{2} \sum_{a=1}^{N(h)} \left\{ \frac{n(u_\alpha)n(u_\alpha+h)}{n(u_\alpha)+n(u_\alpha+h)} \left[ z(u_\alpha) - z(u_\alpha+h) \right]^2 - m^* \right\}
\]

where the different pairs \([z(u_\alpha) - z(u_\alpha+h)]\) are weighted by the corresponding population sizes to homogenize their variance.

### 2.1 Area-to-Area (ATA) Poisson Kriging

In the situation where the geographical entities have very different shapes and sizes, areal data can not be simply collapsed into their respective polygon centroids. Following the terminology in Kyriakidis (2004), ATA kriging refers to the case where
both the prediction and measurement supports are blocks (or areas) instead of points. The PK estimate (1) for the areal risk value \( r(v_\alpha) \) thus becomes:

\[
\hat{r}_{PK}(v_\alpha) = \sum_{i=1}^{K} \lambda_i (v_\alpha) z(v_i) \tag{5}
\]

The Poisson kriging system (2) is now written as:

\[
\sum_{j=1}^{K} \lambda_j (v_\alpha) \left[ \bar{C}_R(v_i, v_j) + \delta_{ij} \frac{m*}{n(v_i)} \right] + \mu(v_\alpha) = \bar{C}_R(v_i, v_\alpha) \quad i = 1, \ldots, K
\]

\[
\sum_{j=1}^{K} \lambda_j (v_\alpha) = 1. \tag{6}
\]

The main change is that point-to-point covariance terms \( C_R(u_i - u_j) \) are replaced by area-to-area covariances \( \bar{C}_R(v_i, v_j) = \text{Cov}(Z(v_i), Z(v_j)) \). Like in the traditional block kriging, those covariances are approximated by the average of the point support covariance \( C(h) \) computed between any two locations discretizing the areas \( v_i \) and \( v_j \):

\[
\bar{C}_R(v_i, v_j) = \frac{1}{P_i \sum_{s=1}^{P_i}} P_j \sum_{s'=1}^{P_j} w_{ss'} C(u_s, u_{s'}) \tag{7}
\]

where \( P_i \) and \( P_j \) are the number of points used to discretize the two areas \( v_i \) and \( v_j \), respectively. For the example of Fig. 1 a grid with a spacing of 5 km was overlaid over the study area, yielding a total of 11 to 2,082 discretizing points per county depending on its area. The high-resolution population map in Fig. 1 clearly shows the heterogeneous distribution of population within counties. To account for spatially varying population density in the computation of the area-to-area covariance, the weights \( w_{ss'} \) were identified to the product of population sizes within the 25 km\(^2\) cells centred on the discretizing point \( u_s \) and \( u_{s'} \):

\[
w_{ss'} = n(u_s) \times n(u_{s'}) \text{ with } \sum_{s=1}^{P_i} n(u_s) = n(v_i) \text{ and } \sum_{s'=1}^{P_j} n(u_{s'}) = n(v_j) \tag{8}
\]

The kriging variance for the areal estimator is computed as:

\[
\sigma^2_{PK}(v_\alpha) = \bar{C}_R(v_\alpha, v_\alpha) - \sum_{i=1}^{K} \lambda_i (v_\alpha) \bar{C}_R(v_i, v_\alpha) - \mu(v_\alpha) \tag{9}
\]
where $\bar{C}_R(v_\alpha, u_\alpha)$ is the within-area covariance that depends on the form of the geographical entity $v_\alpha$ and decreases as its area increases. Thus, ignoring the size of the prediction support in the computation of the kriging variance (3) can lead to a systematic overestimation of the prediction variance of large blocks.

2.2 Area-to-Point (ATP) Poisson Kriging

A major limitation of choropleth maps is the common biased visual perception that larger rural and sparsely populated areas are of greater importance. A solution is to create continuous maps of mortality risk, which amounts to perform a disaggregation or area-to-point interpolation. At each discretizing point $u_s$ within an entity $v_\alpha$, the risk $r(u_s)$ can be estimated as the following linear combination of areal data:

$$\hat{r}_{PK}(u_s) = \sum_{i=1}^{K} \lambda_i(u_s) z(v_i)$$  \hspace{1cm} (10)

The Poisson kriging system is similar to system (6), except for the right-hand-side term where the area-to-area covariances $\bar{C}_R(v_i, v_\alpha)$ is replaced by the area-to-point covariance $\bar{C}_R(v_i, u_s)$. The latter is approximated by a procedure similar to the one described in equation (7). A critical property of the ATP kriging estimator is its coherence, that is the aggregation of the $P_\alpha$ point risk estimates within any given entity $v_\alpha$ yields the areal risk estimate $\hat{r}_{PK}(v_\alpha)$:

$$\hat{r}_{PK}(v_\alpha) = \frac{1}{n(v_\alpha)} \sum_{s=1}^{P_\alpha} n(u_s) \hat{r}_{PK}(u_s)$$  \hspace{1cm} (11)

Condition (11) differs from the constraint commonly found in the geostatistical literature (Kyriakidis, 2004) in that: 1) the observation $z(v_\alpha)$ is uncertain, hence it is the reproduction of the PK risk estimate $\hat{r}_{PK}(v_\alpha)$ that is imposed, and 2) the incorporation of the population density in the computation of the areal covariance implies that it is the population-weighted average of the point risk estimates, not their arithmetical average, that satisfies the coherence condition. The constraint (11) is satisfied if the same $K$ areal data are used for the estimation of the $P_\alpha$ point risk estimates. Indeed, in this case the population-weighted average of the right-hand-side covariance terms of the $K$ ATP kriging systems is equal to the right-hand-side covariance of the single ATA kriging system:

$$\frac{1}{n(v_\alpha)} \sum_{s=1}^{P_\alpha} n(u_s) C_R(v_i, u_s) = \frac{1}{n(v_i)} \sum_{s=1}^{P_i} n(u_s) \left[ \frac{1}{n(v)} \sum_{s'=1}^{P} n(u_s') C(u_s', u_s) \right] \hspace{1cm} (12)$$

$$= C_R(v_i, v_\alpha),$$
per relations (7) and (8). Therefore, the following relationship exists between the two sets of ATA and ATP kriging weights:

$$\lambda_i(v_\alpha) = \frac{1}{n(v_\alpha)} \sum_{s=1}^{P_v} n(u_s) \lambda_i(u_s) \quad i = 1, \ldots, K$$

(13)

which ensures the coherence of the estimation.

### 2.3 Deconvolution of the Semivariogram of the Risk

Both ATA and ATP kriging require knowledge of the point support covariance of the risk $C(h)$, or equivalently the semivariogram $\gamma(h)$. This function cannot be estimated directly from the observed rates, since only aggregated data are available. Derivation of a point support semivariogram from the experimental semivariogram of areal data is called “deconvolution”, an operation that is frequent in mining and has been the topic of much research (Journel and Huijbregts, 1978). However, in typical mining applications all blocks (areas) have the same size and shape, which makes the deconvolution reasonably straightforward. Goovaerts (2008) proposed an iterative approach to conduct the deconvolution in presence of irregular geographical units. This innovative algorithm starts with the derivation of an initial deconvoluted model $\gamma^{(0)}(h)$; for example the model $\gamma_R(h)$ fitted to the areal data. This initial model is then regularized using the following expression:

$$\gamma_{\text{regul}}(h) = \bar{\gamma}^{(0)}(v, v_h) - \bar{\gamma}_h^{(0)}(v, v)$$

(14)

where $\bar{\gamma}^{(0)}(v, v_h)$ is the area-to-area semivariogram value for any two counties separated by a distance $h$. It is approximated by the population-weighted average (7), using $\gamma^{(0)}(h)$ instead of $C(h)$. The second term, $\bar{\gamma}_h^{(0)}(v, v)$, is the within-area semivariogram value. Unlike the expression commonly found in the literature, this term varies as a function of the separation distance since smaller areas tend to be paired at shorter distances. To account for heterogeneous population density, the distance between any two counties is estimated as a population-weighted average of distances between locations discretizing the pair of counties:

$$\text{Dist}(v_i, v_j) = \frac{1}{\sum_{s=1}^{P_i} \sum_{s'=1}^{P_j} n(u_s)n(u_{s'}) ||u_s - u_{s'}||}$$

(15)

Note that the block-to-block distances (15) are numerically very close to the Euclidean distances computed between population-weighted centroids (Goovaerts, 2006b). The theoretically regularized model, $\gamma_{\text{regul}}(h)$, is compared to the model fitted to experimental values, $\gamma_R(h)$, and the relative difference between the two curves,
denoted $D$, is used as optimization criterion. A new candidate point-support semi-
variogram $\gamma^{(1)}(h)$ is derived by rescaling of the initial point-support model $\gamma^{(0)}(h)$, and then regularized according to expression (14). Model $\gamma^{(1)}(h)$ becomes the new optimum if the theoretically regularized semivariogram model $\gamma^{(1)}_{\text{regul}}(h)$ gets closer to the model fitted to areal data, that is if $D^{(1)} < D^{(0)}$. Rescaling coefficients are then updated to account for the difference between $\gamma^{(1)}_{\text{regul}}(h)$ and $\gamma_R(h)$, leading to a new candidate model $\gamma^{(2)}(h)$ for the next iteration. The procedure stops when the maximum number of allowed iterations has been tried (e.g. 35 in this paper) or the decrease in the $D$ statistic becomes negligible from one iteration to the next. The use of lag-specific rescaling coefficients provides enough flexibility to modify the initial shape of the point-support semivariogram and makes the deconvolution insensitive to the initial solution adopted. More details and simulation studies are available in Goovaerts (2006b, 2008).

### 2.4 Application to the Cervix Cancer Mortality Data

Figure 2 (top graph, dark gray curve) shows the experimental and model semi-
variograms of cervix cancer mortality risk computed from aggregated data using
estimator (4) and the distance measure (15). This model is then deconvoluted and, as expected, the resulting model (light gray curve) has a higher sill since the punctual process has a larger variance than its aggregated form. Its regularization using expression (14) yields a semivariogram model that is close to the one fitted to experimental values, which validates the consistency of the deconvolution.

The deconvoluted model was used to estimate aggregated risk values at the county level (ATA kriging) and to map the spatial distribution of risk values within counties (ATP kriging). Both maps are much smoother than the map of raw rates since the noise due to small population sizes is filtered. In particular, the high risk area formed by two central counties in Fig. 1 disappeared, which illustrates how hazardous the interpretation of the map of observed rates can be. The highest risk (4.081 deaths/100,000 habitants) is predicted for Kern County, just west of Santa Barbara County. ATP kriging map shows that the high risk is not confined to this sole county but spreads over four counties, which is important information for designing prevention strategies. By construction, aggregating the ATP kriging estimates within each county using the population density map of Fig. 1 (right medium graph) yields the ATA kriging map.

The map of ATA kriging variance essentially reflects the higher confidence in the mortality risk estimated for counties with large populations. The distribution of population can however be highly heterogeneous in large counties with contrasted urban and rural areas. This information is incorporated in the ATP kriging variance map that shows clearly the location of urban centers, such as Los Angeles, San Francisco, Salt Lake City, Las Vegas or Tucson. The variance of point risk estimates is much larger than the county-level estimates, as expected.
Fig. 2 Experimental semivariogram of the risk estimated from county-level rate data, and the results of its deconvolution (top curve). The regularization of the point support model yields a curve (black dashed line) that is very close to the experimental one. The model is then used to estimate the cervix cancer mortality risk (deaths/100,000 habitants) and associated prediction variance at the county level (ATA kriging) or at the nodes of a 5 km spacing grid (ATP kriging).

3 Detection of Spatial Clusters and Outliers

Mapping cancer risk is a preliminary step towards further analysis that might highlight areas where causative exposures change through geographic space, the