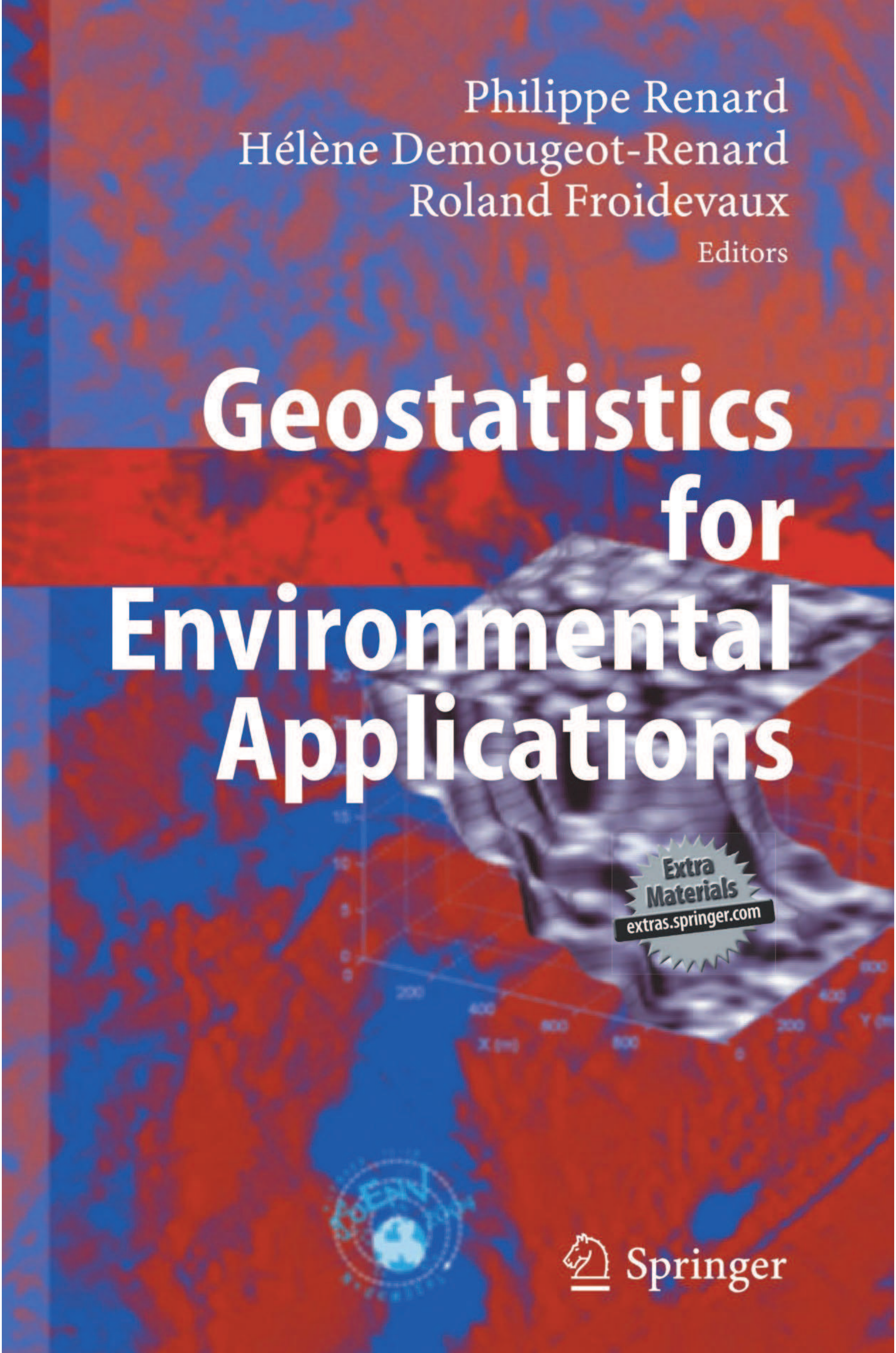
Philippe Renard
Hélène Demougeot-Renard
Roland Froidevaux
Editors

Geostatistics for Environmental Applications



Extra Materials
extras.springer.com

Springer

Philippe Renard
Hélène Demougeot-Renard
Roland Froidevaux

Geostatistics for Environmental Applications

Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications

Philippe Renard
Hélène Demougeot-Renard
Roland Froidevaux
(Editors)

Geostatistics for Environmental Applications

Proceedings of the Fifth European Conference on Geostatistics for Environmental Applications

With 218 Figures and a CD-ROM

Springer

Dr. Philippe Renard
University of Neuchâtel, Centre for Hydrogeology
11 Rue Emile Argand, 2007 Neuchâtel, Switzerland
E-mail: *philippe.renard@unine.ch*

Dr. Hélène Demougeot-Renard
FSS International
7 Chemin de Mont-Riant, 2000 Neuchâtel, Switzerland
E-mail: *demougeot.renard@fssintl.com*

Dr. Roland Froidevaux
FSS International
9 Rue Boissonnas, 1227 Geneva, Switzerland
E-mail: *roland.froidevaux@fssintl.com*

Library of Congress Control Number: 2005927607

ISBN-10 3-540-26533-3 **Springer Berlin Heidelberg New York**
ISBN-13 978-3-540-26533-7 **Springer Berlin Heidelberg New York**

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Cover design: Erich Kirchner
Production: Luisa Tonarelli
Typesetting: Camera-ready by the editors

Printed on acid-free paper 30/2132/LT – 5 4 3 2 1 0

Foreword

While the society becomes everyday more aware of environmental problems, the experts have to deal with a major issue: uncertainty due to incomplete data sets and spatio-temporal variability. Among the techniques used to quantify this uncertainty and to improve environmental management, geostatistics is becoming a recognized standard, applied in fields as different as hydrogeology, remote sensing, ecology or soil contamination. In recent years, the domain of application of these techniques has regularly grown together with the wide availability of Geographical Information Systems and geostatistical software packages.

This book is the outcome of the fifth edition of the European Conference on Geostatistics for Environmental Applications (geoENV V) held in Neuchâtel (Switzerland) from October 13th to October 15th, 2004. The conference attracted more than 100 participants, mostly from Europe, but also from North America, South America, North Africa, Russia and Australia. Among the 140 abstracts submitted to the conference, the organization committee selected 53 for oral presentation and 40 for poster presentation.

The book reflects the current status of the research in the field of geostatistics for environmental applications. It opens with one keynote paper by Carol Gotway-Crawford, senior researcher at the Center for Disease Control and Prevention, Atlanta, which emphasizes the problem of the size of the data support when making spatial statistics. It is then followed by 38 research papers, classified according to their main topics, that describe recent methodological advances and applications. All these papers have been presented orally during the conference and accepted by the reviewers. The final version of the papers were again checked by the editors. Also included in the book is a CDROM containing the original posters and the companion poster papers. This material has not been reviewed.

Finally, the editors wish to thank all the reviewers and the authors without whom this book could not exist, as well as the sponsors of the conference: the Swiss Federal Statistical Office (http://www.bfs.admin.ch), the Swiss Federal Office for Water and Geology (http://www.bwg.admin.ch/e/), the Swiss National Science Foundation (http://www.snf.ch/), the University of Neuchâtel (http://www.unine.ch), the Centre of Hydrogeology of the University of Neuchâtel (http://www.unine.ch/chyn), the Banque Cantonale Neuchâteloise (http://www.bcn.ch), and the NCCR Plant Survival (http://www.unine.ch/nccr/).

Neuchâtel, February 2005

The editors.

Philippe Renard
Hélène Demougeot-Renard
Roland Froidevaux

Organizing and scientific committees

Philippe Renard, *University of Neuchâtel, Switzerland (Chairman)*
Hélène Demougeot-Renard, *FSS International, Neuchâtel, Switzerland*
Roland Froidevaux, *FSS International, Geneva, Switzerland*
Denis Allard, *INRA, Avignon, France*
Jaime Gómez-Hernández, *UPV, Valencia, Spain*
Pascal Monestiez, *INRA, Avignon, France*
Xavier Sánchez-Vila, *UPC, Barcelona, Spain*
Amilcar Soares, *IST, Lisbon, Portugal*

The editors are grateful to the following persons for their work as referees:

Rachid Ababou, *IMFT, Toulouse, France*
Denis Allard, *INRA, Avignon, France*
Peter M. Atkinson, *University of Southampton, United Kingdom*
Ana Bio, *Instituto Superior Tecnico, Portugal*
Patrick Bogaert, *Université Catholique de Louvain, Belgium*
Josè Capilla, *Universidad Politécnica de Valencia, Spain*
Jesus Carrera, *Universidad Politécnica de Catalunya, Spain*
Eduardo Cassiraga, *Universidad Politécnica de Valencia, Spain*
Mario Chica-Olmo, *University of Granada, Spain*
Jean-Paul Chilès, *Ecole des Mines de Paris, France*
George Christakos, *University of North Carolina, USA*
Olaf Cirpka, *University of Stuttgart, Germany*
Noel Cressie, *Ohio State University, USA*
Hélène Demougeot-Renard, *FSS International, Neuchâtel, Switzerland*
Dimitri d'Or, *Université Catholique de Louvain, Belgique*
Souheil Ezzedine, *Lawrence Livermore National Laboratory, USA*
Chantal de Fouquet, *Ecole des Mines de Paris, France*
Luc Feyen, *Katholieke Universiteit Leuven, Belgium*
Aldo Fiori, *University of Roma Tre, Italy*
Roland Froidevaux, *FSS International, Geneva, Switzerland*
Michel Garcia, *FSS International, Paris, France*
Marc Genton , *North Carolina State University, USA*
Tilmand Gneiting, *University of Washington, USA*
Jaime Gómez-Hernández, *Universidad Politécnica de Valencia, Spain*
Pierre Goovaerts, *Biomedware, Ann Arbor, USA*
Carol Gotway-Crawford, *NCEH, Atlanta, USA*
Alberto Guadagnini, *Politecnico de Milano, Italy*
Laura Guadagnini, *Politecnico de Milano, Italy*
Gilles Guillot, *INRA, France*
Harri-Jan Hendricks-Franssen, *ETH Zürich, Switzerland*
Olivier Jaquet, *Colenco Power Consulting AG, Switzerland*
André Journel, *Stanford University, USA*

Phaedon Kiriakidis, *University of California, USA*
Denis Marcotte, *Ecole Polytechnique Montréal, Canada*
Peter Meier, *NOK Baden, Switzerland*
Pascal Monestiez, *INRA, France*
Carla Nunes, *Instituto Superior Tecnico, Portugal*
Maria João Pereira, *Instituto Superior Tecnico, Portugal*
Antonio Pulido-Bosch, *University of Almeria, Spain*
Monica Riva, *Politecnico de Milano, Italy*
Jacques Rivoirard, *Ecole des Mines de Paris, France*
Klaus-Jürgen Röhlig, *GRS, Germany*
Christian Roth, *Airparif, France*
Xavier Sánchez-Vila, *Universidad Politécnica de Catalunya, Spain*
Marc Serre, *University of North Carolina at Chapel Hill, USA*
Amilcar Soares, *Instituto Superior Tecnico, Portugal*
Mohan Srivastava, *FSS Canada, Canada*
Fritz Stauffer, *ETH Zürich, Switzerland*
Alfred Stein, *University of Wageningen, The Netherland*
Hans Wackernagel, *Ecole des Mines de Paris, France*
Richard Webster, *Rothamsted Experimental Station, United Kingdom*

Table of Contents

Keynote paper

Methods

Ecology, air and health

Groundwater

Soil contamination

Remote sensing

Environment

Contents of the CD-ROM

Methods

GEOSSAV: a simulation tool for subsurface applications (article, poster) *C. Regli, P. Huggenberger and L. Rosenthaler*

Testing independence for spatial processes through spectral analysis (article) *P. Juan, E. Porcu and J. Mateu*

Ecology, air and health

Mapping annual nitrogene dioxide concentrations above Mulhouse urban area (article) *C. de Fouquet, D. Gallois, L. Malherbe, G. Cardenas and G. Perron*

Using systematic diffusive sampling campaigns and geostatistics to map air pollution in Portugal (article, poster) *F. Ferreira, S. Mesquita, P. Torres and H. Tente*

Mapping air quality using a geostatistical approach: application to a regional ozone measurement campaign in the North of France (article, poster) *G. Cardenas and L. Malherbe*

Habitat suitability index models for the sympatric soles Solea solea and Solea senegalensis using GIS procedures (poster) *C. Vinagre, V. Fonseca, H.N. Cabral and M.J. Costa*

Modelling and monitoring epidemics by means of spatio-temporal lattices (poster) *E. Järpe*

Agrogeomatic techniques application for a more precise management of corn (Zea mays L.) (article, poster) *S. Bocchi, A. Castrignanò and L.S. Viganò*

Environmental radioactive pollution: biogeochemical assessment of 90Sr in trees using geostatistical methods. Theory, methodology and a case study (article, poster) *C. Hervada-Sala, E. Jarauta, Y.G. Tyutyunnik, S.M. Bednaiya and N.D. Kuchma*

Groundwater

Cokriging of the phreatic level using a digital elevation model (poster) *E. Mendoza, G. Herrera and M. Díaz*

Stochastic inverse modeling of groundwater flow and environmental tracer transport: Baltenswil case study (Switzerland) (poster) *G.A. Onnis, H.J. Hendricks Franssen, F. Stauffer and W. Kinzelbach*

How many transmissivity data are required to safely delineate a protection zone braided alluvial aquifer when using kriging and upscaling? (poster) *J. Kerrou, P. Renard, I. Lunati, S. Madier and H.J. Hendricks-Franssen*

Application of geostatistics to soil and groundwater samples contaminated by petroleum products – A case study (article) *J.M. Carvalho and A. Fiúza*

Temporal and spatial variations of geo-environmental parameters in soil, rock and groundwater samples of the northern calcareous alps in austria (poster) *S. Pfleiderer, H. Reitner, H. Pirkl, P. Klein and M. Heinrich*

Soil

Survey and evaluation of land use and land cover in switzerland (poster) *F. Weibel*

Estimating the frequency of polluted soils in allotment gardens and mapping their spatial distribution using proxy data (article) *C. Hofer, A. Papritz and A. Borer*

Mapping and simulations of geneva soils, using geostatistics and ANN (article) *M. Maignan, M. Kanevski, F. Celardin and A. Besson*

Use of kriging to assess the ground contamination (article) *F. Largueche*

Geostatistical model for total petroleum hydrocarbons (tph) in Santa Alejandrina porous media, Veracruz (México) (poster) *J.H. Flores Ruiz, A. Mejía Ramírez, J. Fucugauchi Urrutia, E. Hernández-Quintero and G. Domínguez Zacarías*

Cokriging field measured soil hydraulic conductivity and texture in a brazilian semi-arid watershed (article, poster) *A.A. Montenegro, W. Lundgren, S.M.G.L Montenegro*

A study of the spatial variability of soil water retention by mixed effects linear models with a spatial continuous autoregressive correlation structure (article, poster) *B. Cafarelli, A. Castrignanò and N. Romano*

Exploring the multivariate spatial structure of soil acidity data (article, poster) *R. Lilla Manzione, G. Camara, A. M. Vieira Monteiro, C.R. Lopes Zimback and S. Druck Fucks*

Sampling plan criteria for the bottom mud characterisation of a drainage channel (article, poster) *S. Sgallari, R. Bruno and C. Zampighi*

Environment

Spatio-temporal geostatistical analyses of runoff and precipitation (article, poster) *J.O. Skøien, G. Blöschl*

A geostatistical approach for rainfall patterns, using proximity indices (poster) *A. Gutiérrez and M. Preciado*

Delineation of estuarine management units: Evaluation of an automatic procedure

F. Bação[1], S. Caeiro[2], M. Painho[1], P. Goovaerts[3] and M. H. Costa[4]

[1]ISEGI/CEGI, Institute for Statistics and Information Management of the New University of Lisbon, Portugal, e-mail: bacao@isegi.unl.pt, painho@isegi.unl.pt
[2]IMAR/Department of Exact and Technological Sciences of the Portuguese Distance Learning University, Lisbon, Portugal, e-mail: scaeiro@univ-ab.pt
[3]BioMedware, Inc. 516 North State Street, Ann Arbor, USA,
e-mail: goovaerts@biomedware.com
[4]IMAR, Faculty of Science and Technology of the New University of Lisbon, Caparica, Portugal, e-mail: mhcosta@fct.unl.pt

1 Introduction

A costal zone management (CZM) program should have well defined zones that can be managed based on their specific characteristics and needs. Once established, these management units constitute the backbone of the whole management strategy. The objective of a zoning plan is to delineate smaller areas that can be managed in a more flexible way (Cicin-Sain and Knecht 1998). Over the last few decades, there has been a move towards developing ways to identifying these units (McGlashan and Duck 2002).

The definition of the transition zone between the ocean and terrestrial environment, ocean and coastal zones, and zones (or units) within the coastal areas is sometimes not an easy task. Physical criteria, political boundaries, administrative boundaries, arbitrary distances or selected environmental units can and are often used (Clark 1996).

Most CZM projects use administrative boundaries instead of adopting an ecosystem approach looking at impacts coming from outside the area considered (Belfiore 2000). Coastal management units are evolving by becoming more inclusive, relying more on processes than on administrative boundaries and by incorporating a wider range of expertise in defining relevant areas (McGlashan and Duck 2002). The correct way to delineate estuarine management units is based on the development of robust and ecological representative processes. Eventually, these processes can be implemented using automatic procedures capable of providing promptly answers to complex problems. The automatic procedure proposed here draws inspiration on previous work developed in geographic zone design. In fact, the task of developing estuary management units can be viewed as a special case of the more general problem of geographical zone design (Martin 2000). Zone design is a long-standing geographical problem that is present in a number of geographical tasks; the best known example probably is electoral districting (Williams 1995, Macmillan and Pierce 1994). Zone design algorithms have also been used in a variety of tasks; such as school districting (Ferland and Guénette 1990), the design of zones with appropriate characteristics for posterior socio-economic and epidemiological analysis (Haining *et al.* 1994, Openshaw and Rao 1995, Openshaw and Alvanides 1999), the design of sales territories (Fleischmann and Paraschis 1988) and census output geography (Martin 1997, Martin 1998).

The aim of this paper is to develop an automatic optimization procedure, based on genetic algorithms, to delineate sediment estuarine management units, and to compare these results with a well-established multivariate geostatistical method. Both approaches will be illustrated using the Sado Estuary. The resulting management units will represent the support infrastructure of an environmental data management framework to monitor this ecosystem.

1.1 Zone design using multivariate geostatistical tools

Geostatistical techniques like kriging allow estimation of attribute values at unsampled locations taking into account the spatial continuity of the data (Soares 2000). Since kriging is preceded by an analysis of the spatial structure of the data, the average spatial variability of the data is already integrated into the estimation/interpolation process (Wackernagel 1995).

Multivariate methods like principal component analysis, cluster analysis and discriminant analysis can be coupled with the different types of kriging (Oliver and Webster 1989, Reed *et al.* 2001, Goovaerts 2002) allowing one to group sampling sites that both have similar properties and are geographically close. With these multivariate geostatistical techniques interpolation is improved, small occurrences of one kind of land within others of fairly similar kind are disregarded and undesirable fragmentation avoided (Goovaerts 1997, Reed *et al.* 2001).

For example (MacDonald *et al.* 2000) developed an ecosystem-based framework for assessing and managing sediment quality conditions in Tampa Bay previously defined management areas. Those areas were delineated using interpolated contour lines based on sediment chemistry data and guidelines of potential adverse effects. Picollo *et al.* (2003) used homogenous units for the coastal zone management of the Ligurian region. These subdivisions of the coast corresponded to physiographic units (topographic elements).

1.2 Zone design using genetic algorithms

The constraints of the zone design problem are similar to the ones that characterize the clustering problem. Let the set of initial areal units be $X = \{x_1, x_2 ..., x_n\}$, where x_i is the ith areal unit. The number of zones is denoted k, and Z_i is the set of all the areal units that belong to zone Z_i. Then:

$$\mathbf{Z}_i \neq \varnothing, \text{ for } i = 1, \ldots, k, \tag{1}$$

$$\mathbf{Z}_i \cap \mathbf{Z}_j = \varnothing \text{, for } i \neq j, \tag{2}$$

$$\bigcup_{i=1}^{k} \mathbf{Z}_i = \mathbf{X} \tag{3}$$

These constitute the set of constraints that can be applied equally in clustering and in zone design. Nevertheless, the typical zone design task usually presents an additional constraint, which accounts for contiguity and creates a more complex problem (Macmillan and Pierce 1994). We will not address the issue of contiguity here, as it is irrelevant in the context of this research. However the automatic procedure described below guaranties that the zones created will be contiguous.

To deal with the zone design problem a number of different algorithms have been proposed (for a thorough review in the context of electoral districting, see Williams 1995). Nevertheless, there are two major problems in applying the existing automatic zone design algorithms to the development of estuary management units. First, most of the algorithms described in the literature are not available as they result from research efforts and most of them were never implemented as software packages. Second, most algorithms are based on an areal perspective of the zone design problem, i.e. they use areas as the basic units for zone design, which conflicts with the point supports used in the context of this research.

The zone design algorithms proposed in the literature use different optimization strategies ranging from hill climbing procedures (Horn 1995) to simulated annealing (Macmillan and Pierce 1994, Openshaw and Rao 1995), tabu search (Openshaw and Rao 1995) and linear programming associated with branch-and-bound (Mehrotra *et al.* 1998). Genetic Algorithms (GA) remain largely unexplored in this field. To our knowledge the only reference is Altman (1998) and no details are provided on how GA's were applied to this particular problem. However, GA's have been used extensively as search procedures in related fields such as the P-Median Problem (Correa *et al.* 2001) and Cluster Analysis (Murthy and Chowdhury 1996). Other fields facing complex optimization problems, such as Pattern Recognition, Image Processing and Machine Learning (Ankenbrandt *et al.* 1990, Belew and Booker 1991, Back *et al.*, 1997) have also benefited from the use of GA's.

One of the reasons why the zone design problem is especially difficult is the size of the solution space. The dimension of a "usual" real world problem makes unfeasible any attempt to enumerate all the possible solutions explicitly (Cliff and Hagget 1970). The total number of possible solutions, *S*, for a zone design problem, is similar to the clustering problem and is calculated as the Stirling number of the second kind (Anderberg 1973, Keane 1975):

$$S(n,k) = \frac{1}{k!}\sum_{i=1}^{k}(-1)^i\left(\frac{k!}{(k-i)!\,i!}\right)(k-i)^n \tag{4}$$

This means that for a medium size problem like the one addressed in the results section of this paper, S (153, 19) yields $3.65*10^{178}$ possible solutions. Additionally, in terms of computational complexity the zone design problem has been shown to be NP-Complete (Altman 1997). Thus, heuristic techniques seem to be the best way available to produce solutions in reasonable computational time.

2 Methods

2.1 Sampling design

Sediment samples were collected at 153 sites using a Global Positioning System, according to a systematic unaligned sampling design (500 × 750 m) to provide a uniform coverage of the area as well as pairs of close observations required for modeling the short-scale variability. In each site 3 sediment characterization attributes were determined: Fine Fraction content (FF), Total Organic Matter (OM) and Redox Potential (Eh) (Caeiro *et al.* 2003a). The central problem consisted in building a small number of regions (areas) based on the original 153 sample points. This will enable the periodic monitoring and adequate management of the estuary. Boundaries of spatially contiguous and homogeneous regions of sediment structure were derived through two alternative approaches described below.

2.2 Multivariate Geostatistical approach

This method starts with a principal component analysis (PCA) of original data (FF, TOM and Eh), followed by the computation and fitting of a spherical model to the experimental semivariogram of principal components (PC) scores. Following Oliver and Webster (1989), the dissimilarity between any two sampling sites i and j is then computed as:

$$d_{ij}^{*} = d_{ij} \times \frac{c}{c_0 + c} \times \left[1.5 \times \frac{|u_{ij}|}{a} - 0.5 \times \left(\frac{|u_{ij}|}{a} \right)^{3} \right] + d_{ij} \times \frac{c_0}{c_0 + c} \quad \text{for } 0 < u_{ij} \leq a \tag{5}$$

$$d_{ij}^{*} = d_{ij} \quad \text{for } u_{ij} > a$$

Where:

d_{ij} - Distance in the attribute space between *i* and *j*

c - Sill of the spherical semivariogram model

c_0 - Nugget variance

a - Range of the spherical semivariogram model

u_{ij} - Euclidean geographic distance between *i* and *j*.

These values are assembled into a dissimilarity matrix that undergoes hierarchical clustering using the complete linkage rule (Everitt and Dunn 2001). The spatial continuity of each cluster is characterized using semivariograms computed on indicators of occurrence of these clusters. Indicator kriging is then used to interpolate the probability of occurrence of the clusters at unsampled locations. Finally, each grid node is assigned to the cluster with the highest probability of occurrence (maximum likelihood classification).

This method generates relatively smooth maps showing locally dominant classes, uncluttered by outliers. This procedure fulfills the purpose of computing contiguous sediment regions for management and monitoring purposes. A detailed description of this method is available in Caeiro *et al.* (2003a). The area corresponding to the sampling points was further clipped with the study area boundary including the coast line (Caeiro *et al.* 2003b) using ArcGIS 8.0 software.

2.3 Genetic algorithms approach

Genetic algorithms (GA) are a subset of a broader and rapidly expanding area known as Evolutionary Computing (Fogel 2000). As the name indicates, these algorithms drew inspiration on Darwin's theory of evolution, and have been used to solve hard optimization and machine learning problems (Goldberg 1989). The basic idea is that each solution to the problem is coded as a bit string, a chromosome, possibly with a number of sub-strings that act as genes. At any given point in time (or generation), a number of such chromosomes are kept, each representing a different "individual" or solution to the problem. Natural selection is simulated by evaluating the fitness (or goodness) of each solution, measured by how well it solves the problem at hand, and giving the best individuals a higher probability of being chosen to breed (crossover) and thus passing their characteristics into the next generation. To obtain new solutions, two operators are used: crossover, and mutation. Crossover is implemented by combining bits of two different chromosomes (possibly divided along genes) to form a new solution, while mutation amounts at randomly changing some bits or chromosomes. The details of how this basic idea is implemented may vary considerably.

Given enough time, a conveniently coded GA will always find an optimal solution. However, to obtain reasonable solutions within reasonable time, care must be taken in the encoding of the problem into chromosomes, and in the choice of the fitness function that will be optimized.

The application of genetic algorithms to the development of estuary management units first requires a strategy for encoding a solution to the problem. In other words, a specific partition of the sampling points into a smaller set of management units needs to be encoded in such a way that genetic operators may be used. This could be performed using a number of different ways (Bação *et al.* 2004). Bearing in mind that in the specific case of estuary management units compactness is not a relevant constraint, the encoding used here enables the seed of each management unit to be placed anywhere within a rectangle comprising the study region.

Each string represents a possible plan configuration, and the fitness of each specific configuration is evaluated using the following expression:

$$min \sum_{i=1}^{k} \sum_{z=1}^{n_z} \sum_{v=1}^{m} \left| x_{izv} - \overline{x_{zv}} \right| \tag{6}$$

where x_{izv} represents the value of sampling point i from management unit z for variable v, and $\overline{x_{zv}}$ represents the mean value of variable v in management unit z. In order to assess the fitness of each solution we have to calculate the sum of the distances between each sampling point and the mean of the management unit to which it belong along all variables. Finally, the quality of the solution is assessed by the sum of the distances within all the management units.

The genetic algorithm is initialized with a random population of size *p*. For the encoding described above, *p* strings of length 2**k* are initialized. This is done by forcing all elements of the strings to be located within a rectangle comprising the study region. GA literature does not provide guidelines for choosing the size of the initial population. In this study 10 parallel populations were used, with different string numbers. Migration of strings between populations can occur with a probability of 0.001. Identical strings are not allowed, so there are no twins in the population.

The type of selection used is tournament (Goldberg 1989). Sensitivity analysis led to the choice of a uniform crossover, with a probability ranging from 0.95 to 1. Mutation rate was 0.001 and 0.002, and an elitist strategy was adopted, assuring that the best individual of the population would always be carried to the next generation. Different stopping criteria were also used.

Thus, the algorithm proceeds as follows:

1. Generate p sets of k points, according to the selected encoding.
2. For each of the *n* sampling points find the closest seed, and assign the sampling point to the seed.
3. Evaluate the fitness of each string, based on equation 6.
4. Apply selection, crossover and mutation operators, creating a new population;
5. Return to step 2 until the stopping criterion is met.

The final result is a classification of the sampling points into *n* zones. To generate areas (management units) an allocation function was computed using the Spatial Analyst extension of ArcGIS 8.0 software.

3 Results

In this section we compare the results provided by the two approaches. The major goal is to identify the differences rather than trying to prove the superiority of one technique over the other. The elusiveness of a strict objective criterion to the delineation of estuary management units forces a somewhat subjective analysis of the results. The use of expert knowledge was one of the ways of circumvent this difficulty.

3.1 Multivariate Geostatistical approach

The hierarchical classification yielded four clusters that are reasonably distinct, with a decline in organic load from Cluster 1 to 4: HO – High Organic; MHO – Medium High Organic; MO – Medium Organic and LO – Low Organic loads (Fig. 1). For each cluster, the indicator semivariogram was computed along four directions and a geometric anisotropy model was fitted visually. All semivariograms display longer ranges in the direction of azimuth 120º, which corresponds to the water flow and is in agreement with other studies (Martins *et al.*

2001). Maximum likelihood classification performed on estimated probabilities generated 70 areas after clipping with the estuary costal line. The areas smaller than the sampling grid size were assigned to the neighboring area with the longest common border, resulting in a final set of 19 management units. A detailed description of the methodology is available in Caeiro *et al.* (2003a). Results of cluster classification of the 153 sampling points and the corresponding 19 management units are displayed in Fig. 1.

The results of this method are generally in agreement with earlier work performed in the estuary (Rodrigues and Quintino 1993). Low organic load sediments correspond essentially to the estuarine entrance and tend to extend to the inside of the estuary, basically through the southern channel (see at the estuary entrance, a large homogenous area of low organic load, LO1 - Fig. 1). In the middle of the estuary bay the gradient splits into two major components, one directed towards the North Channel and the other towards the South Channel in accordance with the water flow. Since high organic load areas are associated with low hydrodynamics and rich organic discharges, they are more common in the North Channel near industrialized zones and the city of Setubal. They are also distributed in small homogenous areas (Fig 1). These results were compared to the classification provided by two other multivariate approaches using map similarity measurements (Caeiro *et al.* 2003a, Caeiro *et al.* 2004). These previous studies demonstrated the robustness of this multivariate approach, indicating that the different methods yield similar results and thus are of equal value to delineate management units in the estuary.

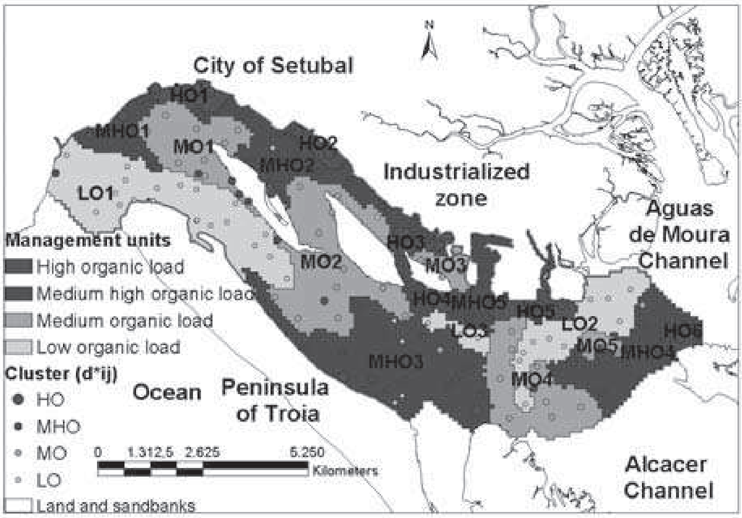


Fig. 1. Classification of the 153 sampling point produced using hierarchical classification (four clusters), and the final set of 19 management units obtained with the multivariate geostatistical approach.

3.2 Genetic algorithms approach

Genetic algorithm was run using five different sets of parameters (see Table 1) and under the constraint of creating 19 areas. Solution S1 achieves the worst result in terms of within cluster variability, nevertheless it is important to note that only 50000 individuals (10 populations of 10 individuals over 500 generations) were evaluated, which constitutes a very small fraction of the solution space ($3.65*10^{178}$). This result indicates that genetic algorithms are quite robust, as they need to search only a small amount of this space in order to find good solutions. In the case of S5 the algorithm stops after 1000 generations without improvements in the objective function, yielding a total of 2907 generations. Only solution S1 did not lower the within cluster variability produced by the geostatistical method (134.17).

Table 1. - Results and parameter specification of the 5 runs performed in the genetic algorithm.

	Sum of Within management unit variability	Population	Mutation	Stopping Criterion	Crossover probability
S1	143.92	10*10	0.001	500 generations	1
S2	123.30	10*15	0.001	1000 generations	1
S3	123.05	10*25	0.001	800 generations	1
S4	126.16	10*25	0.002	800 generations	0.95
S5	117.19	10*25	0.001	2907 generations	0.95

Because of the page limitations, only the best (S5) and worst results (S1) are displayed (see Fig. 2 and 3). The analysis of the maps shows that S5 is able to isolate 6 sampling points into 4 areas, which creates smaller areas while improving the objective function. The two solutions are, however, characterized by similar spatial pattern of the management units.

The genetic algorithm yields 19 management units without clustering them in 4 groups, which makes the comparison with the geostatistical method more difficult. The final result is the classification of each sampling point into one of the 19 management units, unlike the geostatistical approach which produces surface areas. Despite these differences the analysis of the resulting maps shows that similar patterns of small areas are found in the North Channel. In fact, the general results are quite similar in terms of the “macro” trends. The genetic algorithm identifies small management areas (like area nº 19 in Fig. 2 or nº 8 in Fig. 3), which were also detected by the multivariate geostatistical method (see Fig. 1 cluster results), but as the interpolation procedure only yielded an area smaller than the sampling grid size, that was discarded. The split gradients in the North and South channels displayed by the hierarchical classification are less noticeable in the two solutions of the genetic algorithm. This can be explained by the fact that the genetic algorithm does not take into account the anisotropic pattern of variability of the attributes, which can be modeled and incorporated using the geostatistical approach. The estuarine sedimentary environment is strongly controlled by the water flow, which supports the existence of the long areas identified by the geostatistical method, particularly in the south channel where the hydrodynamics is higher, representing the natural physical behavior. Due to distance restrictions imposed by the encoding the genetic algorithm is unable to create elongated areas (see areas nº 7 and 11 in Fig. 2 or nº 7 and 18 in Fig. 3, that correspond to one single area – LO1 in Fig. 1).

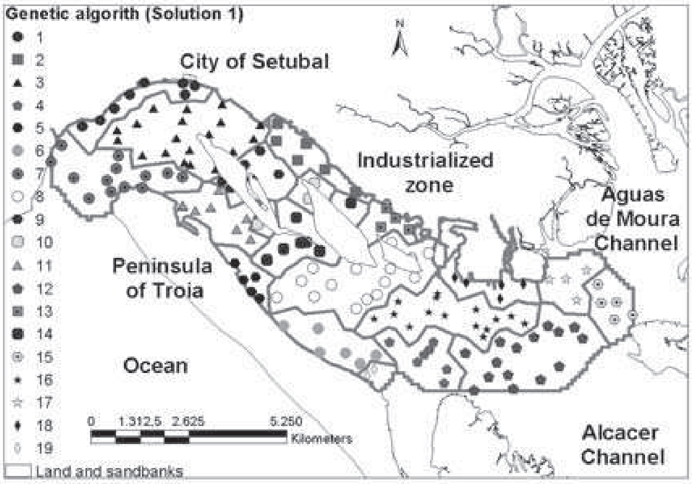


Fig. 2. Classification of the 153 points produced using the genetic algorithm approach – S1.

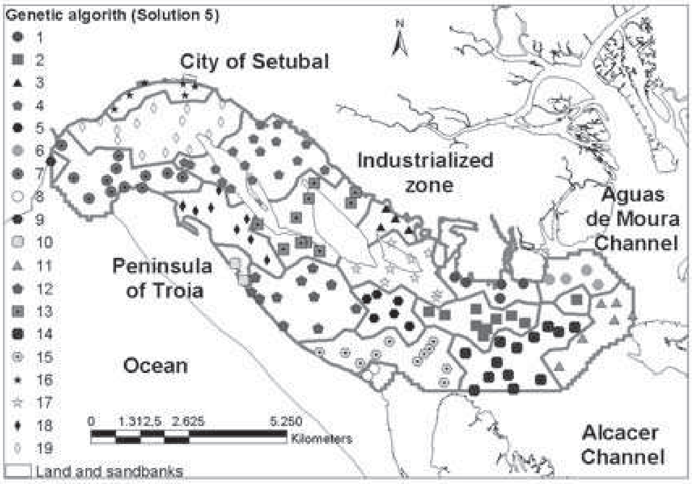


Fig. 3. Classification of the 153 points produced using genetic algorithm approach – S5.

About 60 % of the total number of the GA's areas of both solutions 1 and 5 are mainly formed by one single cluster (i.e. proportion is higher then 80 %). However, in solution 5, only 31.6 % of the areas are 100 percent classified as belonging to one single type of cluster, compared to 15.8 % for Solution 1. In spite of this S1 seems to represent better the estuary hydrodynamic behavior, and is more in accordance with the geostatistical approach. In addition, 47.4 % of the 19 centroids of Solution 1 are associated with one single centroid of the management units of geostatistical approach compared to only 36.8 % in solution 5.

In order to organize the major findings of this study, the quality of each solution was assessed using four point expert evaluation criteria, see Table 2. As expected, the quality of the ecological representation is superior in the geostatistical method, as it accommodates specific needs related to underlying phenomena, such as the hydrodynamic behavior. Nevertheless, it is noteworthy that the genetic algorithm is not inadequate, since for instance the expert assessment is quite encouraging (does not need to have an in-depth understanding of the methods).

Table 2. Assessment scores of the results of the multivariate geostatistical approach (MG) and solutions 1 and 5 provided by GA. Best score is 3.

Method	Sum of within management unit variability	Ecological interpretation (N-S differences)	Ecological interpretation (Hydrodynamic behavior)	Required knowledge about the method from the user*	Expert Assessment	Total
MG	134.17	3	3	1	3	10
S1	143.92	2	2	3	2	9
S5	117.19	2	1	3	2	8

* 1 means that the user needs to have an in-depth understanding of the methods.

4 Conclusions

In this paper, we implemented a genetic algorithm approach for the delineation of management units in estuary and the results were compared to a multivariate geostatistical approach based on a hierarchical clustering of a spatial dissimilarity matrix followed by indicator kriging. Although the task of comparing such different approaches is difficult, due to different types of outputs and the absence of a "reference truth", this study indicates that the results of both methods are relatively similar. The automatic procedure presented here has the potential to become a valuable option in the delineation of estuary management units. The current genetic algorithm can provide a benchmark for other approaches; enabling the possibility of critical assessment of theoretical based approaches.

The first solution S1, although with higher within area variability, seems to represent better the ecological behavior of the estuary (see Table 2). This fact may indicate that using the within management unit variability as the optimization criteria might be misleading, or at least it can be improved. Finally, future work should investigate alternate encoding options which would enable the algorithm to produce more elongated, and still contiguous, areas. This improvement should lead to a better representation of the hydrodynamic effects, which play a relevant role in the definition of the management units.

In this study the issue of defining the number of regions was not addressed. Clearly, in the future, the genetic algorithm should incorporate new strategies for the automatic computation of the number of regions to be built. This should capitalize on the work developed within cluster analysis, for instances the application of a pseudo *F* statistic developed by Calinski and Harabasz (1974) or the cubic clustering criterion proposed by Sarle (1983).

References

Anderberg MR (1973) Cluster Analysis for Applications. Academic Press

Ankenbrandt CA, Buckles BP, Petry FE (1990) Scene recognition using genetic algorithms with semantic nets. Pattern Recognition Letters 11(4):285-293

Altman M (1997) The Computational Complexity of Automated Redistricting: Is Automation the Answer? Rutgers Computer and Technology Law Journal 23 (1):81-142

Altman M (1998) Modeling the Effect of Mandatory District Compactness on Partisan Gerrymanders. Political Geography 17 (8): 989-1012

Bação F, Lobo V, Painho M (2004) Applying Genetic Algorithms to Zone Design. Accepted for publication in Soft Computing: A Fusion of Foundations, Methodologies and Applications

Back T, Fogel DB, Michalewicz Z (eds.) (1997) Handbook of Evolutionary Computation. Oxford University Press NY

Belew RK, Booker LB (eds.) (1991) Proceedings of the Fourth International Conference on Genetic Algorithms. La Jolla, CA: Morgan Kaufmann

Belfiore S (2000) Recent developments in coastal management in the European Union. Ocean & Coastal Management 43:123-135

Caeiro S, Goovaerts P, Painho M, Costa MH (2003a) Delineation of Estuarine management areas using multivariate geostatistics: the case of Sado estuary. Environmental Science and Technology 37:4052-4059

Caeiro S, Goovaerts P, Painho M, Costa MH, Sousa S (2003b) Spatial sampling design for sediment quality assessment in estuaries. Environmental Modeling and Software 18(10): 853 – 859

Caeiro S, Sousa S, Painho H (2004) Map Similarity Measurements and its application to the Sado Estuary. Accepted for publication in Finisterra

Calinski T, Harabasz J (1974) A Dendrite Method for Cluster Analysis, Communications in Statistics 3:1 -27

Cicin-Sain B, Knech RW (1998) Integrated Coastal and Ocean Management. Concepts and Practices. Washington, D. C, Covelo, California, Island Press

Clark JR (1996) Coastal Zone Management Handbook. Boca Raton, Florida, Lewis Publishers

Cliff AD, Hagget P (1970) On the efficiency of alternative aggregations in region-building problems. Environment and Planning 2: 285-294

Correa ES, Steiner MTA, Freitas AA, Carnieri C, (2001) A genetic algorithm for the P-median problem. In: Proc. 2001 Genetic and Evolutionary Computation Conf. (GECCO-2001),Morgan Kaufmann, 1268-1275

Everitt B, Dunn G (2001) Apllied Multivariate Data Analysis. New York, Arnold

Ferland JA, Guénette G (1990) Decision support system for a school districting problem. Operations Research. 38:15-21

Fleischmann BJ, Paraschis N (1988) Solving a large scale districting problem: a case report. Computers and Operations Research. 15:521-533

Fogel DB (2000) Evolutionary Computation: Toward a New Philosophy of Machine Intelligence. 2nd edition, IEEE Press, Piscataway, NJ

Goldberg DE (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, Mass

Goovaerts P (1997) Geostatistics for Natural Resources Evaluation. New Work, Oxford University Press

Goovaerts P (2002) Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. Journal of Geographical Systems 4: 99-111.

Haining R, Wise S, Blake M (1994) Constructing regions for small-area analysis – material deprivation and colorectal cancer. Journal of Public Health Medicine, 16: 457-469.

Horn M (1995) Solution techniques for large regional partitioning problems. Geographical Analysis 27: 230-248.

Keane M (1975) The size of the region-building problem. Environment and Planning A 7: 575-577.

MacDonald DD, Lindskoog RA, Smorong DE, Greening H, Pribble R, Janicki T, Janicki S. Grabe S, Sloane G, Ingersoll CG, Eckenrod S and Long ER (2000) Development of an Ecosystem-Based Framework for Assessing and Managing Sediment Quality Conditions in Tampa Bay, Florida. Florida, Tampa Bay Estuary Program.

Macmillan B, Pierce T, (1994) Optimization Modelling in a GIS Framework: The Problem of Political Redistricting. In: Fotheringham S, Rogerson P (eds.) Spatial analysis and GIS. Taylor & Francis Inc, Bristol, 221-246.

Martin D (1997) From enumeration districts to output areas: experiments in the automated creation of a census output geography. Population Trends, 88:36-42

Martin D (1998) 2001 Census output areas: from concept to prototype. Population Trends, 94:19-24

Martin D (2000) Automated zone design in GIS. In: Atkinson, P. and Martin, D. (eds.) GIS and Geocomputation, Innovations in GIS 7: Taylor and Francis, London, pp. 103-113

Martins F, Leitão P, Silva A and Neves R (2001). 3D modelling in the Sado estuary using a new generic vertical discretization approach. Oceanologica Acta 24:S51 - S62.

McGlashan D. and Duck RW (2002) The Evolution of Coastal Management Units: Towards the PDMU. In: F. Veloso-Gomes, F. Taveira-Pinto, and L. Neves (eds.). Littoral 2002, The Changing Coast. Porto, European Coastal Zone Association for Science and Technology (EUROCOAST)- Portugal Association /EUCC - The Coastal Union, 29 – 33.

Mehrotra A, Johnson EL, Nemhauser GL (1998) An optimization based heuristic for political districting. Management Science, 44:1100-1114.

Murthy CA, Chowdhury N (1996) In search of optimal clusters using genetic algorithms. Pattern Recognition Letters, 17:825-832

Oliver MA and R. Webster, R (1989) A Geostatistical Basis for Spatial Weighting in Multivariate Classification. Mathematical Geology 21:15-35.

Openshaw S, Rao L, (1995) Algorithms for reengineering 1991 Census Geography. Environment and Planning A 27:425-446

Openshaw S, Alvanides S (1999) Applying geocomputation to the analysis of spatial distributions, In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW, (eds.) Geographical Information Systems Principles, Techniques, Management and Applications, Wiley, Chichester, 267-282.

Picollo A, Albertelli G, Bava S and Cappo S. (2003) The Role of Geographic Information Systems (GIS) and of DPSIR Model in Ligurian Coastal Zone Management.In: Proceedings of 5th International Symposium on GIS and Computer Cartography for Coastal Zone Management Italy, Geographical Information Systems International Group (GISIG/International Center for Coastal and Ocean Policy Studies(ICOOPS). 1 – 5

Reed J, Chappel, A, French JR and Oliver MA (2001) A geostatistical analysis of PCB-contaminated sediment in a commercial dock, Swansea, UK. P. In: Monestiez, P., Allard, D.,. Froidevaux, R. (eds.) *geoENV III* - Geostatistics for environmental application*s*. Kluwer Academic Publishers, Dordrecht. 487 – 498.

Rodrigues AM and Quintino VMS (1993) Horizontal Biosedimentary Gradients Across the Sado Estuary, W. Portugal. Netherlands Journal of Aquatic Ecology 27:449-464.

Sarle WS (1983) Cubic Clustering Criterion, SAS Technical Report A-108, Cary, NC: SAS Institute Inc.

Soares A (2000) Geostatística para as Ciências da Terra e do Ambiente. IST Press Lisboa.

Williams JC (1995) Political Redistricting - A Review. Papers in Regional Science 74(1): 13-39

Wackernagel H (1995) Multivariate Geostatistics: an Introduction with Applications. Springer-Verlag.