Book of Abstracts of geoENV 2010

8th International Conference
on
Geostatistics for Environmental Applications

Ghent University, Belgium

13 - 15 September 2010

Editors: L. Cockx, M. Van Meirvenne, P. Bogaert, D. D`Or

ISBN: 978-94-9069-539-2

8th International Conference on Geostatistics for Environmental Applications

WELCOME TO GEOENV2010 !

On behalf of the geoENV2010 Organising Committee and the geoENVIA patronising organisation we are pleased to welcome you at the 8th International Conference on Geostatistics for Environmental Applications in Gent, Belgium.

GeoENV conferences have been held biennially at venues across Europe. From the first conference in Lisbon in 1996, the event has been in Valencia (1998), Avignon (2000), Barcelona (2002), Neuchâtel (2004), Rhodes (2006) and Southampton (2008). It has become a leading forum for Scientists across a broad range of disciplines to share their experiences on the application of geostatistics to environmental problems.

This book contains the extended abstracts of the oral and poster contributions to the geoENV2010 conference. They have been selected, but not edited, by the organising committee. So the presented abstracts are included as they have been submitted by the authors. Papers are included in the order in which they appear in the programme.

In the past, each geoENV conference produced a book of full papers, creating a prestigious and authorative series of geostatistical literature. In this conference it was decided to bundle the full papers after review in a special issue of an international journal. The choice fell on the Springer journal *Stochastic Environmental Research and Risk Assessment* (SERRA). Authors who have presented their work at the conference are invited to submit a paper to the geoENV2010 organising committee who will then proceed with the reviewing process and make recommendations to the editor of SERRA. Every participant of geoENV2010 will receive a copy of this special issue.

Other differences with previous geoENV conferences are the omission of parallel sessions which is possible because we reduced the time per contribution. In this way all participants can attend all talks, but talks will have to be more concise. Ample time is foreseen for a poster session with poster authors presenting their work in front of a group of delegates.

Many persons, mainly from the research group Soil Spatial Inventory Techniques of the Department of Soil Management of the Faculty of Bio-Science Engineering, UGent, have contributed to the organisation of this conference. One person in particular, Dr. Liesbet Cockx, is thanked for taking most of the load on her shoulders and for preparing this book of abstracts.

We hope that this issue in the series of geoENV conferences will be equally fruitful and rewarding to its participants and the readers of this book and the special issue of SERRA.

Marc Van Meirvenne, Patrick Bogaert and Dimitri D'Or
Chair of the Organising Committee of GeoENV2010

With financial support from:

Fonds Wetenschappelijk Onderzoek
Research Foundation – Flanders

UGent Doctoral Schools

Faculty
of Bioscience Engineering

8th International Conference on Geostatistics for Environmental Applications

PROGRAMME

Monday 13 September 2010

8h-9h	Registration at the Faculty of Bioscience Engineering, Building E	
9h	Conference opening by geoENVia President Ph. Renard	
	Chair: P. Goovaerts	
9h15	**Keynote I: C. Scheidt & J. Caers**	**Spatial modeling with ensembles in metric space** **p. 1**
	Session 1: Theory I	**p. 4**
10h	G. Mariethoz, P. Renard, J. Straubhaar	The directed sampling method to perform multivariate multiple-point geostatistical simulations
10h15	A. Boucher, G. Phelps, C. Cronkite-Ratcliff	Simulating complex 3D alluvial stratigraphy by combining training-image-based algorithms and vertical variograms
10h30	J. Strauhbaar, Ph. Renard, G. Mariethoz, R. Froidevaux, O. Besson	List-based algorithm for multiple-point statistics simulation
10h45	**Coffee break**	
	Chair: J. Boisvert	
	Session 2: Theory II	**p. 13**
11h05	R. Labourdette, T. Chugunova	Multiplepoint statistics with auxiliary properties applied to sustainable development. Example from urban expansion prediction of Pau agglomeration (France)
11h20	M. Huysmans, A. Dassargues	Direct multiple-point geostatistical simulation of edge properties for modeling thin irregularly-shaped surfaces
11h35	C. Scheidt, J. Caers, Y. Chen, L.J. Durlofsky	Rapid construction of ensembles of high-resolution models constrained to dynamic data
11h50	A. Parra, J.M. Ortiz	Conditional multiple point simulation with texture synthesis
12h05	R. Tolosana-Delgado, J.J. Egozcue, A. Sánchez-Arcilla, J. Gómez	Classifying wave forecasts with model-based geostatistics and the Aitchison distribution
12h20	**Lunch**	

Monday 13 September 2010

	Chair: D. D`Or		
	Session 3: Soil		**p. 28**
13h30	E. Meerschman, M. Van Meirvenne	Application of multiple point geostatistics in soil science: the reconstruction of polygonal networks of ice-wedge pseudomorphs	
13h45	G.M. Siqueira, J.Dafonte Dafonte, J. Paz-Ferreiro	Spatial correlation between weeds and soil apparent electrical conductivity measured using electromagnetic induction	
14h	M. Robinson, J. McKinley, A. Ruffell, M. Young	The improvement of peat depth models for Northern Ireland through the investigation of Tellus data	
14h15	J. Van de Wauw, P.A. Finke	Screening the Flemish (Belgian) soil drainage class map for its currency	
14h30	R. Kerry, B. Rawlkins, P. Goovaerts	Area-to-point kriging of soil texture: an efficient use of legacy soil data from polygon maps	
14h45	**Poster session: Tour of posters, 5 min per poster** ***Chairs*: *Ph. Renard & J. McKinley***		**p. 165**
16h20	**Coffee break**		
	Chair: A. Boucher		
	Session 4: Geology		**p. 47**
16h40	L. Foresti, V. Demyanov, M. Christie, M. Kanevski	Multiple kernel models for reservoir characterization	
16h55	S. Bhowmik, C.A. Mantilla, S. Srinivasan	Tracking CO_2 plume migration during geologic sequestration using a probabilistic history matching approach	
17h05	J. McKinley, A.Ruffell, M.Young	Determining local spatial variability from a regional airborne radiometric survey	
17h20	P. Tr. Simard, E. Gloaguen, D. Marcotte, A. Boucher	Wavelet-based porosity simulation using high resolution analogue model and low-resolution tomographic data	
18h30	**Evening activity**:	Guided tour through the historical centre of Ghent Meeting point: St-Baafsplein at the fountain	

Tuesday 14 September 2010

	Chair: P. Bogaert		
8h45	**Keynote II: G. Christakos**	**Spatiotemporal geostatistics and the environment: crossing Parmenides' gates and Morrison's doors**	**p. 59**
	Session 1: Hydrology I		**p. 62**
9h30	D. Pedretti, M. Barahona-Palomo, D. Bolster, D. Fernàndez-Garcia, X. Sanchez-Vila	Spatial assessment of infiltration capacity of soils for artificial recharge practices using Google Earth images	
9h45	L. Li, H. Zhou, J.J. Gómez-Hernández	Transport upscaling using multi-rate mass transfer in three-dimensional highly heterogeneous porous media	
10h	R.L. Manzione, E. Wendland, M.Knotters	Stochastic simulation of time series models combined with universal kriging: a framework to predict risks of water table depths in time and space	
10h15	H. Zhou , J.J. Gómez-Hernández, L. Li	Performance of modified ensemble Kalman filter in non-Gaussian heterogeneous media	
10h30	E. Savelyeva, S. Utkin	Stochastic modeling for liquid radioactive storage reservoir	
10h45	**Coffee break**		
	Chair: A. Bárdossy		
	Session 2: Hydrology II		**p. 77**
11h05	B. Bourgine, B. Lopez, N. Baran, D. Ratehau	Characterization of nitrate long term trends at regional scale using statistical and geostatistical tools. Example of the Loire_Brittany basin	
11h20	C. Cronkite-Ratcliff, G. Phelps, A. Boucher	Contribution of geological heterogeneity characterization to uncertainty in predicting unsaturated zone flow	
11h35	C. P. Haslauer, P. Guthke, A. Bárdossy, E.A. Sudicky	Effects of non-gaussian spatial dependence of hydraulic conductivity on hydrodynamic macrodispersion	
11h50	L. Li, H. Zhou, J.J. Gómez-Hernández	A comparative study of three-dimensional hydraulic conductivity upscaling at the macrodispersion experiment (made) site, on Colombus Air Force Base in Mississipi	

Tuesday 14 September 2010

	Session 2: Hydrology II (continued)	
12h05	B. Rogiers, D. Mallants, O. Batelaan M. Gedeon, M. Huysmans, A. Dassargues	Geostatistical analysis of primary and secondary data in a sandy aquifer at Mol/Dessel, Belgium
12h20	N. Jeannée, D. Tapsoba	Improved mapping of daily precipitation over Quebec using the moving-geostatistics approach
12h35	**Group photo geoENV10**	
13h	**Lunch**	
	Chair: M. Huysmans	
	Session 3: Pollution I	**p. 98**
14h	H. Kazianka, J.Pilz	Objective bayesian analysis for the correlation parameters in gaussian copula-based spatial models
14h15	D. D'Or, D. Allard, R. Froidevaux	Incorporating non stationarity in categorical variables simulations with MCS
14h30	P. Harris, C. Brunsdon, M. Charlton	Visualisation techniques for moving window kriging
14h45	A. Gribov, K. Krivoruchko	Local polynomials for data detrending and interpolation in the presence of barriers
15h	G. Mariethoz, J. Caers, C. Scheidt	Multi-way sensitivity analysis using clustering techniques
15h15	M. Kanevski, V. Timonin, S. Robert, L. Foresti	Pattern recognition in environmental data using general regression neural networks
15h30	J.B. Boisvert, C.V. Deutch	Modelling locally varying anisotropy of CO_2 emissions
15h45	**Coffee break**	
	Chair: M. Kanevski	
	Session 4: Pollution II	**p. 122**
16h15	K. Henneböhl, E.R. Pebesma, W.G. Müller	Efficient parametric variogram estimation for real-time interpolation of environmental monitoring data
16h30	A. Govaerts, A. Vervoort	Automatic semivariogram modeling by weighted least squares and by cross-validation

Tuesday 14 September 2010

	Session 4: Pollution II (continued)	
16h45	P. Bossew	Probalistic spatial prediction of indoor Rn from soil air concentration
17h	Y. Desnoyers, J.-P. Chilès, D. Dubot, N. Jeannée, J.-M. Idasiak	Geostatistics for radiological evaluation: methodological developments and outcomes
17h15	E. Dons, C. Beckx, Th. Arentze, G. Wets, L. Int Panis	Shop opening hours and population exposure to NO_2 assessed with an activity-based transportation model
17h30	**General Assembly**	
19h15	**Conference Dinner**	Het Pand – Onderbergen - Gent

Wednesday 15 September 2010

	Chair: R. Froidevaux		
9h	**Keynote III: A. Bardossy**	**Geostatistical interpolation using copulas**	**p. 137**
	Session 1: Pollution III		**p. 140**
9h45	L.E. Gerharz, E. Pebesma	Accounting for uncertainties and change of support in spatio-temporal modelling of individual exposure to air pollution	
10h	D.P. Robinson, C.D. Lloyd, J.M. McKinley	Applying an exhaustive secondary dataset for geostatistical analysis of air pollution at regional level	
10h15	A. Kolovos, A. Skupin, W. Pang, G. Christakos, M. Jerrett	Spatiotemporal geostatistics in cross-methodological monitoring of air pollution	
10h30	U. Leopold, L. Drouet, D.S. Zachary	Accounting for spatio-temporal uncertainty associated to emission allocation in an energy-air quality assessment model	
10h45	**Coffee break**		
	Chair: I. Clark		
	Session 2: Health		**p. 151**
11h05	P. Goovaerts	Merging areal and point data in geostatistical interpolation: applications to soil science and medical geography	
11h20	L.B. Nucci,. A.C.C.N. Mafra, C. Stephan, L.T.O. Zangirolani, R. Cordeiro	Distribution of overweight and identification of spatial risk	
11h35	N.A. Samat, D.F. Percy	Relative risk estimation of Dengue disease mapping based on discrete time-space stochastic Sir-Si models for disease transmission with tract-count data	
11h50	C. Stephan, A.C.C.N. Mafra, L.B. Nucci, R. Cordeiro	Spatial risk distribution of work related accidents in southeast Brazilian city under a multinomial approach.	
12h05	Y. Yano, U. Mueller, A. Hinwood	Generalised linear modelling of the asthma hospitalisation risk and air pollutant concentration in Perth	
12h20	**Closing**		
12h45	**Lunch**		

MONDAY 13 SEPTEMBER 2010

ORAL PRESENTATIONS

SPATIAL MODELING WITH ENSEMBLES IN METRIC SPACE

C. Scheidt and J. Caers
Stanford University

State-of-the-art

The construction of earth models using geostatistical estimation and simulation methods is now routinely applied in areas such as reservoir modeling, aquifer modeling or mineral deposit estimation. These models can be constrained to a variety of data (seismic data, dynamic data, well logs, etc.). Due to the sophistication of modern geomodeling software, very complex Earth models containing complex surfaces describing layering and faults, reservoir structure, spatial variation of petrophysical properties, etc., can be constructed in a relatively straightforward manner. Increasing complexity, however desirable for accurate modeling, results in potentially hundreds to millions of different model parameters to describe the earth phenomenon of interest. The complexity and high-dimensionality of these models necessitate the construction of multiple model realizations (ensembles) to perform properly a study.

On the other hand, the goal of modeling studies is rarely the construction of the models themselves, but rather to respond to often basic engineering questions - do we clean-up or not, do we drill, what are oil and gas reserves, should we obtain more data, etc. In these and many other applications, the question to be answered is much less complex than the Earth models that are constructed. Our task then is to analyze multiple, complex, high-dimensional Earth models to answer often very simple questions. Analyzing ensembles of Earth models requires a response evaluation for each model in the ensemble (Monte-Carlo approach). This is generally infeasible, mostly due to the large number of uncertain parameters involved in Earth modeling. This is especially the case when the response function is CPU demanding, for example in cases where finite element or finite difference solutions are required (multiphase flow simulations, climate models).

What is missing in the modeling process is a link between the construction of the Earth models and the purpose of the study. Currently, models are considered simply as input to a transfer function, and are built with little regard to the type of response or decision variable considered. This no doubt has an impact on the efficiency and effectiveness of the studies we undertake.

Metric space

To overcome the disconnection between large-scale, 3D spatial Earth modeling and the actual purpose for which these models are constructed, we propose to introduce a metric space. This metric space is defined using a dissimilarity distance function between models. The distance is tailored to the application at hand - it should be chosen such that it correlates with the difference in response of interest between any two models. The dissimilarity distance function therefore relates the modeling process to the purpose of the modeling, and provides the key missing link

between model uncertainty, and uncertainty in the model responses (oil recovery, climate change, ground water recharge efficiency). If this distance is chosen with respect to the purpose of the modeling effort, then we can show how large complexity and dimensionality in Earth models can be significantly reduced in the metric space defined by this distance (Figure 1).

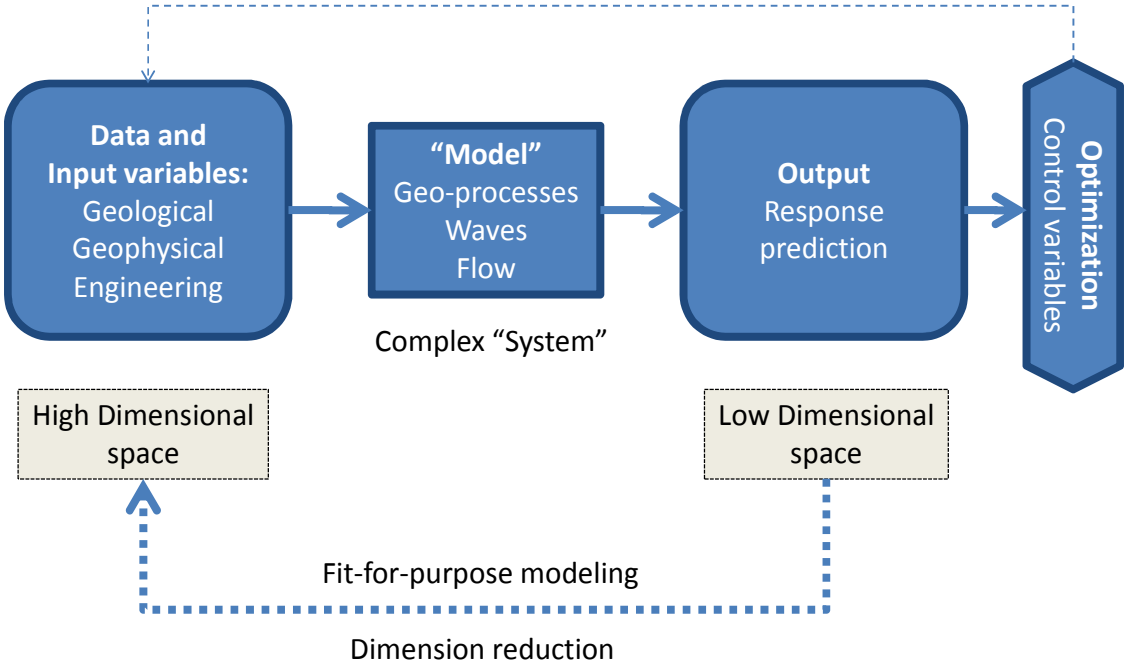


Figure 1: *Schematic diagram of modeling process and link with fit-for-purpose modeling using metric spaces*.

In this presentation, we will show that many of the current Cartesian-based modeling problems and methodologies, such as inverse modeling, model selection and screening, model updating and response uncertainty evaluation can be redefined in metric space. We demonstrate how such a redefinition greatly simplifies as well as increases effectiveness and efficiency of any modeling effort, particularly those that require addressing the matter of model and response uncertainty.
In particular, we will illustrate that such distance allows to:

1. Map complex high-dimensional models (model space) into a lower dimensional metric space (response-dependant) using a statistical tool called multi-dimensional scaling (MDS).
2. Incorporate the purpose of the application into the modeling effort.
3. Focus on the ensemble of earth models and not on one model at a time

We will lay out the basic theory behind modeling with ensemble of models in metric space. We will illustrate the simplicity and power of this novel theory by various practical applications on large dataset, including:

1. How to assess response uncertainty when the forward model is CPU demanding. Model selection can be done efficiently in the metric space in order to identify a few representative models for which response evaluation is performed.
2. How an ensemble of models can be parameterized with a few standard Gaussian variables through a distance-based Karhuenen-Loeve expansion (KLE). Such parameterization is response-dependant and therefore can be useful in solving optimization problems under uncertainty.
3. How an ensemble of models can be updated by using a modified version of the Ensemble Kalman Filtering technique in the metric space.
4. How a multi-resolution framework can be employed within distance-based techniques to circumvent CPU-intensive transfer functions. Distances can be computed from coarsened models while constructing equivalent fine-scale models.

These applications are possible whenever an effective distance between models can be determined. Indeed, we expect many more applications to come to light in this new, rich field of modeling of ensembles in metric space.

THE DIRECT SAMPLING METHOD TO PERFORM MULTIVARIATE MULTIPLE-POINTS GEOSTATISTICAL SIMULATIONS

G. Mariethoz[1,2], P. Renard[1], J. Straubhaar[1]

[1]Centre for Hydrogeology, University of Neuchâtel, 11 Rue Emile Argand, CP 158, CH-2000 Neuchâtel, Switzerland, E-mail: gregoire.mariethoz@minds.ch
[2]Stanford University, ERE Department, 367 Panama st, rm 65, Stanford, CA 94305, USA.

Recent advances in the domain of geostatistical simulations include multiple-points simulation methods (MPS), which infer spatial structures from conceptual example images, namely training images. Their main advantage is to allow using high-order spatial statistics (as opposed to traditional two-point statistics), that can represent a wide range of spatially structured, low entropy phenomena.

We present a novel MPS algorithm, named direct sampling (DS), which is drastically simplified compared to previous approaches and has more possibilities (Mariethoz and Renard, in press; Mariethoz, et al., submitted). It produces conditional realizations honoring the high-order statistics of uni- or multivariate training images. In addition to having virtually no RAM requirement, the method is straightforward to parallelize (Mariethoz, in press). Hence it is computationally efficient and can produce very large and complex realizations.

With the traditional MPS approaches (e.g. Strebelle, 2002; Zhang, et al., 2006; Straubhaar, et al., 2008), the training image is usually scanned and all pixels configurations of a certain size (the template size) are stored in a catalogue of data events having a tree or a list structure. This structure is then used to compute the conditional probabilities at each simulated node. Since the memory load increases dramatically with the size of the template and the number of facies, the use of MPS is generally restricted to relatively small problems with limited structural complexity. Moreover, the template is often not large enough to capture large-scale structures, and multi-grids have been introduced to palliate this problem by simulating the large-scale structures first, and later the small-scale features. Limited memory usage restricts the approach to the simulation of univariate fields.

The DS algorithm takes a different approach. Instead of counting and storing the configurations found in the training image, it directly samples the training image for a given data event. The method is based on a sampling method introduced by Shannon (1948), strictly equivalent to the original MPS algorithm of Guardiano and Srivastava (1993), but that does not explicitly need to compute conditional probabilities and therefore does not need to store them. Starting from an initial randomly located point, the training image is scanned. For each of the successive samples, the distance between the data event observed in the simulation and the one sampled from the training image is calculated. If the distance is lower than a given threshold, the sampling process is stopped and the value at the central node of the data event in the training image is directly used

for the simulation. Since nothing is stored, neighborhoods can have virtually any size and are not restricted to a template, making the use of multiple-grids unnecessary. In fact, multiple-grids are replaced by a continuous variation in the size of data events during the simulation process. This reduces artifacts and leads to a simpler implementation.

Similarly to other methods, DS can reproduce the structures of complex training images and deal with a wide range of non-stationary problems. Additionally, it can simulate both discrete and continuous variables by choosing specific distances between data events, and can deal with specific cases of non-stationarity. But the richest feature of DS is that multivariate data configurations can be considered, allowing to generate realizations presenting a given multivariate multiple-point dependence. The ability to treat multiple variables simultaneously in a multiple-point framework allows co-simulation without formulating the dependency between the simulated attributes. Instead, the multiple-point dependence is reproduced as it is in the multivariate training image.

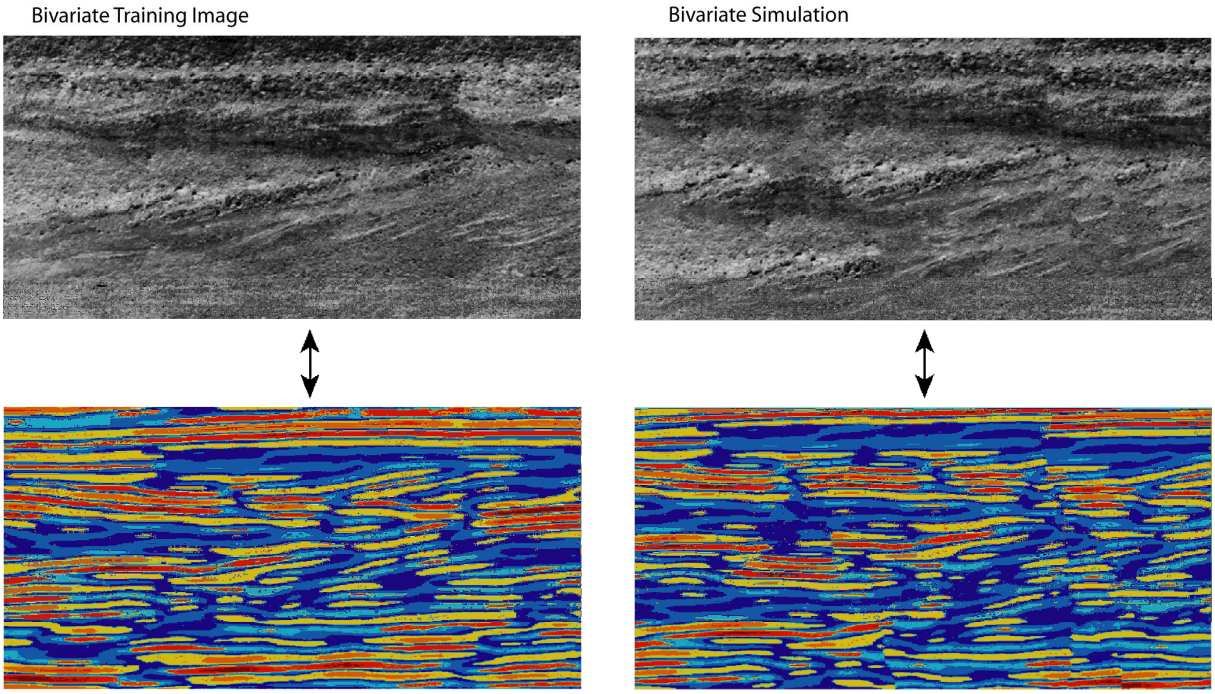


Figure 1: *Multivariate training image and one corresponding multivariate simulation*

An example is shown in figure 1. The left images represent a bivariate training image consisting of the photograph of an outcrop and a georadar image taken at the same location (Bayer, unpublished). The dependence between both variables is shown by the double-headed arrow. One bivariate realization is displayed on the right. Note that the top right image is not an actual photograph, but a simulation of what the photograph could be. Similarly, the bottom right image is a simulation of a georadar image. The dependence between variables is preserved (double-headed arrow), since the simulated photograph of the outcrop presents interfaces at locations that are coherent with the interfaces in the simulated georadar image.

References

Bayer, P. (unpublished), 3-D high-resolution characterization of a sedimentary gravel body for use in hydrogeological and reservoir modeling.

Guardiano, F., and Srivastava, M. (1993), Multivariate geostatistics: Beyond bivariate moments, in Geostatistics-Troia, pp. 133-144, Kluwier Academic.

Mariethoz, G. (in press), A general parallelization strategy for random path based geostatistical simulation methods, Computers & Geosciences.

Mariethoz, G., and Renard, P. (in press), Reconstruction of incomplete data sets or images using Direct Sampling, Mathematical Geosciences.

Mariethoz, G., Renard, P., and Straubhaar, J. (submitted), The direct sampling method to perform multiple-points simulations, Water Resour. Res.

Shannon, C. E. (1948), A mathematical theory of communication, The Bell system technical journal(27): 379-423.

Straubhaar, J., Walgenwitz, A., Renard, P., and Froidevaux, R. (2008), Optimization issues in 3D multipoint statistics simulation, paper presented at Geostats 2008, 1-5 Dec. 2008, Santiago, Chile.

Strebelle, S. (2002), Conditional Simulation of Complex Geological Structures Using Multiple-Point Statistics, Mathematical Geology, 34(1): 1-22.

Zhang, T., Switzer, P., and Journel, A. (2006), Filter-Based Classification of Training Image Patterns for Spatial Simulation, Mathematical Geology, 38(1): 63-80.

SIMULATING COMPLEX 3D ALLUVIAL STRATIGRAPHY BY COMBINING TRAINING- IMAGE-BASED ALGORITHMS AND VERTICAL VARIOGRAMS

A. Boucher[1], G.Phelps[2], C. Cronkite-Ratcliff[3]

[1]Dept of Environmental Earth System Science, Stanford University, CA, USA,
aboucher@stanford.edu
[2]U.S. Geological Survey 345 Middlefield Road, MS 989 Menlo Park, CA,
[3]Dept of Geological and Environmental Sciences, Stanford University, CA, USA,

Alluvial and volcanic basin deposits, prevalent throughout the southwestern United States, can be important groundwater repositories, and the deposit stratigraphy significantly affects groundwater storage, recharge and withdrawal. Mapping alluvial deposits in the subsurface is central to quantifying the availability and modeling the use of this critical resource. However, alluvial units observed at the surface are difficult to model within the subsurface because of their complex sedimentological structure. New tools are needed to map alluvium in the subsurface in order to improve models of soil moisture and groundwater recharge, flow, and transport, which in turn improve the understanding of available water resources.

Training image-based algorithms have been shown to offer great potential to both map and characterize the uncertainty associated with mapping subsurface contacts with complex connectivity. However, finding suitable 3D training images remains one of the main challenges for training image-based algorithms such as the SNESIM algorithm (Strebelle, 2002). Information for mapping 3D alluvial units is often available only in 2D or 1D, and creating explicit 3D training images may not be feasible or desirable. Moreover, training image requirements for standard simulation algorithms are restrictive (no trends or location specific features); consequently, images of natural systems are abandoned in favor of simplistic training images often generated by Boolean techniques. Furthermore, natural sedimentation processes and rates are known to change over time, creating sedimentological pattern trends within a basin that are not captured by simple vertically varying facies proportions.

These issues are addressed by using widely available geological maps as training images and a structured simulation path to mimic the sedimentary environments through time and space. Nonstationary 3D geomorphic patterns are simulated by coupling several 2D geological maps with experimental 1D vertical variograms; the geological maps are selected to model changes in sedimentation regime over time. These methods are demonstrated in Yucca Flat, Nevada, a nuclear test site where radionuclide migration is a concern for groundwater north of the Las Vegas, NV metropolitan area.

Two different training images, selected to represent different depositional regimes, are used as to mimic the length, sinuosity, and adjacency of geologic unit deposition that has occurred in Yucca Flat since the Pliocene. The first training image is the USGS quaternary geologic map of Yucca

Flat basin (Slate and others, 2000). It identifies patterns occurring in a hydrological regime where the sedimentation rate is low, as it is today (Figure 1 left). The second training image (Figure 1 right; D.M. Miller, USGS, written communication), close to Valjean, CA, in the Death Valley regional basin south of Yucca Flat, models a hydrological regime with an inferred higher sedimentation rate. Both training images are derived directly from observed geologic environments and therefore preserve the natural geologic shape, continuity, and stratigraphic relations. These complex training images are decomposed with the search tree partitioning approach (Boucher, 2008), where topographic variables, such as gradient, distance to source, and elevation from source, serve as a proxy for the sedimentary processes.

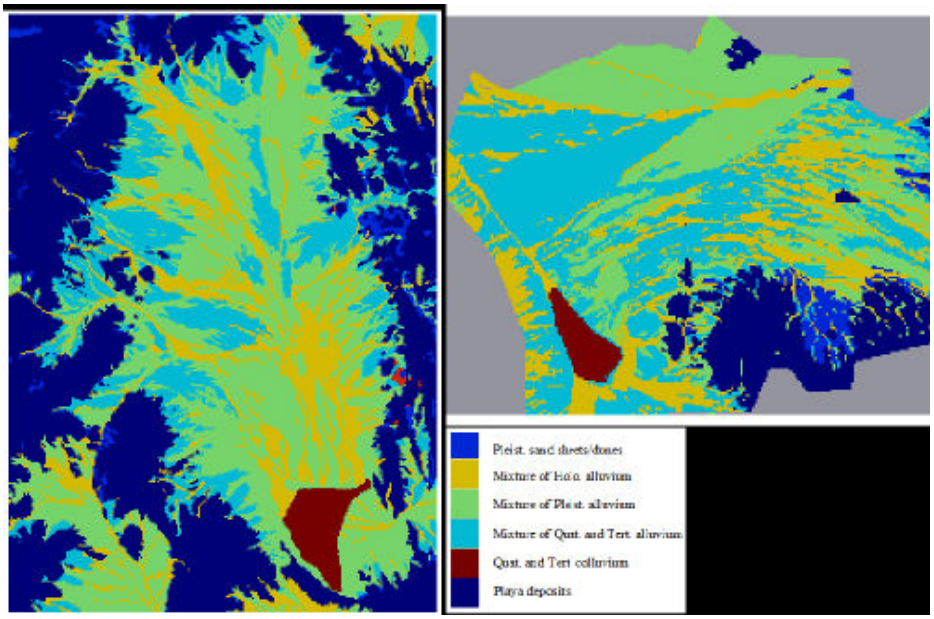


Figure 1: *Training image from the (Left) surface geology of the Yucca Flat (25km x 35 km) and (Right) the Valjean area in the Death Valley Basin (21.4km x 17.3km).*

Simulations are performed layer-by-layer, starting at depth and moving upwards. Each 2D layer is simulated by applying one of the two training images, using the search-tree partitioning method. For each layer a stochastic process determines the sedimentation rate by using one of the two training images. Contact boundaries are modeled vertically using indicator variograms inferred from borehole data and horizontally from a 2D training image. Future improvements to the method could seek to include conditioning the depositional rate to paleoclimate data.

The resulting 3D realizations (Figure 2) (a) are conditioned to borehole data, (b) visually appear to reproduce sedimentary structures, and (c) include variable sedimentation rates through time. The reproduction of the vertical variograms for the three most important sedimentary units is shown in Figure 3; the reproductions are good, but tend to under-estimate the short-scale, withinunit continuity.

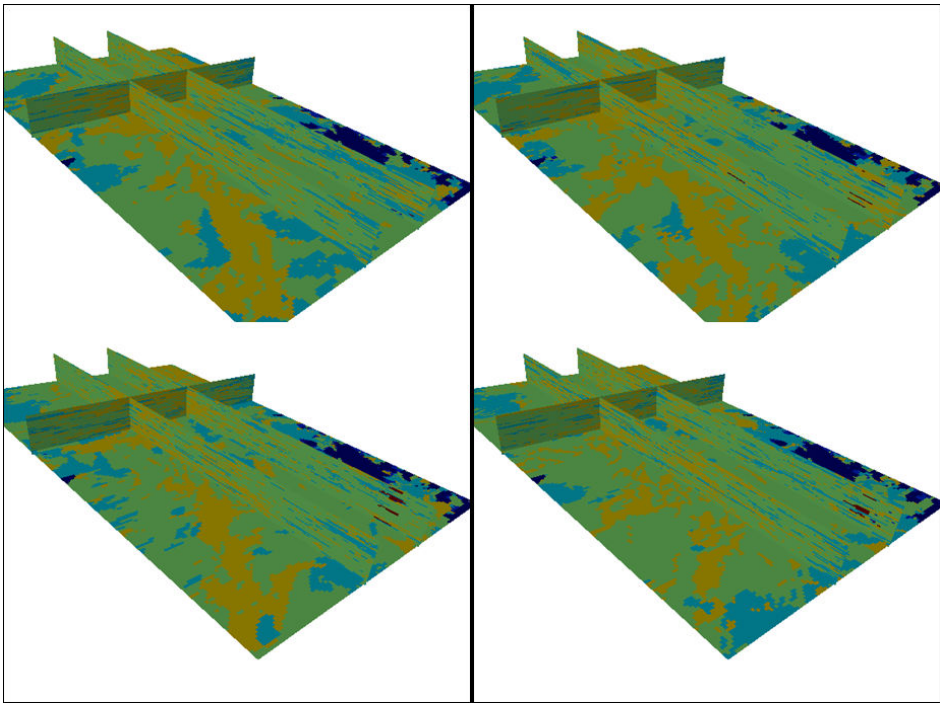


Figure 2: *Four simulations of the Yucca Flat: Dimension 20km by 12 km by 800 meters. See Figure 1 for the legend*

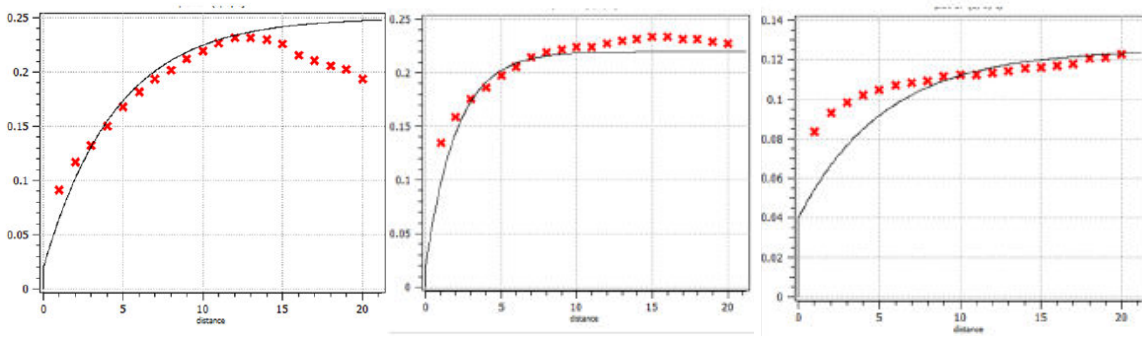


Figure 3: *Vertical variogram reproduction; (LEFT) Quaternary and Tertiary; (MIDDLE) Pleistocene; (RIGHT) Holocene*

Boucher, A., Considering complex training images with search tree partitioning. *Computers & Geosciences*, 2008, 35, 1151-1158

Slate, J.L., Berry, M.E., Rowley, P.D., Fridrich, C.J., Morgan, K.S., Workman, J.B., Young, O.D, Dixon, G.L., Williams, V.S., McKee, E.H., Ponce, D.A., Hildenbrand, T.G., Swadley, W, Digital Geologic Map of the Nevada Test Site and Vicinity, Nye, Lincoln, and Clark Counties, Nevada and Inyo County, California, Revision 4: U.S. Geological Survey, Open- File Report 99-554-A, http://pubs.usgs.gov/of/1999/ofr-99-0554, 1999

Strebelle, S. Conditional simulation of complex geological structures using multiple-point statistics *Mathematical Geology,* 2002*, 34*, 1-21

LIST–BASED ALGORITHM FOR MULTIPLE–POINT STATISTICS SIMULATION

Julien Straubhaar[1,5], Philippe Renard[1], Grégoire Mariethoz[2], Roland Froidevaux[3], and Olivier Besson[4]

[1]Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, rue Emile–Argand 11, CH–2009 Neuchâtel, Switzerland

[2]University of Neuchâtel; now at ERE Department, Stanford University, 367 Panama st, rm 65, Stanford, CA 94305, USA

[3]Ephesia Consult SA, 9, rue Boissonnas, CH–1227 Geneva, Switzerland

[4]Institute of Mathematics, University of Neuchâtel, rue Emile–Argand 11, CH–2009 Neuchâtel, Switzerland

[5]E-mail: julien.straubhaar@unine.ch, Phone: +41 32 718 26 10

August 16, 2010

1 Introduction

Flow and transport processes are highly sensitive to the shape of the structures and their connectivity (Journel and Zhang 2006; Renard 2007). Hence, modeling realistic heterogeneous geological reservoirs is a crucial issue. Multiple–point statistics is a flexible technique for simulating categorical variables, allowing to generate complex geological structures provided by the user via a training image (TI) (Guardiano and Strivastava 1993). In classical implementations, the multiple–point statistics inferred from the training image are stored in a search tree (Strebelle 2002). This provides a fast but very RAM demanding algorithm. Indeed, in particular for 3D cases, the search templates used for retrieving the statistics have to be small enough, because of the RAM limitation. Complex structures are then not properly reproduced, despite the use of multigrids (Tran 1994) for capturing large structures within the TI while keeping a search template of small size.

In this paper, we propose to replace the tree by a list. This new implementation presents several advantages. First, the list structure requires much less RAM than the tree structure, and the search templates can be chosen large enough to reproduce the complex structures within the TI properly. Furthermore, since the memory burden is low, all the lists required by a simulation can be stored simultaneously. In addition, a method inspired from Chugunova and Hu (2008) for handling non–stationary TI can be developped easily; in this case, we use a continuous auxiliary variable for describing the non–stationarity and extended lists taking into account for this secondary variable (Straubhaar et al. 2009). Finally, the list–based algorithm is straightforwardly parallelizable and hence also efficient in terms of CPU time.

2 Using lists for multiple–point statistics simulation

Here, we defined the list for a given TI having M facies numbered from 0 to $M-1$ (categorical variable s), and a given search template defined by its lag vectors, h_1,..., h_N. In this situation, an element of the list is a pair of vectors (d, c) where $d = (s_1, \ldots, s_N)$ defines a data event and $c = (c_0, \ldots, c_{M-1})$ is a list of occurrence counters for each facies. The number c_i is the number of replicates of the data events d found in the TI with facies i at the central node, *i.e.* the number of nodes v in the TI satisfying $s(v+h_1) = s_1, \ldots, s(v+h_N) = s_N$ and $s(v) = i$. Note that an additional virtual facies, *e.g.* -1, can be used for considering data event partially informed provided by nodes near the border of the TI. Such a list allows to compute the conditional probability distibution function (cpdf) needed for simulating a node u in the simulation grid and given by

$$\mathbb{P}(s(u) = k \mid d(u)) = \frac{\#\left\{v \in TI \;:\; s\left(v+h_{i_j}\right) = s\left(u+h_{i_j}\right),\ 1 \leqslant j \leqslant m \text{ and } s(v) = k\right\}}{\#\left\{v \in TI \;:\; s\left(v+h_{i_j}\right) = s\left(u+h_{i_j}\right),\ 1 \leqslant j \leqslant m\right\}}, \qquad (1)$$

where $u+h_{i_1}, \ldots, u+h_{i_m}$ are the informed (simulated) nodes in the neighborhood of u. This cpdf is simply computed by retrieving the counters of the compatible elements in the list.

3 RAM requirements

Only the complete data events found in the TI, corresponding to the leaves of the classical search tree, are stored in the list. In the case where M facies are present and a search template of size N is used, the classical search tree is a M–ary tree with N levels, then the reduction of RAM usage allowed by the list storage can be significative. For illustrating that purpose, we consider search templates of increasing size containing the N closest nodes to the central node, and we retrieve the size of the corresponding search trees and lists in gigabytes. The results are shown for two 3D TI in figure 1a and 1b, and the ratios between the size of the tree and the size of the list for these two cases are drawn on figure 1c. We rapidly observe a factor of 10 and more for the size of the tree in comparison with the size of the list.

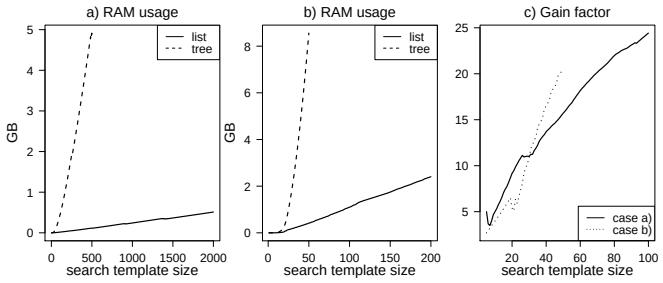


Figure 1: a) RAM usage for a 3D training image having 600′000 nodes and 4 facies; b) RAM usage for a 3D training image having 14′128′385 nodes and 6 facies; c) Gain factor with the list in comparison with the search tree (ratio: size of tree over size of list).

Illustrative example. We consider here a simple example using the well–known TI of Strebelle channels, 3 multigrids and 116 zones for rotations and dilatations (figure 2). The simulation requires 348 lists, one per zone and multigrid. Since the list structure is parsimonious, all of them can be stored simultaneously.

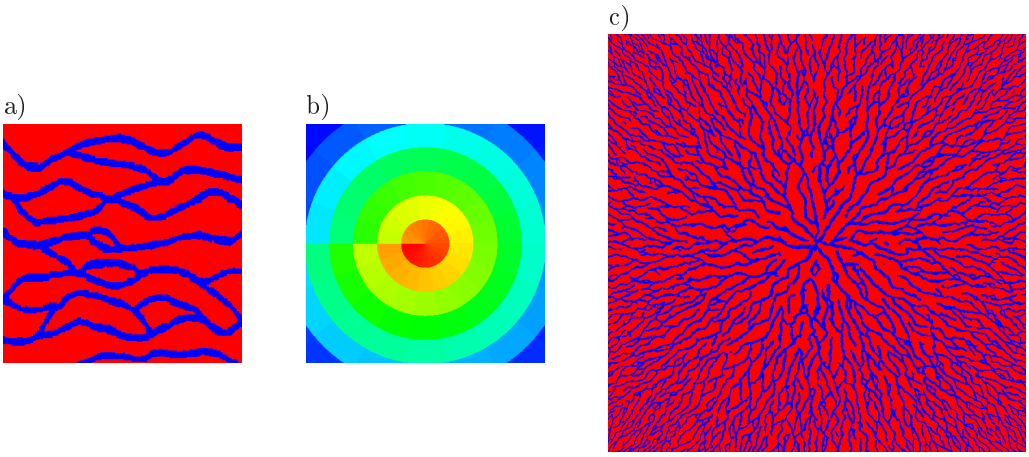


Figure 2: Multiple–zones simulation: a) Training image (Strebelle 2002) (250×250); b) Zones map ($1'000 \times 1'000$, 116 zones) for the simulation grid, defining orientation and dilatation: orientations go from 0^o to 360^o in each circle and dilatation from a factor of 1 to 0.4, from the centre to the borders; c) one simulation ($1'000 \times 1'000$).

4 Parallelization and CPU performances

We show in this section how a list–based multiple–point statistics algorithm can be parallelized on shared or distributed memory machines (using OpenMP or MPI technology). Assume that p processors are available. In this case, each list required by the simulation is divided in p parts of balanced size and each one of theses p partial lists is stored in the memory associated to a single processor. Then, the simulation of each node in the simulation grid is parallelized as follows. For computing the cpdf (1), each processor retrieves the needed information (counters) by scanning its own local list. The information retrieved by all processors is then merged (using a parallel computing communication), and the cpdf is computed. The simulated facies is then drawn from this cpdf, and the simulation grid is updated on all processors (communication). The resulting parallel list–based algorithm presents good performances in terms of CPU ressources: it approximatively spends the same amount of CPU time with two processors as a serial algorithm using search trees (Straubhaar et al. 2009).

References

Chugunova, T. and L. Hu (2008). Multiple–point statistical simulations constrained by continuous auxiliary data. *Mathematical Geosciences 40*(2), 133–146.

Guardiano, F. and R. Srivastava (1993). Multivariate geostatistics: beyond bivariate moments. In A. Soares (Ed.), *Geostatistics Troia*, Volume 1, pp. 133–144. Kluwer Academic, Dordrecht.

Journel, A. and T. Zhang (2006). Necessity of a multiple-point prior model. *Mathematical Geology 38*(5), 591–610.

Renard, P. (2007). Stochastic hydrogeology: what professionals really need? *Ground Water 45*(5), 531–541.

Straubhaar, J., P. Renard, G. Mariethoz, R. Froidevaux, and O. Besson (2009). An improved parallel multiple–point algorithm using a list approach. *Mathematical Geosciences*. submitted.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-points statistics. *Mathematical Geology 34*(1), 1–21.

Tran, T. T. (1994). Improving variogram reproduction on dense simulation grids. *Computers & Geosciences 20*(7), 1161–1168.

MULTIPOINT STATISTICS WITH AUXILIARY PROPERTIES APPLIED TO SUSTAINABLE DEVELOPMENT. EXAMPLE FROM URBAN EXPANSION PREDICTION OF PAU AGGLOMERATION (SOUTHWEST FRANCE)

R. Labourdette[1], T. Chugunova[2]

[1] TOTAL S.A. GSR/TG/G&I, Avenue Larribau, Pau, France, richard.labourdette@total.com, Tel: (0) 5 59 83 53 76, Fax: (0) 5 59 83 60 07

[2] TOTAL S.A. GSR/VDG, Avenue Larribau, Pau, France

In industrialized countries, the current pattern of urban development is increasingly taking the form of low-density, decentralized residential and commercial development. This form of development, the environmental and quality-of-life impacts of which are becoming central to debates over land use and land cover in urban and suburban areas is now commonly known as "sprawl" (Oguz et al. 2007). Many classic symptoms are loss and fragmentation of the natural resource, declining water quality and traffic congestion (Burchell et al. 1998). Land cover is an important element of ecological function (Wicklam et al., 2002). While urbanization has occurred, natural resource lands, such as forest, wetlands and agriculture, have been replaced by land uses with more impervious surfaces. Predicting future environmental consequences requires being able to predict the spatial pattern of land use change.

An emerging branch of geocomputing involves the modelling of spatial processes. A variety of techniques are being used, the most important being traditional cellular automata. Most of these models tempt to capture the evolving nature of real cities and region, but the most pressing concerns evaluation of their results and they cannot be strictly predictive.

As in numerous agglomerations, Pau is typically organised following fractal organisations. This organisation requires neighbourhood interaction but also long-distance spatial interactions to be simulated. Multipoint statistic algorithms are now sufficiently elaborated to simulate complex features such as these fractal organisations.

Our case study focuses on multipoint statistics application for urban expansion prediction. Multipoint statistics classically use training images in stationary mode. By integrating auxiliary properties, training images can evolve to a non-stationary state (Chugunova, 2008). This allows integrating various properties to guide multipoint simulations. The multi-realisation scheme associated with classical stochastic simulations enable to assess a variety of results and thereby a probability of urban expansion.

Initial dataset used are composed of aerial photographs taken respectively in 1997 and 2007. These images coupled with a digital elevation model are interpreted to create on one side the initial training image (Fig. 1a) and on the other side to calibrate multipoint statistic simulations (Fig.2a).

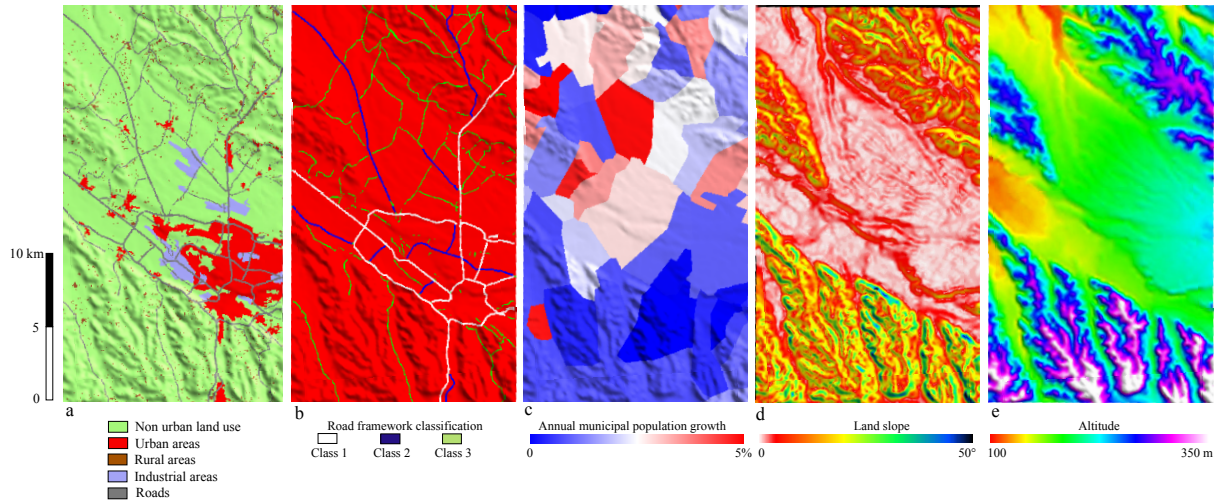


Figure 1: *(a) 1997 aerial photographs interpretations draped on the digital elevation model used as training image and associated auxiliary properties used for simulations: (b) Road framework classification, (c) Annual municipal population growth, (d) Landscape slope and (e) Altitude.*

A statistical analysis of land-uses according to altitude and land-slope shows good correlations between urban or industrial areas and topographic characteristics. One of important constraints of urban expansion is the road framework. To integrate this information in the simulation process, a complementary discrete property was added to represent this phenomenon. In order to predict the urban expansion, population growth statistics for each municipality between 1997 and 2007 were extrapolated, using the same extracted laws, until 2027.

Following these observations, the simulations are performed using altitude, land slope, road framework and local annual population growth as auxiliary properties (Fig. 1b to 1e). The simulation process is therefore guided by the constraints of different nature extracted from different sources. Neither combination nor transformation of those constraints is necessary. MPS algorithm allows to “predict” the type of land-use basing on neighbourhood land-use areas as well as the combination of constraints.

The results show the urban expansion as multi-realisations and the derived probability of this expansion. In order to verify road framework distribution and simulation algorithms, the road are simulated together with different land-use areas. In our case roads are strongly conditioned by the auxiliary property described as “road framework classification”.

Two simulations were run. The first one is estimating urban expansion for 2007, which can be compared with aerial photographs interpretation (Fig. 2a and 2b). The second one for 2027 to predict urban expansion with a constant population growth (Fig. 2c). The results show strong similarities between multipoint statistic realisations and aerial photographs interpretations (Fig. 2 a and 2b). This similarity is explained by the realistic population growth calculated between 1997 and 2007 for each municipality but also by the correct integration of altitude/slope maps

and road frameworks information: urban areas are developed in low altitude and slope, near to road frameworks. Predictive simulation for 2027 shows the same general land-use distribution, with a strong growth of small municipalities away from the main urban pole (Pau agglomeration) and thereby an extension of the "sprawl" effect. This observation is realistic according to Pau urban saturation. This prediction could lead to major traffic congestions in the northern and western part of the urban pole.

In this work a static modelling was used to represent an evaluation in time of a spatial structure. This is possible by integrating a property which represents a spatial evaluation. In our case it was an annual population growth. The proposed approach can be applied to other spatial processes to predict its evaluation while integrating constraints of different sources.

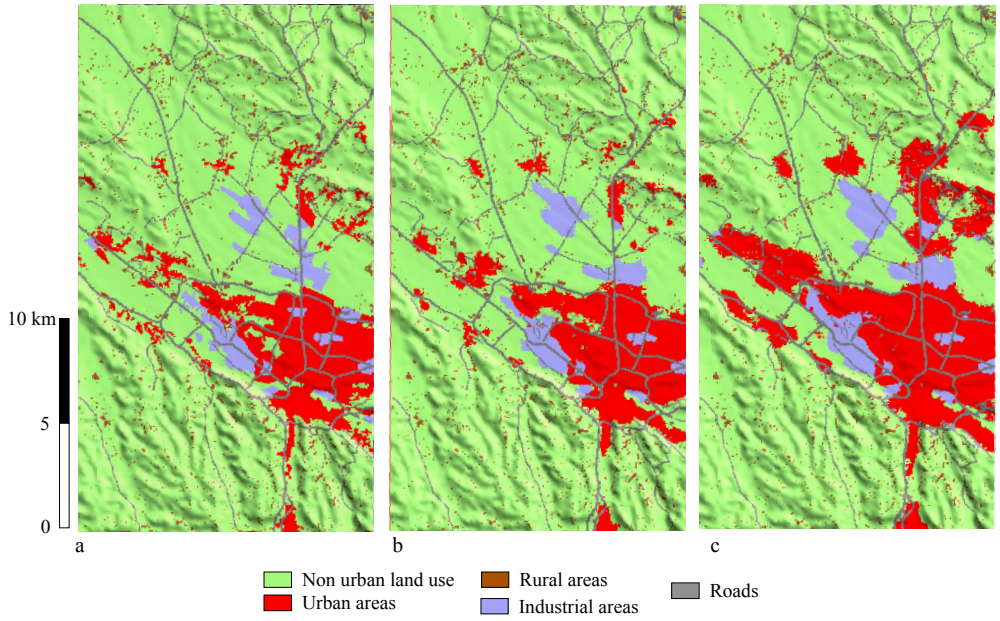


***Figure 2**: (a) 2007 aerial photograph interpretation draped on a digital elevation model and simulation results (most probable) for (b) 2007 and (c) 2027 urban distributions.*

References

Burchell RW, Shad NA, Philips H, et al (1998) The costs of sprawl-revisited. Transit Cooperation Research Program, Report 39, Washington DC, pp: 268

Chugunova T (2008) Contrainte des modèles génétiques de reservoirs par une approche de reconnaissance statistique de forme. PhD, Ecole des Mines de Paris

Hakan O, Klein AG, Srinivasan R (2007) Using the Sleuth urban growth model to simulate the impacts of future policy scenarios on urban land use in the Houston-Galveston-Brazoria CMSA. Research Journal of Social Sciences 2: 72-82

Wickham JD, O'Neill RV, Riitters KH, et al (2002) Geographic targeting of increases in nutrient export due to future urbanization. Ecological Applications 12: 93-106

DIRECT MULTIPLE-POINT GEOSTATISTICAL SIMULATION OF EDGE PROPERTIES FOR MODELING THIN IRREGULARLY-SHAPED SURFACES

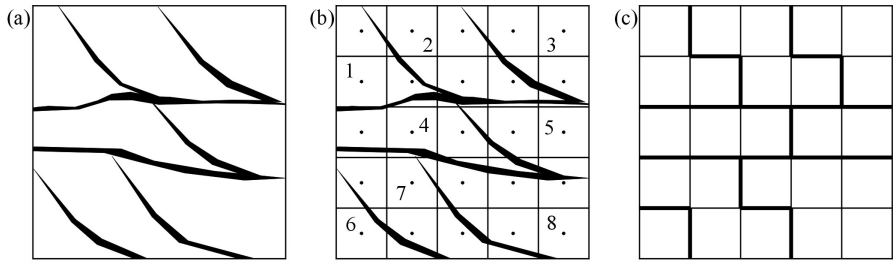
M. Huysmans[1], A. Dassargues[1,2]

1 KULeuven, Earth and Environmental Sciences, Leuven, Belgium,
marijke.huysmans@ees.kuleuven.be, Tel: 0032 16 326449, Fax: 0032 16 322980
2 Université de Liège, Department of Architecture, Geology, Environment and Civil Engineering (ArGEnCo), Hydrogeology and Environmental Geology, Liège, Belgium

Thin irregularly-shaped surfaces such as clay drapes often have a major control on flow and transport in heterogeneous porous media. Clay drapes are often complex curvilinear 3-dimensional surfaces and display a very complex spatial distribution. Variogram-based stochastic approaches are often also not able to describe the spatial distribution of clay drapes since complex, curvilinear, continuous and interconnected structures cannot be characterized using only two-point statistics. Multiple-point geostatistics aims to overcome the limitations of the variogram. The premise of multiple-point geostatistics is to move beyond two-point correlations between variables and to obtain (cross) correlation moments at three or more locations at a time using "training images" to characterize the patterns of geological heterogeneity. Multiple-point geostatistics is able to reproduce thin irregularly-shaped surfaces such as clay drapes but is often computationally intensive. To capture the thin surfaces, a small grid cell size should be adopted for the training image. This results in large training images and a large search template size and thus a large CPU and RAM demand (Huysmans and Dassargues, 2009).

This paper describes and applies a methodology to simulate thin irregularly-shaped surfaces with a smaller CPU and RAM demand than the conventional multiple-point statistical methods. The proposed method is based on the idea proposed by Stright et al. (2006) of using edge properties for upscaled flow simulation without losing information about thin irregularly-shaped flow barriers. To preserve these important fine-scale features at the flow simulation block scale, Stright et al. (2006) introduces an additional modelling variable as the edge of a model cell. The edge of a model cell is a continuous or categorical value associated with the cell face. This idea is extended in this paper to not only use edge properties in the flow simulation step but also in multiple-point geostatistical facies simulation step. This work explores the idea of directly simulating edge properties instead of pixel properties to make it possible to perform multiple-point geostatistical simulations with a larger cell size and thus a smaller computation time and memory demand. The proposed methodology of directly simulating edge properties is explained by applying it to a small and simple training image. Next, the method is applied to simulation of clay drapes in a real cross-bedded aquifer in Belgium.

The method starts from a training image depicting the thin irregularly-shaped surfaces of interest. The grid scale of this training image should be fine enough so that the thin surfaces can be depicted using pixels. Fig 1a shows a simple training image to illustrate the method. It shows sand bodies separated by a few horizontal and inclined clay drapes.



(d)

0	1	0	1	0
1	1	1	1	1
1	1	1	1	1
1	0	1	0	0
0	0	0	0	0

(e)

1	0	1	0	0
0	1	0	1	0
0	0	1	0	0
0	1	0	0	0
1	0	1	0	0

(f)

1	2	1	2	0
2	3	2	3	2
2	2	3	2	2
2	1	2	0	0
1	0	1	0	0

Second, the training image is indexed. Each large scale facies body is indexed individually, from 1 to the number of facies bodies, so that the thin irregularly-shaped surfaces of interest separate the individual facies bodies. In our example, each sand body is given a number from one to eight (Fig 1b). Third, a model grid size for simulation is chosen. This grid should be chosen so that it allows flow simulation within a reasonable computation time and so that it preserves the flow behaviour and connectivity of the relevant structures of the fine scale training image. This coarser grid is superimposed over the training image and grid nodes are placed at the center of each grid cell (Fig 1b). In the fourth step of the method, the grid is searched in the x, y and z-direction for the presence of clay drapes between grid nodes of the coarser grid (Stright et al., 2006). If the index value of the two nodes differs, then the edge between the two nodes is flagged as “1”, indicating that a clay drape is present. Otherwise the edge between the two nodes is flagged as “0”, indicating that no clay drape is present (Fig 1c). After this process, the edge properties can be represented by two matrices (or three 3D matrices in 3D) with the same size as the coarse grid: matrix B displaying the edge properties at the bottom of a grid cell (Fig 1d) and matrix R displaying the edge properties at the right side of a grid cell (Fig 1e). The edge properties at the top and left side of a cell are not stored since they can be obtained from the edge properties at the bottom and right side of neighbouring cells. In the fifth step of the method, matrices B and R are combined into one matrix E by applying the coding shown in Table 1 describing the four different possibilities for each cell. This allows obtaining a single training image displaying all edge properties. If, for example, a cell displays no flow barrier, this cell is

coded “0”. If, for example, a cell displays a flow barrier at the right and not at the bottom, this cell is coded “1”. The E-coded training image of our simple example is shown in Figure 1f. The E-coded training image can now be used as input training image to perform SNESIM simulations with 4 categories, i.e. the four possible outcomes for E. This E-coded training image (Fig 1f) is much coarser than the original training image depicting thin irregularly-shaped surfaces of interest using pixels (Fig 1a), while still preserving all relevant information about the presence of the clay drapes.

Table 1 *Definition of E-coding*

B	R	E
0	0	0
0	1	1
1	0	2
1	1	3

The proposed method is also applied to the cross-bedded Brussels Sands aquifer in Belgium. The small grid cell size training image (Fig. 2a) is converted into an edge training image with a larger grid cell size (Figure 2b). This edge training image is used as training image to perform SNESIM simulations with 4 categories. In this case, direct SNESIM simulation of edge properties is 60 times faster than ordinary SNESIM simulation of facies.

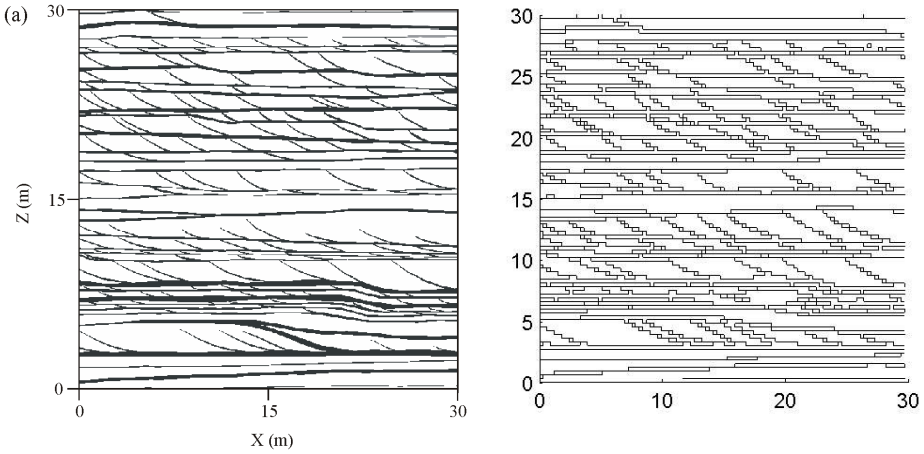


Figure 2 *(a) Vertical 2D training image of 30 m x 30 m in N40°E direction (white = sand facies, black = clay-rich facies) (adapted from Huysmans and Dassargues, 2009); (b) Edge training image*

Huysmans M. and Dassargues A., 2009, Application of multiple-point geostatistics on modeling groundwater flow and transport in a cross-bedded aquifer, Hydrogeology Journal 17(8), 1901-1911

Stright, L. , Caers, J. and Li, H. (2006) Coupled geological modeling and history matching of fine-scale curvilinear flow barriers, 19th SCRF annual meeting, Stanford Center for Reservoir Forecasting, Stanford, May, 2006.

RAPID CONSTRUCTION OF ENSEMBLES OF HIGH-RESOLUTION MODELS CONSTRAINED TO DYNAMIC DATA

C. Scheidt[1], J. Caers[1], Y.Chen[2] and L.J. Durlofsky[1]

[1] Energy Resources Engineering, Panama St., Stanford University, USA, scheidtc@stanford.edu
[2] Chevron Energy Technology Company, San Ramon, CA, USA

Introduction

Distance-based spatial stochastic modeling techniques have recently emerged in the context of ensemble modeling, in particular for data integration, model selection and uncertainty quantification. The objective of this work is to present a workflow which can rapidly generate multiple high-resolution models that are sufficiently constrained to dynamic data. This allows for a realistic assessment of uncertainty. We employ a distance-based ensemble modeling technique for data integration and propose a multi-resolution framework. Here distances are computed on coarse-scale models to efficiently evaluate transfer functions, while accurate fine-scale models are generated constrained to available dynamic data. Flow-based upscaling techniques are applied to generate the coarse models, and potential upscaling errors are accounted for in the workflow through an error modeling-based adjustment in the distance calculations.

Description of Method

Upscaling and Error Modeling. Ensemble modeling techniques usually require the computation of a transfer function **g** for each model in the ensemble. The function g is often a subsurface flow simulation, which for high-resolution models, can be CPU prohibitive. To reduce the computational intensity, we compute **g** on coarse-scale models $\mathbf{x}_i^c$ which are generated from the associated fine-scale models $\mathbf{x}_i$ (where i represents one model/realization) using flow-based upscaling techniques, which are widely applied in subsurface flow simulations (see e.g., Durlofsky, 2005).

Although there exist a variety of upscaling techniques which can provide adequate accuracy in many cases, upscaling may induce some errors in the coarse models and thus in the transfer functions. The objective here is to appropriately account for these errors in an efficient manner, e.g., without evaluating **g** at the fine scale. We introduce an error modeling approach based on the assumption that models with similar responses on the coarse scale result in similar upscaling errors. Thus, if we group the models with similar coarse-scale responses, we can estimate the error from a typical model in the group, and associate the estimated error to other models in that group. A k-means clustering technique (see Scheidt and Caers, 2009a) is applied to classify models with similar coarse-scale responses in order to estimate the upscaling error **e**. We then assign a single upscaling error $\mathbf{e}_k = \mathbf{g}(\mathbf{x}_k) - \mathbf{g}(\mathbf{x}_k^c)$ to every model in a particular cluster k. Note

that the estimation of error thus requires only a few fine-scale flow simulations. If this error estimation is not sufficiently accurate, new fine-grid simulations can be iteratively added in order to obtain a better estimation of **e** for each cluster.

Distance-Based Modeling. The purpose of distance-based modeling is to parameterize the large variability between reservoir models in a given ensemble. A distance relevant to data integration is simply the difference in responses between any two models in the set. Then, once we have an estimate of the upscaling error **e** for each coarse-grid model, we can approximate the distance between any two high-resolution models as the distance between the coarse models, corrected for upscaling error: $d_{ij}^{c+e} = d(\mathbf{g}(\mathbf{x}_i^c) + \mathbf{e}_i, \mathbf{g}(\mathbf{x}_j^c) + \mathbf{e}_j) \ \forall i,j$. In this paper, we apply a post-image problem in the multi-resolution framework. The post-image problem is a distance-based modeling technique that provides a stochastic parameterization of an ensemble of models. It constructs a distance matrix and employs the classical Karhunen-Loeve expansion (KLE) and kernel-based techniques, which allows new realizations to be generated and adjusted to new data (Scheidt *et al*., 2008). A full account of this approach is published in Caers *et al*. (2010).

Application

The proposed workflow is applied to a 3D test case, involving 100 realizations of log-normally distributed permeability fields. We consider a five-spot pattern, where the injector and four producers operate at fixed bottom hole pressures. The fine-scale models of dimensions 55×55×25 are uniformly coarsened to 11×11×5 using standard local permeability upscaling. In this particular example, the use of coarse-scale models introduces some errors. This results from the relatively large upscaling ratio and because we do not apply near-well upscaling. Figure 1a shows the estimated P10, P50 and P90 quantiles of the field water rate, for fine and coarse-scale flow simulations. Figure 1b shows the corrected coarse-scale predictions.

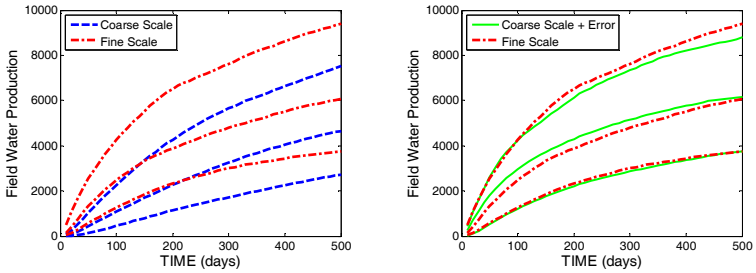


Figure 1: *P10, P50 and P90 quantiles of field water production for (a) fine and coarse-scale models, (b) fine and corrected coarse-scale flow simulations*.

The objective is to construct new high-resolution models that are constrained to water production at each well. In order to estimate the upscaling errors, six fine-scale simulations are performed based on the k-means clustering. The post-image problem is then applied to construct new models that are constrained to the water injection and production data.

Figure 2a shows the field water production for the initial models (gray lines), the ten new models (red lines) and the production data (green dots). Figure 2b displays histograms of the objective function (distance to the "truth") for the initial models (red) and the newly generated models (blue). We observe that the new models show a clear improvement (i.e., smaller objective function values), in terms of agreement with the dynamic data, compared to the initial ensemble.

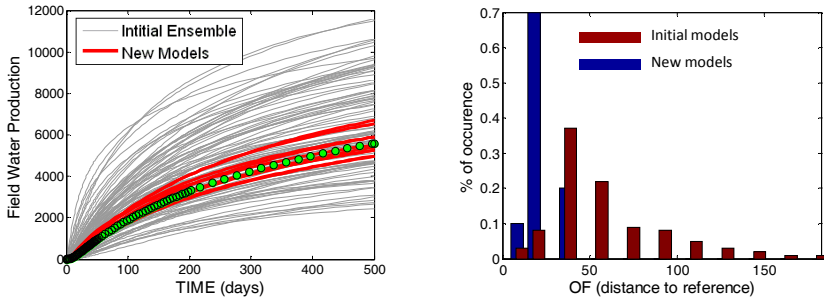


Figure 2: *(a) Field water production for initial and new models, (b) Histogram of the objective function for the initial ensemble of models and the new set of models*

Conclusions

This paper presents a novel method to generate ensembles of models using distance-based stochastic modeling techniques in a multi-resolution framework. Starting with an initial set of model realizations, we demonstrate how to construct new high-resolution models that are constrained to dynamic data by evaluating the transfer function on the associated coarse models. The method employs distance-based modeling techniques, where the distance between any two models is defined from the coarse-scale simulations. An error modeling procedure is introduced in the distance calculations to account for the potential error in upscaling. The method is general in that it can handle simulation errors, due to the use of either coarse-scale models or proxy models, through the updating of the distance function. We note finally that the method can be applied not only to the post-image problem, but also to other distance-based techniques, such as distance-based Ensemble Kalman Filter (Park *et al*., 2008).

References

Caers J., Park K. and Scheidt C. (2010), Modeling Uncertainty of Complex Earth Systems in Metric Space. Handbook of Geomathematics, eds. T. Sonar, Z. Nashed and W. Freeden, in press

Durlofsky, L. (2005), Upscaling and Gridding of Fine Scale Geological Models for Flow Simulation, 8th International Forum on Reservoir Simulation, Iles Borromees, Italy, June 20-24

Park, K., Scheidt, C., and Caers, J. (2008). Simultaneous Conditioning of Multiple Non-Gaussian Geostatistical Models to Highly Non-Linear Data Using Distances in Kernel Space, VIII International Geostatistics Congress, Santiago, Chile, December 1-5

Scheidt, C., Park, K. and Caers, J. (2008). Defining a Random Function from a Given Set of Model Realizations, VIII International Geostatistics Congress, Santiago, Chile, December 1-5

Scheidt C. and Caers J. (2009a). Representing Spatial Uncertainty Using Distances and Kernels, Mathematical Geosciences 41(4): 397-419

CONDITIONAL MULTIPLE POINT SIMULATION WITH TEXTURE SYNTHESIS

A. Parra[1,2], J. M. Ortiz[1,2]

[1]ALGES Lab, Advanced Mining Technology Center, Universidad de Chile
[2]Department of Mining Engineering, University of Chile, Av. Tupper 2069, Santiago, Chile,
aparra@dcc.uchile.cl, jortiz@ing.uchile.cl,+56(2)9784585

Abstract

Understanding the spatial distribution of categorical data is relevant, since many processes in the Earth Sciences depend upon the connectivity of facies, soil and rock types. Geostatistical simulation aims at characterizing the spatial distribution and its uncertainty by generating alternate scenarios of these variables, exploiting the spatial correlation. This is traditionally done by using a variogram or spatial covariance, but these methods have proven to be insufficient in some cases, where the connectivity shows patterns that are not well reproduced with these traditional techniques. Multiple point simulation provides a new set of tools to characterize these variables. The goal is to reproduce the relationships of points in specific patterns, so that globally the generated realizations show the desired connectivity, which is corroborated by the statistical reproduction of specific patterns frequencies (Strebelle, 2002; Ortiz, 2008).

We present a conditional algorithm based on texture synthesis (Wei, 2001; Daly and Knudby, 2007) and discuss an approach to parallelize its implementation. The idea in computer vision is to reproduce a texture, provided in a training image. The main difference is that the computer vision problem is not concerned with conditioning to hard data.

The algorithm works by defining a unilateral path with a pattern containing a causal and non-causal region (Figure 1). The highlighted node is simulated conditional to all previously simulated information in the template, called causal region. Conditioning is ensured by adding a non-causal region, which looks ahead of the unilateral simulation path and ensures that all simulated values are consistent with conditioning data within the template. Pattern statistics are retrieved from the training image and the frequencies of the patterns are stored in a linked list. The distribution of categories in the causal region is used as a key to identify patterns that are to be used during simulation. Conditioning data in the non-causal region are used as a filter when the frequencies are retrieved.

The simulation grid is visited in order, rather than randomly, so that the causal region of the pattern is always informed by previously simulated nodes (except at the beginning). The non-causal region informs about conditioning data.

At every visited node, the information of the causal pattern is gathered (categories) and the frequencies of a set of similar patterns are retrieved from the training image. If conditioning information is available in the non-causal region, only the patterns that honor the hard data are considered. The conditional probabilities for the categories are computed using the most similar pattern, using a weighted distance scheme. A value is drawn from the conditional distribution and added as conditioning information in the causal region for subsequent nodes.

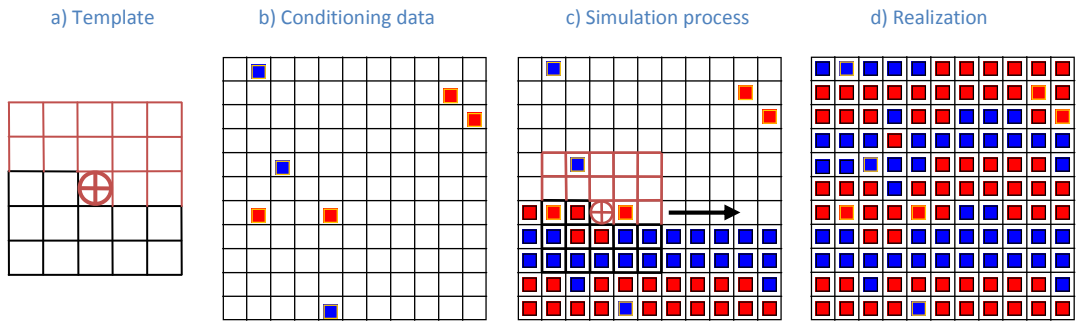


Figure 1: *Illustration of the implementation of the causal and non-causal region for conditioning during simulation.*

The algorithm works well and is able to reproduce long range connectivity with a relatively small search pattern. Furthermore, this implementation does not require using a multiple grid approach (Parra and Ortiz, 2009).

To improve the algorithm performance, parallelization can be achieved in two ways. The first approach consists in parallelizing at the realization level, that is, computing each realization in a separate process, avoiding all communication between processes. The second and most challenging way is to parallelize at the path level, in order to simulate a single realization using multiple processes. This approach is described next.

The domain to be simulated must be divided into regions, so that in every region, one or more nodes can be simulated simultaneously, without communication problems. Since the path to visiting the nodes is fixed, we know which conditioning nodes are required to simulate a specific location. Parallelization is achieved by proceeding row-wise (Figure 2). The single-process algorithm starts from the first node at the bottom row of the realization grid. According to the causal-template, it is possible to define how many nodes r have to be simulated before simulating nodes in the second row. When r nodes have been simulated in the first row, process 2 can start simulating the first node of the second row, since the required conditioning information is available. Simultaneously, the $r+1$ node in the first row is being simulated by process 1. This approach can be repeated in order to achieve a pre-defined maximum of concurrent processes.

The speed up of this approach depends on the number of processes run in parallel, the size of the search template and of the simulation grid. Some examples are presented to demonstrate the algorithm performance.

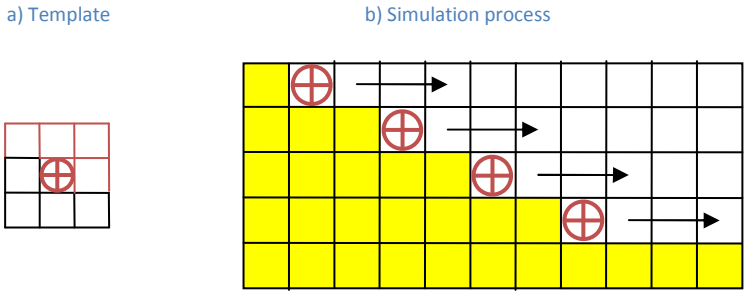


Figure 2: *Illustration of parallel simulation. Yellow zone represents simulated nodes, the 4 marked nodes will be simulated at the same time.*

References

Daly C, Knudby C (2007) Multipoint Statistics in Reservoir Modelling and in Computer Vision. Petroleum Geostatistics 2007, A32

Ortiz JM (2008) An Overview of the Challenges of Multiple-Point Geostatistics. In Geostats 2008 – Proceedings of the Eighth International Geostatistics Congress, JM Ortiz and X Emery (eds.), Gecamin Ltda., Santiago, Chile, vol. 1, p. 11-20

Parra A, Ortiz JM (2009) Conditional Multiple-Point Simulation with a Texture Synthesis Algorithm. IAMG 09 Conference, Stanford University

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34(1): 1-21

Wei LY (2001) Texture Synthesis by fixed neighborhood searching. Ph. D. thesis, Stanford University, 132 p

CLASSIFYING WAVE FORECASTS WITH MODEL-BASED GEOSTATISTICS AND THE AITCHISON DISTRIBUTION

R. Tolosana-Delgado[1], J.J. Egozcue[2], A. Sánchez-Arcilla[1], J. Gómez[1]

[1]Maritime Engineering Laboratory (LIM), raimon.tolosana@upc.edu, agustin.arcilla@upc.edu, jesus.gomez@upc.edu, phone: (+34 93 401) 6468, fax: (...) 1861

[2]Applied Mathematics III, juan.jose.egozcue@upc.edu, phone: (...) 6908, fax: (...) 1825

[1,2]Technical University of Catalonia (UPC), c/Jordi Girona 1-3, Barcelona, Spain.

Introduction and goal

Maps of the numerical forecasts of waves at the Western Mediterranean Sea during the last 5 years are stored at the LIM premises. We would like to find an automatic way to classify these maps in the four dominant regimes relative to the interaction of waves with the Catalan Coast: calm situations, E-storms, N/NW-storms, and SW-storms. For this goal, each forecast map is represented by some *characteristic features*, the wave energy (actually, the log of *significant wave height*) at $p = 8$ pre-specified pixels. Thus each forecast becomes a point $\vec{x}_i$ in the *feature space* $\mathbb{R}^p$, a p-dimensional real space. Notice that the feature space is not the geographic space. $N = 970$ of these maps has been classified by experts in the $D = 4$ regimes. We aim thus to assign a probability of classification in a regime class to each point of the feature space. In such cases, we commonly apply a discriminant technique. However, these methods produce probability vectors with a component near to 1 (highly informative) at feature points very far from the observed ones. Geostatistical interpolation techniques (e.g. Indicator Kriging (IK), Journel, 1983) would be better adapted to our needs: results would become non-informative probabilities beyond the range of influence of the available data, and somehow "follow" them within their range. However, classical IK is discarded because it frequently delivers negative probabilities. In this contribution we consider that the probability that a map i (as a whole) belongs to each of the four groups is a vector $\mathbf{z}(\vec{x}_i)$ of the four-part simplex $\mathcal{S}^D$, the sample space of D-part probability vectors (Aitchison, 1986). We assume that these are realizations of a random function $\mathbf{Z}(\vec{x})$, defined on the feature space $\mathbb{R}^p$. For each of the N labelled forecasts, we observe one of the D categories, as an indirect information about $\mathbf{z}(\vec{x}_i)$. A Bayesian approach within a model-based framework (Diggle et al., 1998) is accordingly used to estimate the probabilities $\mathbf{z}(\vec{x}_i)$ for labelled forecasts $\vec{x}_i$. These probabilitis can be then interpolated to any feature-point.

Prior and data model

Diggle et al. (1998) present a fairly similar case, in a success/failure-classification case. They assume that the number of successes observed at each location $\vec{x}$ follows a binomial model with underlying probability $p(\vec{x})$, independent between locations. The spatial dependence is introduced by assuming that $\log(p/(1-p))(\vec{x})$ is a Gaussian random

field, second-order stationary. To generalize this model to D categories, we make use of the Euclidean space structure of $\mathcal{S}^D$ introduced by, e.g. Pawlowsky-Glahn and Egozcue (2001): D-component probability vectors $\mathbf{z}$ are fully identified with their *orthonormal coordinate vectors* $\boldsymbol{\zeta}$, computed as $\boldsymbol{\zeta} = \mathbf{V} \cdot \ln(\mathbf{z})$, and inverted with $\mathbf{z} = \mathcal{C}[\exp(\mathbf{V}^t \cdot \boldsymbol{\zeta})] = \exp(\mathbf{V}^t \cdot \boldsymbol{\zeta})/\mathbf{1}^t \cdot \exp(\mathbf{V}^t \cdot \boldsymbol{\zeta})$, where logs and exponentials operate component-wise, and $\mathbf{V}$ is a $(D-1) \times D$ *matrix of contrasts*, i.e. such that its rows sum up to 0, $\mathbf{V} \cdot \mathbf{1} = \mathbf{0}$, and form a set of orthonormal vectors. Take now the observed labels $\mathbf{J}(\vec{x}_i)$ as vectors of 3 zeroes and 1 one, placed at the component of the category observed at $\vec{x}_i$. We assume for them a multinomial model $\mathbf{J}(\vec{x}_i) \sim \mathcal{M}u(n = 1, \mathbf{z}(\vec{x}_i))$, again independent between locations. The log-likelihood of $\mathbf{z}$ is then $\ln L(\mathbf{z}|\mathbf{J}) = \kappa_{\mathcal{D}}(\mathbf{J}) + \sum_i \mathbf{J}^t(\vec{x}_i) \cdot \ln(\mathbf{z}(\vec{x}_i))$, or in terms of coordinates $\boldsymbol{\zeta}$,

$$\ln L(\boldsymbol{\zeta}|\mathbf{J}) = \kappa_{\mathcal{D}}(\mathbf{J}) + \sum_{i=1}^{N} \mathbf{J}^t(\vec{x}_i) \cdot \ln \frac{\exp(\mathbf{V}^t \cdot \boldsymbol{\zeta}(\vec{x}_i))}{\mathbf{1}^t \cdot e^{\mathbf{V}^t \cdot \boldsymbol{\zeta}(\vec{x}_i)}} = \kappa_{\mathcal{D}}(\mathbf{J}) + \sum_{i=1}^{N} \mathbf{J}^t(\vec{x}_i) \cdot \mathbf{V}^t \cdot \boldsymbol{\zeta}(\vec{x}_i) - g(\boldsymbol{\zeta}) \quad (1)$$

where $g(\boldsymbol{\zeta}) = \sum_{i=1}^{N} \ln(\mathbf{1}^t \cdot \exp(\mathbf{V}^t \cdot \boldsymbol{\zeta}(\vec{x}_i)))$. The prior distribution for $\mathbf{z}(\vec{x})$ is stated in terms of its coordinates, by saying that the vector of coordinates $\boldsymbol{\zeta}(\vec{x})$ forms a multivariate Gaussian random field. Thus, the prior density $\pi^0(\boldsymbol{\zeta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ of the vector $\boldsymbol{\zeta} = [\boldsymbol{\zeta}(\vec{x}_i), i = 1, \ldots, N]$ satisfies

$$\ln \pi^0(\boldsymbol{\zeta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \kappa_{\mathcal{N}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \boldsymbol{\mu}^t \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\zeta} - \frac{1}{2} \boldsymbol{\zeta}^t \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\zeta}, \quad (2)$$

as usual for normal variates. For our application, we assume that $\boldsymbol{\zeta}(\vec{x})$ is second order stationary, with known mean $\boldsymbol{\mu}(\vec{x}) = \mathbf{V} \cdot \ln(0.5, 1, 1, 1)$ (i.e., giving a priori lower likelihood to calm situations, which are less dangerous but over-represented in the data set) and covariance $\Sigma_{ij}(\vec{x}_n, \vec{x}_m) = C_{ij} \times \rho_M(\vec{x}_n - \vec{x}_m|\alpha, a)$ given by linear model of corregionalization of an identity matrix $C_{ij} = \delta_{ij}$, and a Matérn correlogram, with smoothness $\alpha = 2.5$ and an effective range of $a = 0.4$ (e.g. Diggle et al., 1998).

Posterior distribution for the probability vector

By Bayes Theorem, we add Eqs. (1) and (2) to obtain the log-posterior density for $\boldsymbol{\zeta}$

$$\ln \pi(\boldsymbol{\zeta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{J}) = \kappa(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{J}) + \boldsymbol{\mu}^t \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\zeta} - \frac{1}{2} \boldsymbol{\zeta}^t \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\zeta} + \sum_{i=1} \mathbf{J}^t(\vec{x}_i) \cdot \mathbf{V}^t \cdot \boldsymbol{\zeta}(\vec{x}_i) - g(\boldsymbol{\zeta}), \quad (3)$$

i.e. an Aitchison distribution (Aitchison, 1986). We may then estimate the coordinate vector of probabilities by maximizing Eq. (3), equivalent to solving the non-linear system of equations $\boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\zeta} - \hat{\boldsymbol{\mu}}_{MP}) = \sum_i \mathbf{V} \cdot (\mathbf{J}(\vec{x}_i) - \hat{\mathbf{z}}_{MP}(\vec{x}_i))$. Fig. 1(top) shows these maximum posterior estimates $\hat{\boldsymbol{\zeta}}_{MP}(\vec{x}_i)$ obtained for a training set of 1/3 of the sample size N.

Interpolating the probabilities

Up to here, we obtained probability estimates for the labelled feature combinations (in geostatistical jargon, the *sampled locations*). But we can easily interpolate these estimates

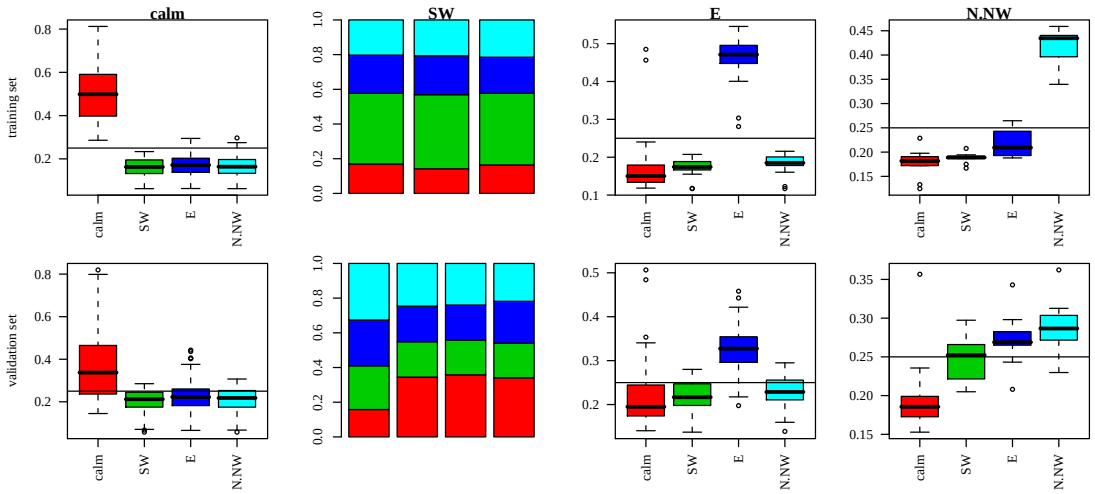


Figure 1: Probability estimates for each dominant regime: maximum posterior estimates for the training set, and interpolated with simple kriging for the validation set. The training and validation sets contain a stratified sample of 1/3 and 2/3 of the data.

to an unlabelled feature $\vec{x}_0$ (unsampled location). Note that the vectors of observations $\mathbf{J}$, of latent probabilities $\mathbf{z}$ (and its mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$) can be considered to have a block related to the non-labelled cases: the only difference is that for the unlabelled block $\mathbf{J}(x_0) = \mathbf{0}$. Then the posterior (Eq. 3) can be factorized in the sum of a term (with an Aitchison distribution) linked to sampled locations, and another term (with an unchanged prior distribution) related to the unsampled location. Because the second term carries no further information beyond the prior, the maximization of the first term can be obtained as before ignoring the block $\mathbf{J}(x_0) = \mathbf{0}$. Finally, conditional on the values estimated for the first block, note that the second block is a normal on the simplex with known mean and variance. Thus the BLUE for $\boldsymbol{\zeta}(\vec{x}_0)$ is obtained by simple kriging of $\hat{\boldsymbol{\zeta}}_{MP}(\vec{x}_i)$ at sampled locations, i.e. by simplicial IK (Tolosana-Delgado et al., 2007). Fig. 1 shows also the obtained probabilities against the actual labels of the validation set.

References

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). 416 pp.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed (1998). Model-based geostatistics. Journal of the Royal Statistical Society, Series C 47: 299–350.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. Mathematical Geology 15: 445–468.

Pawlowsky-Glahn, V. and J. J. Egozcue (2001). Geometric approach to statistical analysis on the simplex. Stochastic Environmental Research and Risk Assessment 15: 384–398.

Tolosana-Delgado, R., V. Pawlowsky-Glahn, and J. J. Egozcue (2007). Indicator kriging without order relation violations. Mathematical Geology 40: 327-347.

APPLICATION OF MULTIPLE-POINT GEOSTATISTICS IN SOIL SCIENCE: THE RECONSTRUCTION OF POLYGONAL NETWORKS OF ICE-WEDGE PSEUDOMORPHS

E. Meerschman, M. Van Meirvenne

Research Group Soil Spatial Inventory Techniques, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, B-9000 Gent, Belgium, eef.meerschman@ugent.be, Tel: 0032 09 2646042, Fax: 0032 09 2646247

Introduction

The recently developed multiple-point geostatistical tool SNESIM aims at reconstructing spatially continuous patterns by replacing the traditional two-point variogram by a multiple-point training image (TI) (Strebelle, 2002). Although several applications can be found in petroleum- and hydrogeology, little research has been done on the use of multiple-point geostatistics (MPG) in soil science. However, also in soil science repetitive or curvilinear patterns appear that are hard to model with the traditional two-point geostatistical algorithms.

Polygonal networks of ice-wedge pseudomorphs are an example of such repetitive soil features. These structures form hexagonal networks which are the imprint of thermal contraction cracks formed under periglacial circumstances. They were filled with ice and wind-blown sediment, resulting in wedge-shaped bodies extending several meters deep into the soil. At present their pseudomorphs are found across much of northern Europe. They are indicative for the severity of the paleoclimate during glacial periods since they are considered as indicators for permafrost (Ewertowski, 2009).

Our aim was to evaluate the performance of traditional and multiple-point geostatistical algorithms in the interpolation of the presence of ice-wedge pseudomorphs as indicator data. First, we interpolated point data without prior knowledge, using ordinary kriging (OK) with a variogram model fitted to the experimental variogram. Then, we translated prior knowledge into a conceptual wave variogram model for OK and a polygonal network TI for SNESIM. For each method we assessed how the pattern reconstruction and the local accuracy of the estimation maps changed with an increasing number of data points.

Construction of a reference image

As a test case, we selected a field in the sandy silt region of Flanders, Belgium. The field was chosen because of a seldom available aerial photograph showing crop marks indicating the occurrence of pseudomorphs. To observe the differences in soil composition between the Tertiary host material and the Quaternary wedge filling, a part of the field (ca. 6 m by 6 m) was excavated to a depth of 90 cm. The host material had a sandy clay texture and dated from the Early Eocene-Ypresian (55.8 Ma - 48.6 Ma), whereas the younger wedge filling clearly had a larger sand percentage.

To map the wedges, the field was measured with an electromagnetic induction soil sensor (EM38-MK2) (Cockx et al., 2006). After the noise and a significant trend had been removed

from the electrical conductivity measurements, the experimental variogram of the residuals showed a hole effect which is characteristic for repetitive features. To create a binary reference map, a coefficient of variation filter (kernel 3x3) was run on the interpolated map followed by a reclassification (Figure 2-left). From this binary reference map three indicator data series of each 13 data sets with an increasing number of data (from 84 to 840) were extracted according to a stratified random sampling scheme.

Reconstruction with traditional and multiple-point geostatistics

Without prior knowledge, the most obvious way to create an estimation map based on indicator data for two categories is OK. The classified kriged map using 168 indicator data and a spherical variogram model fitted to the experimental variogram (Figure 1a) is shown in Figure 2. As could be expected, the polygonal network was not well reconstructed. It is known that the diameter of the pseudomorph hexagonal cells is between 5 and 15 m (Ghysels, 2008). Hence, we interpolated the same data points with a conceptual wave variogram model (Figure 1b) with a first minimum at a lag distance of 10 m and a sill at 0.25 since we assumed the pseudomorphs occurrence to be about 50 %. At first sight, the polygonal network was well reconstructed, but a closer look revealed that the algorithm could not make a distinction between the typical shapes for indicator 0 (host material) and 1 (wedge material): the map also shows white pseudomorphs with a black filling (Figure 2). With SNESIM prior knowledge about the spatial structure of the soil features can be translated into a TI. We designed a TI of a network of hexagons with a diameter of 10 m and a width of ca. 1 m, resulting in a pseudomorph occurrence of 52 % (Figure 1c). Although less detailed than the reference map, the classified E-type map reconstructs the shape of some dominant pseudomorphs reasonably well (Figure 2).

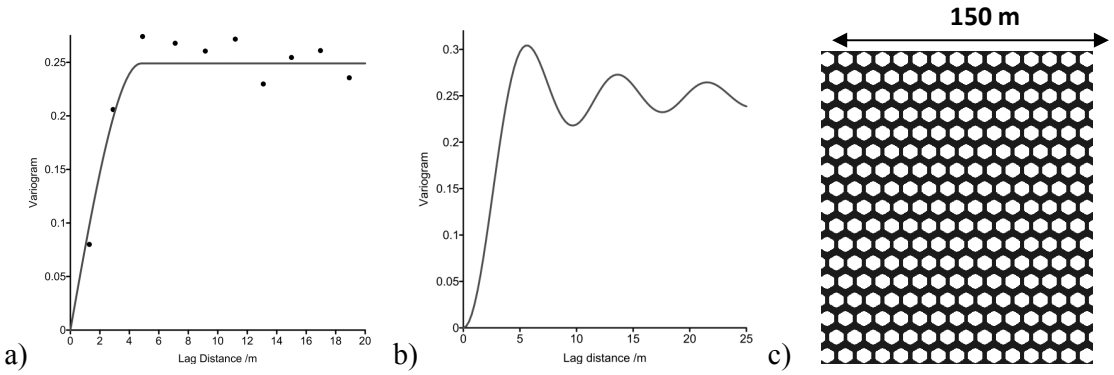
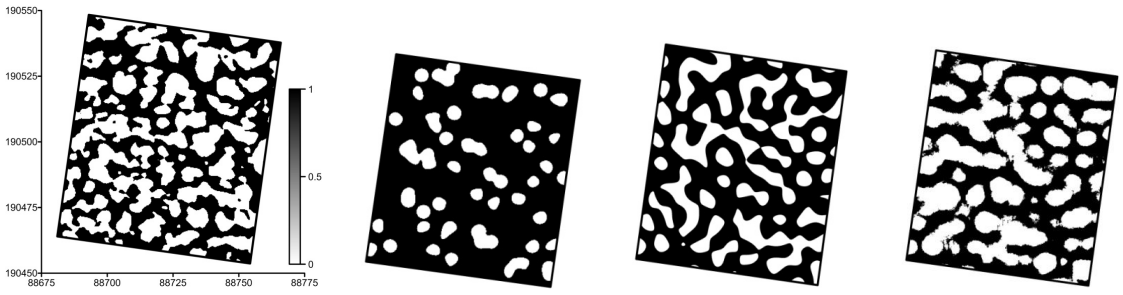


Figure 1: *(a) Data driven variogram model, (b) conceptual wave variogram model and (c) conceptual TI.*

The previous analysis was repeated for all the 13 data sets of the three data series. Both OK with a wave variogram model and SNESIM reconstructed the polygonal network. However, no matter the number of data points, the former continued to estimate black and white pseudomorphs.



Reference map | **OK estimation** data driven variogram model | **OK estimation** conceptual wave variogram model | **SNESIM estimation** conceptual TI

Figure 2: *Binary reference map and estimation maps of the pseudomorph occurrence interpolated from 168 indicator data.*

In soil science, local accuracy is also a requirement next to pattern reconstruction. To quantify the cell-by-cell agreement between the reference map and the estimated maps, we used the kappa statistic as implemented in the Map Comparison Kit software (Visser & Nijs, 2006) (Figure 3). As long as the data points did not reveal the hexagonal pattern, SNESIM gave the most accurate estimation. If the number of data became large enough to reveal the polygonal pattern, traditional OK with a data driven variogram model was the most accurate estimation method.

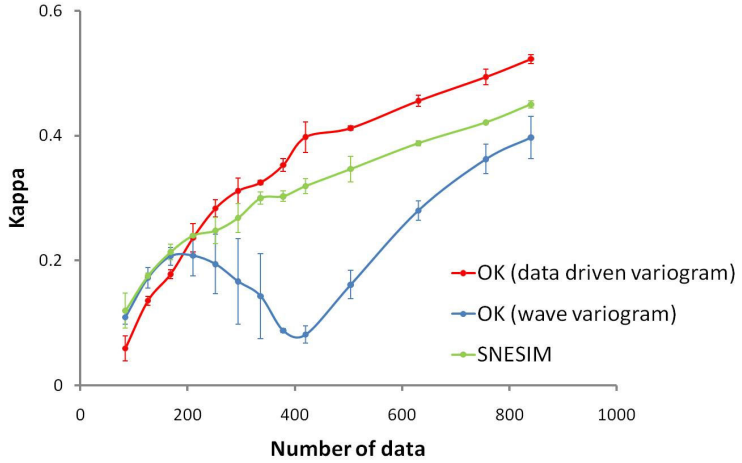


Figure 3: *Kappa statistic for the agreement between the reference and estimation maps in function of the number of data points for the three different estimation methods applied. The error flags represent the standard deviations.*

Conclusion

MPG can be a valuable contribution to soil science. Ice-wedge pseudomorphs are one example of repetitive soil features. Especially in situations where the number of data is too small to reveal the underlying pattern and where prior knowledge about the spatial structure is available, SNESIM is better able to reconstruct the pattern than two-point geostatistics. However, once the number of data becomes very large, traditional geostatistics performed best.

References

Cockx L, Ghysels G, Van Meirvenne M, Heyse I (2006) Prospecting frost-wedge pseudomorphs and their polygonal network using the electromagnetic induction sensor EM38DD. Permafrost and Periglacial Processes 17:163-168

Ewertowski M (2009) Ice-wedge pseudomorphs and frost-cracking structures in Weichselian sediments, Central-West Poland. Permafrost and Periglacial Processes 20:316-330

Ghysels G (2008) Bijdrage tot de studie van de kenmerken, de genese en de datering van periglaciale polygonale wigstructuren in België. Ph.D. thesis, Dep. of Geography, Ghent University, 266 p

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Mathematical Geology 34:1-21

Visser H, de Nijs T (2006) The Map Comparison Kit. Environmental Modeling & Software 21:346-358

SPATIAL CORRELATION BETWEEN WEEDS AND SOIL APPARENT ELECTRICAL CONDUCTIVITY MEASURED USING ELECTROMAGNETIC INDUCTION

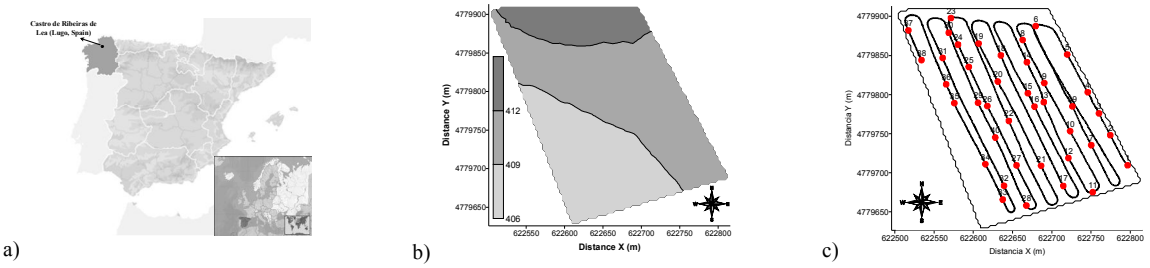
G.M. Siqueira[1], J.Dafonte Dafonte[1], J. Paz-Ferreiro[2]

1 Universidad de Santiago de Compostela, Escuela Politécnica Superior, Laboratorio de Hidráulica, 27002, Lugo, Spain. E-mail: jorge.dafonte@usc.es

2 Universidad de La Coruña, Facultad de Ciencias, A Zapateira, 15071, La Coruña, Spain. E-mail: jpaz@udc.es

Abstract

Weed control is an important issue of all production systems. Usually, herbicides are uniformly spread, based on a visual evaluation of the weed density, over the entire field aiming at weed control (Oliveira et al., 2008). On the other hand, there is no information about the weed population and its relation to soil properties. In this sense, it is possible to map in detail soil apparent electrical conductivity (EC_a) measured by electromagnetic induction. According to McNeill (1980) EC_a is influenced by soil water content, texture, organic matter, size and pore distribution, salinity, cationic exchange capacity, dissolved electrolytes concentration, soil temperature and colloid composition. The objective of this work was to determine the linear and spatial correlation between weed populations and soil apparent electrical conductivity (EC_a) measured with electromagnetic induction. The study area surface is a 6 ha plot located in Castro Ribeiras de Lea, Lugo, Spain (Fig. 1a). The geographical coordinates of the study area are: 43º 09' 49" N and 7º 29' 47" W, with an average elevation of 410 m.a.s.l. and an average slope of 2 % (Fig. 1b). The soil of the area is classified as Gleyc Cambisol.



a) b) c)

Figure 1. *Location map of the study area (a); digital elevation map of the study area (b); Sampling scheme of soil apparent electrical conductivity (EC_a) for 1886 points and weeds population at 40 points (c).*

The crop at the moment of measurement was maize for silage under no tillage system. In the last years the crop was permanent grassland for silage. Apparent soil electrical conductivity (EC_a, mS m^{-1}) was measured with an induction electromagnetic device EM38-DD (Geonics Limited). The equipment consists of two units of measurement, one in a horizontal dipole (EC_a-H) to provide and effective measurement at a depth of approximately 1.5 m and the other one in vertical dipole

(EC_a-V) with an effective measurement depth of approximately 0.75 m (McNeill, 1980). The data were collected on 23/06/2008 in 1859 points (Fig. 1c), using a field computer and a GPS RTK to georeference the electrical conductivity measurements. The software ESAP 2.35 (Lesch et al., 2000) was also used to determine the optimum position of the 40 sample points (Fig. 1c) for sampling the weeds populations, according with Lutman and Perry (1999).

Table 1. *Statistical parameters*.

Variable	Unit	N	Min	Max	Media	Variance	CV	Skew	Kurt	D
EC_a-V	mS m^{-1}	1886	4.13	20.13	11.21	6.12	22.07	0.485	-0.243	0.071Ln
EC_a-H		1886	6.63	20.00	12.12	3.22	14.81	0.839	1.285	0.092Ln
Dactylis glomerata	plants m^{-2}	30	5	51	24	229.00	64.01	0.503	-0.986	0.160n
Trifolium sp.		18	7	26	10	26.50	50.47	0.686	-0.124	0.246Ln
Bellis perennis		25	5	31	13	32.40	43.15	0.286	-0.107	0.191Ln
Plantago lanceolata		24	4	31	13	43.20	52.41	1.029	1.087	0.209Ln
Taraxacum officinale		22	5	20	10	20.10	43.29	0.749	0.209	0.290Ln
Achillea millefolium		17	1	15	9	18.80	48.61	0.534	-1.416	0.311n
Other species		27	2	36	13	73.00	67.47	1.737	4.393	0.218Ln
Weed density		40	26	77	52	185.00	26.00	0.308	-0..761	0.113n
Number of species	species m^{-2}	40	1	6	4	1.60	35.30	-0.403	-0.302	0.204Ln

N: number of measurements; Min: minimum value; Max: maximum value; CV: coefficient of variation (%); Skew: skewness; Kurt: kurtosis; D: Data normality for Kolmogorov-Smirnov test ($p < 0.01$, n: normal and Ln: lognormal).

The statistical test of Kolgomorov-Smirnov ($p<0.01$) (Table 1) shows that the majority of the parameters measured have a lognormal distribution, but *Dactylis glomerata* and weed density data have a normal distribution. Coefficient of variation values are medium (CV = 12.00-60.00 %) for all of the studied parameters except for *Dactylis glomerata* (CV = 64.01 %) and other species (67.47 %), being ECa-V e ECa-H the properties with lowest values of CV, 22.07 and 14.81% respectively.

Table 2. *Semivariogram parameters.*

Variable	Model	C_0	C_1	a (m)	SD (%)
Log EC_a-V Residual	Spherical	0.0001	3.14	105.00	0.00
Log EC_a-H Residual	Spherical	0.14	0.302	44.00	31.67
Dactylis glomerata	Spherical	1.90	12.00	210.00	13.67
Trifolium sp	Exponential	0.85	0.65	30.00	56.67
Bellis perennis	Spherical	0.85	1.65	50.00	34.00
Plantago lanceolata	Exponential	0.40	1.70	40.00	19.05
Taraxacum officinale	Exponential	0.30	0.75	40.00	28.57
Achillea millefolium	Spherical	0.30	0.40	80.00	42.86
Otras especies	Exponential	1.00	1.60	50.00	38.46
Weed density	Exponential	20.00	160.00	60.00	11.11
Number of species	Spherical	0.83	0.83	85.00	50.00

C_0: nugget effect; C_1: structural variance; a: range (m); SD: spatial dependence (%).

The values for Pearson´s correlation coefficient are not showed at this study because the correlations are weak or very weak, being the higher value of correlation calculated for EC_a-V x Others species ($|r|$=0.488), EC_a-H x *Bellis perennis* ($|r|$=-0.370) and EC_a-V x *Bellis perennis* ($|r|$=-0.352). The spherical model of semivariogram (Table 2) showed the best fitting for the majority of the studied parameters, although the exponential model was used for *Trifolium sp.*, *Plantago lanceolata*, *Taraxacum officinale*, Others species and weed density. The range value (a, Table 2) oscillates from 30.00 m (*Trifolium sp.*) to 210.00 m (*Dactylis glomerata*).

All the maps (Fig. 2) were obtained using ordinary kriging. Spatial variability shows a similar pattern between the EC_a-V and EC_a-H maps (Fig. 2). In these maps the higher values of EC_a are in the southwest side of the field. It is interesting to notice that only *Dactylis glomerata* and weeds density maps show the highest values in the same area where the values of EC_a were higher. On the other hand, the maps of *Bellis perennis*, *Plantago lanceolata*, *Taraxacum officinale* and other species show the lowest values in the areas with highest EC_a values. The topography of the field influences in the soil water flow, and soil water content is the main property which influences in the occurrence of the different weed species at this area. The use of EC_a data for the optimizing the weed sampling scheme was an efficient tool. Although the linear correlation values are low between EC_a and weed species, it is possible to identify a common spatial pattern in distribution of the different weed species and soil EC_a measured with electromagnetic induction.

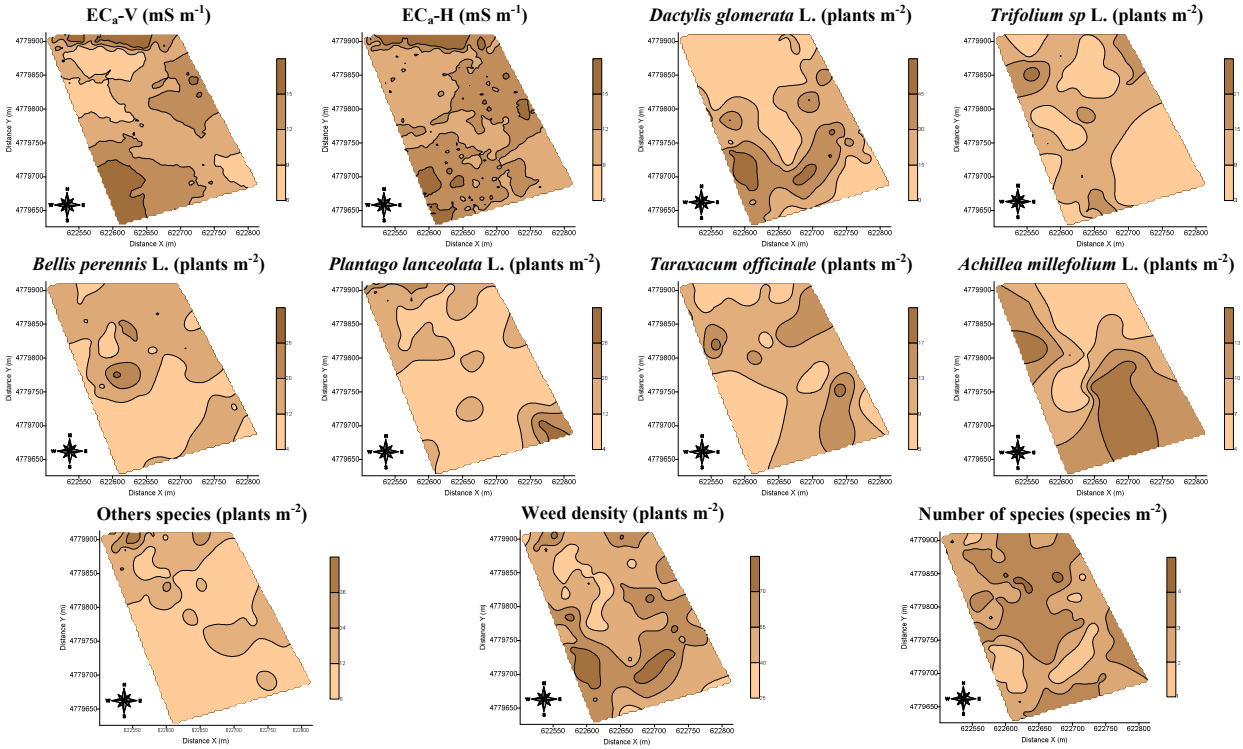


Figure 2. *Maps of spatial variability for the studied attributes*.

References

Lesch SM, Rhoades JD, Corwin DL (2000) ESAP Version 2.01R, user manual and tutorial guide.

Research Report No.146. George E. Brown Jr., Salinity Laboratory, Riverside, CA, 153pp.

Lutman PJW, Perry NH (1999) Methods of weed patch detection in cereal crops. In: The 1999 Brighton Conference – Weeds, Brighton 1999. Proceedings BCPC, 1999. P.627-634.

McNeill J.D (1980) Electrical conductivity of soils and rocks. Technical Note, TN-5, Geonics Ltda, Ontario, 22p.

Oliveira JP, Batista LF, Siqueira GM (2008) Variabilidade espacial de plantas daninhas num latossolo vermelho sob sistema plantio direto. Revista Brasileira de Engenharia de Biossistemas 2: 273-281.

THE IMPROVEMENT OF PEAT DEPTH MODELS FOR NORTHERN IRELAND THROUGH THE INVESTIGATION OF TELLUS DATA

M. Robinson[1], J. McKinley[1], A. Ruffell[1], M. Young[2]

[1] School of Geography, Archaeology and Palaeoecology, Queen's University, Elmwood Building, BT7 1NN, Belfast, N. Ireland, mrobinson34@qub.ac.uk, +44 (0)28 9097 3312

[2] Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Malone Lower, BT9 5BF, Belfast, N. Ireland, gsni@detini.gov.uk, +44 (0)28 9038 8462, +44 (0)28 9038 8461

Introduction

The government of the United Kingdom (UK), under the UN Framework Convention on Climate Change, is committed to the implementation of policies that protect and enhance reservoirs of greenhouse gases. CO_2 is considered the main anthropogenous greenhouse gas, with its atmospheric concentration being limited by the sequestration of Carbon (C) in vegetation and soils.

The high proportion of soil carbon within peat (42% of soil C in NI (Tomlinson and Milne, 2006)) is explained by the relatively high carbon density of peat and peaty soils. Soil C density, the mass of organic C per unit area, is calculated using %C, depth and bulk density values.

This research investigates the use of Tellus gamma radiation data to produce peat depth models for Northern Ireland and improve estimates of carbon stocks. The gamma spectrometer system utilised in the Tellus airborne geophysical survey collected information on naturally-occurring potassium (K), uranium (eU) and thorium (eTh) in the ground surface. Gamma-radiation from rocks is attenuated by overlying peat, with previous tests indicating some correlation between gamma-radiation and peat depths to at least 2m. Current work is assessing the relationships between the Tellus data and previous peat depth measurements using geostatistical techniques within GIS. Hyvonen *et al* (2005) reports the successful use of gamma radiation for limited depth mapping in Finland while unpublished preliminary tests in Northern Ireland have also been positive. An improved peat depth model, produced through the integration of the Tellus data and known peat depths, could improve carbon stock calculations.

Methods

Before geostatistical analysis could be utilised, peat depths from historical sources (1956) and recent surveys (1997/98) had to be incorporated into GIS through georeferencing and digitisation. This comprehensive dataset allows analytical techniques to be applied to every site simultaneously.

Results

Case Study: Armoy Peat Bog

Analysis will now focus on the pilot study site of Armoy peat bog, situated approximately 0.5 miles north of Armoy village, Co. Antrim. The bog exhibits a typical raised-moss stage which has developed within a local depression (Double, 1956).

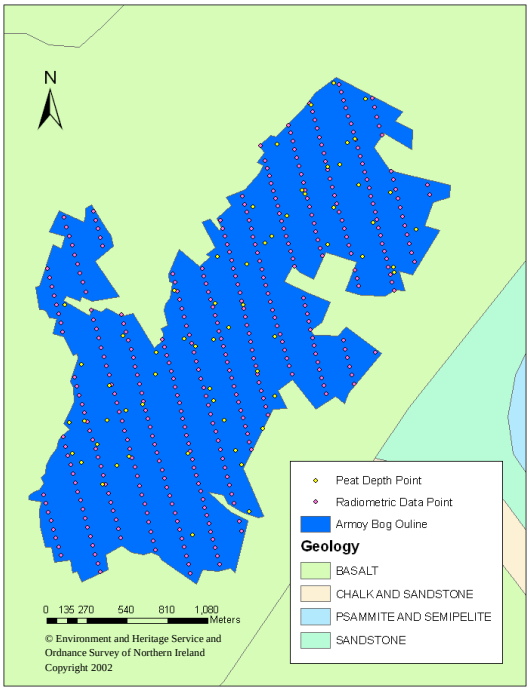


Fig.1: Map displaying annotated image of Armoy peat bog

The bog outline, as presented in the source material, was firstly digitised to restrict spatial interpolation to inside the area of study. Fig.1 displays the boundary of the Armoy site, with the peat depth sample locations and radiometric data points being clearly illustrated.

The image indicates that the entire site is underlain by Basalt. The homogeneity of the underlying geology was an important factor when determining a suitable location for the preliminary study as the level of gamma radiation emitted from bedrock varies with geology type. Recent research shows that while an attenuated count rate of 300 cps from Dalradian Shists indicates a peat depth of 2m, a count of 300 cps from underlying Basalt, as in Fig.1, suggests a peat depth of approximately 0.5m. Gamma-ray levels from a site composed of a complex underlying geology will be interpreted in relation to peat depth in future research.

Spherical semi-variogram models were produced using the radiometric total count data (Fig.2) and peat depth measurements (Fig.3). Fig.2 was composed of two models to generate a more accurate representation of the spatial variation within the dataset.

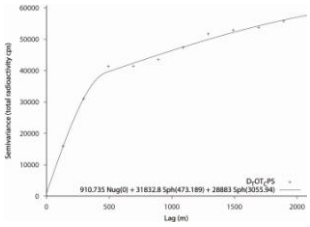


Fig.2: Spherical semi-variogram model of gamma radiation total cps values

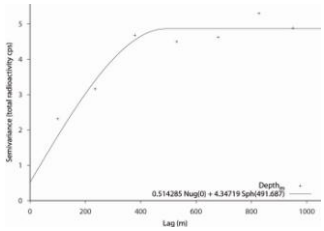


Fig.3: Spherical semi-variogram model of peat depth values

Values obtained from the model equations, illustrated in Fig.2 and Fig.3, were incorporated into GIS to produce outputs using Ordinary Kriging (OK). Fig.4 displays interpolated radiometric total count values using OK while Fig.5 shows interpolated peat depths generated using the same technique.

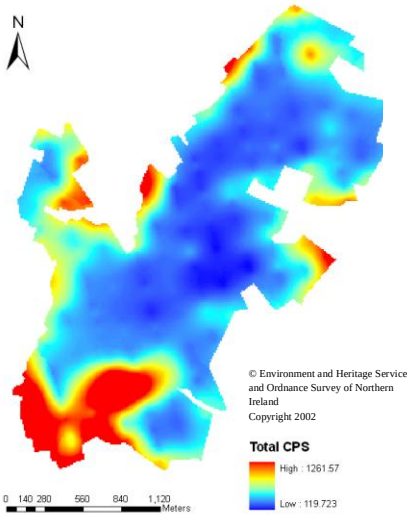


Fig.4: Map displaying interpolated total count values (OK) for Armoy

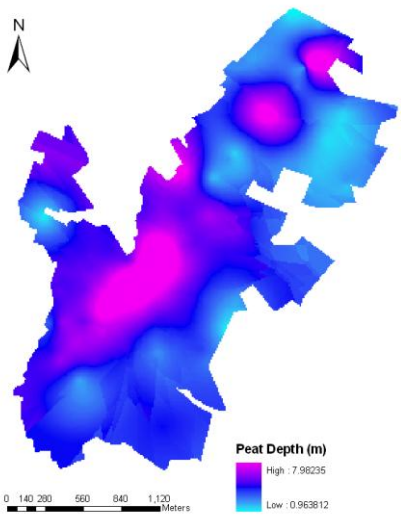


Fig.5: Map displaying interpolated peat depth values (OK) for Armoy

The interpolated images indicate a general correlation between gamma radiation and peat depth, with low total cps values being associated with high peat depths. Deep areas of peat located in the centre of the bog appear to correspond with low total count figures, however the radiation output suggests the areas of highest peat depth should be located further east. Several anomalies can be identified when comparing the outputs, including high radiation values in the south west and a point in the north containing both high peat depths and high total count values. These irregularities may be due to the historical nature of the depth data. Future work will involve crosschecking radiation levels using portable gamma-ray spectroscopy, as well as identifying temporal changes in peat cover through the analysis of orthographic imagery and field observations.

Conclusion

Northern Ireland, with considerable volumes of existing depth data, is an ideal location for establishing, through the employment of geostatistics, the relationship between radiation attenuation and peat depth. These results will significantly improve our national estimates of existing carbon-in-storage.

References

Double KWW (1956) The Northern Ireland Peat Bog Survey. Council of Scientific Research and Development. Belfast: Her Majesty's Stationery Office.

Hyvonen E, Turunen P, Vanhanen E, Arkimaa H, Sutinen R (2005) Airborne Gamma-ray Surveys in Finland. Geological Survey of Finland, Special Paper 39: 119-134.

Tomlinson RW, Milne RM (2006) Soil carbon stocks and land cover in Northern Ireland from 1939 to 2000. Applied Geography 26(1): 18-39.

SCREENING THE FLEMISH (BELGIAN) SOIL DRAINAGE CLASS MAP FOR ITS CURRENCY

J. Van de Wauw, P.A. Finke

Department of Geology and Soil Science, Universiteit Gent, Krijgslaan 281/S8, B-9000 Gent, Belgium

Abstract

The currency of the Flanders drainage class map was evaluated using data from two monitoring networks: one with good spatial coverage but poor temporal coverage and another with better temporal but poor spatial coverage. We combine both networks to obtain point expressions for mean highest (MHW) and mean lowest water tables (MLW) by applying time series analysis and total least squares regression. The resulting MHW and MLW point data set was used to evaluate the currency of the existing map and to identify regional differences.

Introduction

Phreatic groundwater dynamics are one of the most important land characteristics for agriculture, nature development and other land uses. In Flanders these dynamics are usually estimated from the natural drainage classes that are indicated on the Belgian soil map (1/20.000), based on data collected during the national soil survey (1947-1971). The natural drainage condition on the soil maps was derived from the depth of gley mottles and a reduction horizon and their position in the landscape. They are indicated using combined classes of the depth of reduction and the depth of mottling. However, these morphogenetic features do not always reflect recent changes in the hydrology and their expression also strongly depend on other soil properties like pH, parent material and organic carbon content.

Even though the original definition was morphological, a common interpretation (eg Van Damme 1969, Boucneau 1996) of these drainage classes is a set of mean highest and mean lowest groundwater levels. The mean highest water level (MHL) and mean lowest water level (MLW) are defined as the mean value of the three highest and lowest groundwater levels measured biweekly for at least 8 years, preferably longer (30 years) for climate representativeness (Van der Sluijs and De Gruijter, 1985).

To check the currency of the drainage class maps, mapped classes need to be compared to recent point estimates of MHW/MLW. Such estimates can be obtained from the recently established phreatic groundwater monitoring network. This network has a good spatial coverage (average density of 1/340 ha), but a poor temporal coverage since the monitoring started only in 2004-2006 and only 2 observations are made per year. This means it has to be combined with data from monitoring networks with a better temporal coverage to derive MHW/MLW statistics.

Reference time series

Two monitoring networks are used as a reference series for calculating the groundwater statistics: the shallow filters (<5m) of the primary network of Flemish Environmental Agency (VMM) and the monitoring networks installed in nature reserves. A selection was made: all locations with at least 2 years of measurements and 24 observations are included, further visual selection was made to exclude series with trends. A total number of 150 reference times series could be retained, which have a good coverage of both the drainage classes and the different considered regions.

Time Series Analysis

The monitoring period in some reference series was too short for the direct calculation of groundwater statistics. Moreover the monitoring periods have different lengths. This can lead to biased estimations since precipitation surplus is a major driving factor in groundwater fluctuations in western European climate (Knotters and Bierkens, 2000), and precipitation surplus varies by year. Therefore a time series model was fitted to the phreatic and precipitation surplus data and used to estimate the MHW and MLW over a fixed climate-representative period of 30 years (1978-2008). The time series model used is PIRFICT (von Asmuth et al., 2002). This is an impulse- response model that uses precipitation surplus with a stochastic component.
A different objective for using a time series model is that data can be interpolated for days without measurements. This is needed when the data of the reference series is combined with the measurements in the phreatic groundwater monitoring network.

Phreatic groundwater monitoring network

The phreatic groundwater monitoring network (figure 1), implemented and managed by the Flemish Environmental Agency (VMM) is built as an implementation of the European Nitrate Directive (91/676/EEG). The network contains 2100 monitoring points, or a density of 1/340 ha. These points are measured 2 times per year, generally in spring and autumn. Monitoring piezometers were located more or less equally over Flanders, but with a higher density in regions more vulnerable to nitrate leaching (Eppinger and Thomas, 2007).

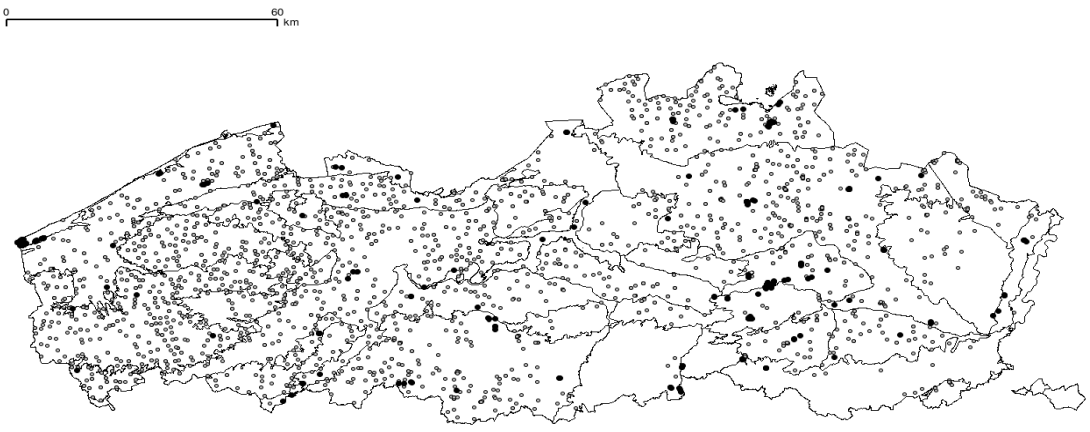


Figure 1 *Location of the long reference series (black) and the phreatic groundwater monitoring network (gray)*

Combining the reference time series and the phreatic network data

To estimate ground water statistics in case only a small number of observations is available, two different methods have been proposed. Both methods use linear regression to create a relationship between observations in the reference piezometers and the observations in the short time series.

The first method (te Riele and Brus, 1991; Finke et al., 2004) was developed for estimating the mean highest and lowest water table using only two well timed observations in the short time series, in winter and summer, when the groundwater table is close to the MHW and MLW respectively. A regression relation is derived between the MHW/MLW and the observation at date X in the reference series, and applied to the measurement at the same date in the short series.

The second method (Oude Voshaar and Stolp, 1997) can be used when a larger number of observations is available. In this case linear regression is used for to fit a relationship between the observations in the short time series and the observations in each of the reference series. The best relationship (i.e. with the best fitting reference series) is than retained and used for prediction.

Total Weighted Regression

Since in many cases the reference series were not measured on the same day as the short time series, both methods had to be adjusted: they rely on a relationship between the water level measured in the reference series on a specific date and the groundwater statistics derived in these series. If no observation is available a simulated water table on that specific date had to be used. This introduces uncertainty in the explanatory variable in both methods, which excludes the standard least-squares algorithms. Instead a weighted total least square method (Markovsky and Van Huffel, 2007) is used, which incorporates (and weighs for) the uncertainty in the explanatory variable. Compared to ordinary least squares, this method suffers less from

regression towards the mean (thus covering a larger range of drainage classes), and gives less weight to the less certain predictions, also influencing the confidence levels. The resulting MHW and MLW point data set is not dense enough for mapping, but allows the identification of regional differences.

References

Boucneau G, Van Meirvenne M, Desmet J, Hofman G (1996) Regressive modelling of soil water table fluctuations. Mathematics and Computers in Simulation 42(2-3): 271-277.

Eppinger R, Thomas P (2007) Hydrological homogeneous zones to derive the nitrate vulnerability of ground water. (in Dutch) Tijdschrift Water 28, no. 7: 1-6.

Finke PA, Brus DJ, Bierkens MFP, Hoogland T, Knotters M, de Vries F (2004) Mapping groundwater dynamics using multiple sources of exhaustive high resolution data. Geoderma 123(1-2): 23-39..

Knotters M, Bierkens MFP (2000). Physical Basis of Time Series Models for Water Table Depths. Water Resour. Res., 36(1), 181–188.

Markovsky I, Van Huffel S (2007) Overview of total least-squares methods. Signal Processing 87(10): 2283-2302.

Oude Voshaar J, Stolp J (1997) Estimating MHW and MLW from temporary piezometers with short time series (in Dutch). Wageningen, SC-DLO, Techn. Doc. 30, 38pp.

Vandamme J, De Leenheer L(1969) Fluctuations of the ground water tabel over a period of 5 years (1963-68) in the sandy soils of the Campine in the province of Antwerp (Belgium) (in French). Pedologie 19(3): 275-320.

Van der Sluijs P, De Gruijter JJ (1985) Water table classes: A method to describe seasonal fluctuation and duration of water tables on Dutch soil maps. Agricultural Water Management 10(2): 109-125.

von Asmuth JR, Bierkens MFP, Maas K (2002) Transfer function-noise modeling in continuous time using predefined impulse response functions. Water Resour. Res., 38(12): 1287.

AREA-TO-POINT KRIGING OF SOIL TEXTURE: AN EFFICIENT USE OF LEGACY SOIL DATA FROM POLYGON MAPS

R. Kerry[1], B. Rawlins[2], P. Goovaerts[3]

[1]Brigham Young University, Provo, UT, USA. Email: ruth_kerry@byu.edu
[2]British Geological Survey, Keyworth, UK. Email: bgr@bgs.ac.uk
[3]Biomedware Inc., Ann Arbor, MI, USA. Email: goovaerts.pierre@gmail.com

Polygon maps representing soil classes with associated typical values of key soil properties (legacy data) exist in many locations. Effective methods are required for their incorporation into digital soil mapping (DSM) at national and regional scales so that additional sampling effort can be minimized and so that this useful information is not overlooked.

Here we outline an approach to incorporating soil organic carbon (OC) values from polygon maps into a digital soil mapping procedure. Rawlins et al. (2009) demonstrated that airborne radiometric and altitude data (ancillary data) could improve soil organic carbon (OC) estimates for the whole of Northern Ireland within a linear mixed modeling framework. The models used by Rawlins et al. (2009) are summarized in Table 1. Here we use the same data as Rawlins et al. (2009) from the Tellus survey of Northern Ireland (6862 soil samples) to investigate whether a 1:250,000 scale map of soil polygons with associated typical OC values (AFBINI) could further improve estimation of OC. We use the raw OC values associated with each polygon and Area-to-Point (AtoP) kriged (Kyriakidis, 2004) OC values as additional fixed effects in the six models used by Rawlins et al. (2009). The AtoP approach involves kriging from irregularly shaped polygons to points. Unlike a centroid-based approach, the irregular shape and area of the polygons is taken into account during variogram deconvolution and kriging. Figure 1 shows maps of OC measured in samples from the TELLUS survey, the polygon map and AtoP kriging, the altitude and radiometric K data are also shown along with their correlations with TELLUS Log OC.

Table 1.

Fixed effects for Rawlins et al. (2009) models

Model	Fixed effects for Rawlins et al. (2009)
1	Constant (ie the mean)
2	Constant and altitude
3	Constant and K
4	Constant, K and K^2
5	Constant, altitude and K
6	Constant, altitude, K and K^2

Table 2 shows that when considering all soil types the mean absolute errors were smaller than those obtained by Rawlins et al. (2009) for all six models when the raw polygon OC was incorporated into regression kriging, and smaller still when AtoP kriged OC was used as the extra fixed effect. This was also the case for the peat soils. For organo-mineral and mineral soils

the results were less consistent, but using the information from the polygon maps provided improvement on Rawlins et al.'s (2009) models in some cases.

A jack-knife procedure performed for all soils (results not shown) confirmed that using raw OC in models 1-6 reduced MAEs and using AtoP OC reduced these further. A jackknife for Model 6 considering all soils and different soil types showed that 300 soil sample points was optimal to map OC for the whole of Northern Ireland (Figure 2). It also showed that using AtoP OC as an extra fixed effect reduced MAEs for all soil types except for some small sample sizes for mineral soils.

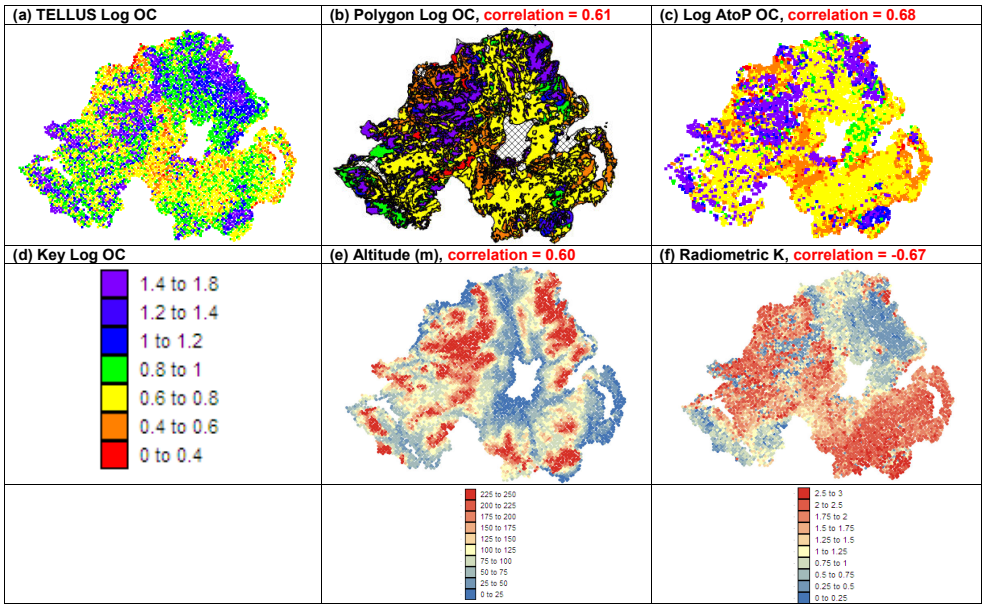


Figure 1. *Maps of TELLUS Log OC and various covariates*.

Table 2. *Mean absolute errors (MAE) from regression kriging of soil organic carbon using Models 1-6 outlined in Table 1 (Rawlins) and Models 1-6 plus raw polygon OC (raw OC) and Models 1-6 plus Area to Point kriged OC (AtoP OC).*

		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
All	Rawlins	6.46	6.08	4.43	3.86	4.53	4.02
	Raw OC	6.13	5.94	4.32	3.82	4.43	3.96
	AtoP OC	5.60	5.37	4.11	3.66	4.18	3.77
Mineral	Rawlins	1.93	1.85	1.62	1.57	1.61	1.57
	Raw OC	1.88	1.82	1.62	1.56	1.60	1.56
	AtoP OC	1.81	1.75	1.62	1.58	1.59	1.57
OM	Rawlins	3.32	3.40	2.66	2.64	2.86	2.69
	Raw OC	3.35	3.35	2.72	2.65	2.82	2.68
	AtoP OC	3.43	3.27	2.86	2.81	2.78	2.70
Peat	Rawlins	4.35	4.20	3.58	3.15	3.62	3.30
	Raw OC	4.21	4.17	3.46	3.11	3.56	3.25
	AtoP OC	3.89	3.90	3.19	2.82	3.36	3.04

We the extend approach used with OC to mapping of soil texture. AtoP kriged soil texture data are used with geochemical data as covariates in regression kriging. The percentages of sand, silt

and clay – important soil properties in DSM - are compositional data; we apply the additive log ratio transformation (alr) so that these data can be treated as unbounded random variables. The alr transforms of the typical soil texture values for each soil association in the 1:250,000 scale soil polygon map are used in AtoP kriging. Values are back transformed to give the percentages of sand, silt and clay and textural class for cross-validation at 800 sites.

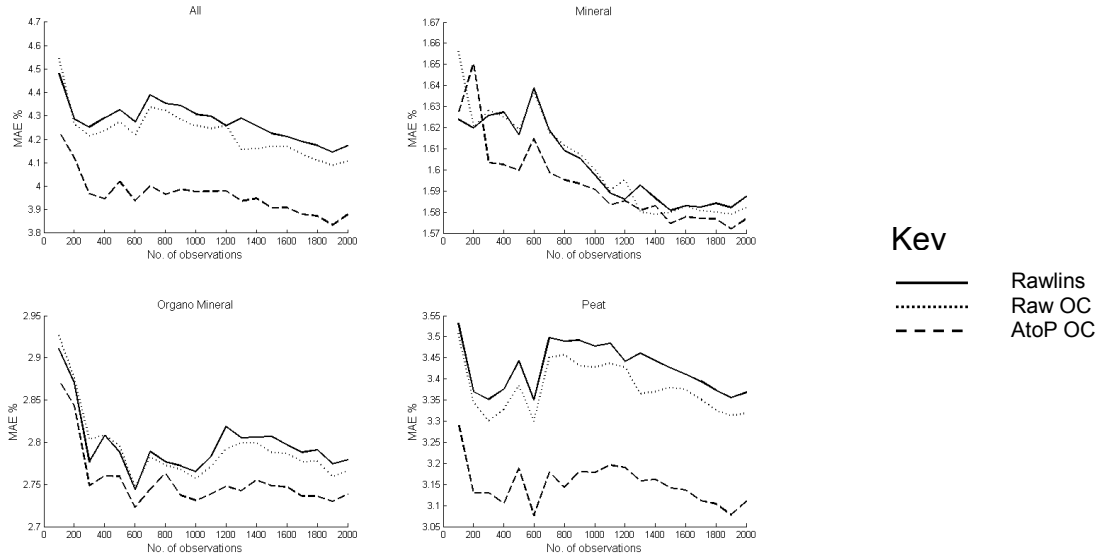


Figure 3. *Jack-knife results for all, mineral, organo-mineral and peat soils using model 6 outlined in Table and model 6 plus (raw OC) and plus Area to Point kriged OC (AtoP OC).*

We also assess the performance regression kriging with geochemical data and incorporating the AtoP kriged data as an extra fixed effect. This approach is promising when using the geochemical data and parent material codes. However, the geochemical data are expensive to collect and are not available as ancillary data in many locations. The radiometric and topography data collected as part of the Tellus survey were effective covariates for OC but not for soil texture. Other cheaper covariates need to be found to effectively map soil texture, but AtoP kriged values of texture provide useful approximations at the soil texture which should not be discarded. Electrical conductivity (EC_a) measurements, Aerial imagery or derived topographic attributes have proved useful as covariates for mapping within field variations in soil texture in precision farming. Perhaps these could also be useful inexpensive covariates for mapping soil texture. These possibilities need to be further investigated.

References

Kyriakidis P (2004) A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis, 36:259–289

Rawlins B G, Marchant B P, Smyth D, Scheib C, Lark R M, Jordan C (2009) Airbourne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. European Journal of Soil Science. 60:44-54

Acknowledgements

We thank Dr Ben Marchant for completing the jack-knife procedure for this work. We thank Michael Young of DETINI for arranging access to the Tellus data. The Tellus project was funded by the Department of Enterprise, Trade and Investment and by the Building Sustainable Prosperity Scheme of the Rural Development Programme (Department of Agriculture and Rural Development of Northern Ireland).

MULTIPLE KERNEL MODELS FOR RESERVOIR CHARACTERIZATION

L. Foresti[1], V. Demyanov[2], M. Christie[2], M. Kanevski[1]

[1] Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland
Loris.Foresti@unil.ch, Tel: +41 (0)21 692 3536, Fax: +41 (0)21 692 3535
[2] Institute of Petroleum Engineering, Heriot-Watt University, Riccarton, UK
Vasily.Demyanov@pet.hw.ac.uk, Tel: +44 (0)131 451 8298, Fax: +44 (0)131 451 3127

Introduction

The present study explores the use of multiple kernel learning (MKL) for spatial prediction of porosity and permeability fields in a subsurface reservoir from seismic information. Spatial distribution of these petrophysical properties is an essential component of a reservoir model used in oil production forecasting. The oil production model prediction is matched to the available measured production history, which is an inverse problem. Thus, there may be several porosity and permeability fields which would match the historical data well. Uncertainty about oil production forecasts can be estimated by running flow simulations with a set of possible realistic porosity and permeability fields. Several approaches can be adopted to build a set of realizations of such fields using 2-point geostatistics, multi-point geostatistics, object-based models, and more recently – data-driven models (Kanevski et al. 2002; Demyanov et al. 2008). In this study we exploit the flexibility of multiple kernel models to provide realizations of porosity and permeability fields. By wrapping a Support Vector Regression model MKL provides more interpretable solutions in terms of the relevance of contributing seismic inputs. MKL power in integrating different information sources is tested by extracting geological features from noisy seismic images. The flexibility of the approach is very adapted for modelling under uncertainty.

Methods

Support Vector Regression (SVR) is a kernel method for non-linear regression estimation (Vapnik 1995). Particularly adapted for high-dimensional data, SVR minimizes a robust cost function comprising empirical risk and model's complexity. SVR prediction is basically given by a weighted expansion of kernel functions

$$f(\vec{x}) = \sum_{i=1}^{N} (\alpha_i - \beta_i) \cdot K(\vec{x}, \vec{x}_i) + b$$

where the coefficients α and β are found by quadratic programming and b is the bias. Kernel functions provide an implicit mapping of the data to a higher dimensional space (feature space) where linear regression is possible, thus finding the correct mapping of the data is crucial and affects model's performance. Basis kernels such as the Gaussian radial basis function may encode a rigid representation of the data. A more flexible and interpretable kernel can be built using a convex combination of m basis kernels

$$K(\vec{x},\vec{x}') = \sum_{m=1}^{M} d_m K_m(\vec{x},\vec{x}') \qquad \text{with} \qquad d_m \geq 0 \qquad \text{and} \qquad \sum_{m=1}^{M} d_m = 1$$

The optimal weighted combination is found by SimpleMKL (Rakotomamonjy et al. 2008) which is an efficient gradient-based algorithm for the joint optimization of the kernel weights d_m's and of the SVR coefficients α and β.

Data

A 2D section of the synthetic Stanford VI dataset (Castro et al. 2005) composed of meandering channels with high porosity and permeability was selected to demonstrate the applicability of the approach. The layer is perforated by 45 wells which form the training set. The multiple kernel is built using one kernel for XY coordinates and one kernel for the seismic attribute. In order to have a more realistic situation seismic data were perturbed with noise. Based on the perturbed field we applied a family of convolution filters to describe geological structures at different scales. The computed *geo-features* are stacked to the original noisy seismic and dedicated to a kernel.

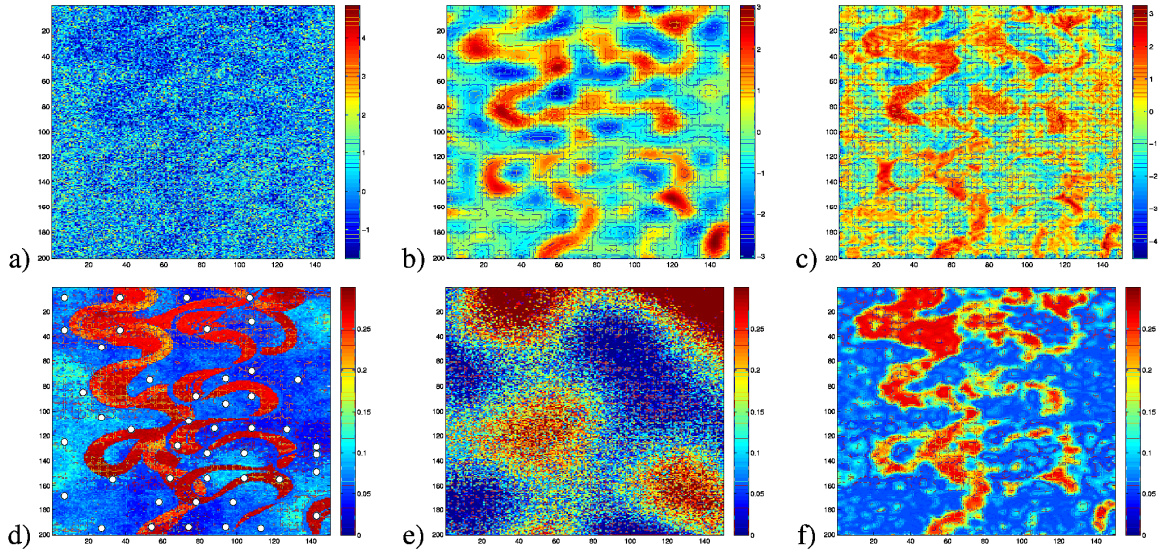


Figure 1*. a) Raw noisy seismic; b,c) geo-features derived from a); d) target porosity; e) predicted porosity from XY+noisy seismic; f) predicted porosity from XY+noisy seismic+geo-features; wells are shown with circles*

Experimental Results

Tab. 1 demonstrates that XY coordinates and the most noisy seismic input only (Fig. 1a) do not give enough information to capture and to model the channel structure (cross-validation error is very high). MKL learns from the data that the noisy input does not help to model the porosity field. Therefore, higher weights are given to the XY kernel. The effect is visible in Fig. 1e where porosity patterns are mainly a function of XY. However, this model was very unstable and did not find acceptable results. Advanced models integrating the extended group of geological features describing channel structure have much lower errors, higher correlations (Tab. 1) and more realistic patterns (Fig. 1f) which are close to the target porosity field (Fig. 1d). In such case,

the influence of XY and the noisy seismic in the models is irrelevant and most of the weights are given to the *geo-features* (see Tab. 1).

Models	cvRMSE (st.dev.)	cvCorr (st.dev.)	Kernel weights	
$SVR_{xy+seismic}$	1.030 (0.008)	-0.041 (0.085)	d_m of K_{xy}	d_m of $K_{seismic}$
$MKL_{xy+seismic}$	1.026 (0.025)	-0.218 (0.221)	0.816 (0.393)	0.185 (0.393)
$SVR_{xy+seismic+features}$	0.674 (0.013)	0.752 (0.011)	Σd_m of (K_{xy},$K_{seismic}$)	Σd_m of $K_{features}$
$MKL_{xy+seismic+features}$	0.645 (0.017)	0.779 (0.017)	0.031 (0.038)	0.969 (0.038)

Table 1. Statistics for 50 best fitting models (mean and standard deviation of errors)

Porosity model of Fig. 1f is only one among many best fitting models, i.e. models with low cross-validation errors. Depending on MKL parameters, the weights are distributed differently over the input features, which results in different possible porosity patterns. At the same time, matching these porosity fields to historical oil production data and not only to the porosity measurements from the wells is necessary to quantify uncertainties in prediction of reservoir performance.

Conclusion

MKL has demonstrated to be a powerful method for regression estimation and was easily applicable for porosity mapping in complex geological environments. Compared to the standard SVR model, MKL provides more flexibility in the kernel representation and allows to understand how different inputs/features contribute to the solution. In addition, there is a need for processing seismic data to extract geologically plausible features, especially when dealing with noisy seismic data. Flexibility of MKL eases its adaptation to uncertain quantification. Actually, the posterior probability of MKL parameters according to scales of different inputs/features can be evaluated by matching the history of oil production.

The research is partly supported by the Swiss NSF project N^o 200020-121835/1, and by the industrial sponsors of the Heriot-Watt Uncertainty Quantification JIP, Phase III.

References

Demyanov V, Pozdnoukhov A, Kanevski M, Christie M (2008) Geomodelling of a fluvial system with semi-supervised support vector regression. Proceedings of the VII International Geostatistics Congress GEOSTATS'2008, pp. 627-636

Kanevski M, Pozdnoukhov A, Canu S, Maignan M, Wong PM, Shibli SAR (2002) Support vector machine for classification and mapping of reservoir data. In: Wong P et al. (ed) Soft Computing for Reservoir Characterization and Modelling, Springer, pp. 531-558

Rakotomamonjy A, Bach FR, Canu S, Grandvalet Y (2008) Simple MKL. Journal of Machine Learning Research 9:2491-2521

Vapnik V (1995) The Nature of Statistical Learning Theory. Springer

Castro SA, Caers J, Mukerji T (2005) The Stanford VI reservoir. 18th Annual Report, Stanford Center for Reservoir Forecasting, Stanford University

TRACKING CO_2 PLUME MIGRATION DURING GEOLOGIC SEQUESTRATION USING A PROBABILISTIC HISTORY MATCHING APPROACH

S. Bhowmik[1], C.A. Mantilla[1] and S. Srinivasan[1]

[1] University of Texas at Austin, 1 University Station C0300, Austin, TX 78712, USA, phone: +1 512 788 1256, email: ssriniva@mail.utexas.edu

Introduction

Geologic carbon storage seeks to minimize the impact of CO_2 emissions by sequestering CO_2 in deep saline aquifers. Large-scale implementation of geologic carbon storage (GCS) will require monitoring on a project-by-project basis. Apart from purely regulatory concerns, operators are likely to avail themselves of any opportunity to track the CO2 plume so that they can better manage the storage operation. However, the large cost of most monitoring technologies, such as time-lapse seismic, limits their application. In this paper we investigate the application of probabilistic history matching using routine measurements of injection data as an inexpensive alternative to track the migration of CO_2 plume in an aquifer.

Probabilistic History Matching

In the petroleum industry, history matching has been widely used to calibrate prior geologic models using production data for more accurate predictions of field performance. Since it is an ill-posed inverse problem with inherent non-unique solutions, it is important to address history matching in a probabilistic framework in order to characterize the remaining uncertainty in the final set of models. In Pro-HMS- a software for assisted history matching developed at the University of Texas at Austin (Srinivasan and Bryant, 2004), the main objective is to assess the conditional probability $P\{A \mid B,C\}$ of a reservoir attribute (A) given the prior geological data (B) and the dynamic field data (C) by using the permanence of ratio hypothesis proposed by Journel (2002).

$$P\{A \mid B,C\} = \frac{a}{a+bc} \quad \text{where } a = \frac{1-P(A)}{P(A)}, b = \frac{1-P(A \mid B)}{P(A \mid B)} \text{ and } c = \frac{1-P(A \mid C)}{P(A \mid C)}.$$

The conditional probability P{A|B} is obtained from sequential indicator simulation, which ensures consistency of the merged models with prior geological information by honoring the conditional data and variogram model(s). The conditional probability P{A|C}is obtained by the gradual deformation method proposed by Caers (1999). The prior probability of permeability at any location in the reservoir P{A} is iteratively perturbed using a deformation parameter to yield the probability P{A|C}. The criterion for perturbing the deformation parameter is the best match to the available dynamic data. Subsequent to merging the two conditional distributions, new geostatistical realizations sampled from that posterior probability distribution reflect better the observed dynamic data while honoring the prior

geological information. The residual uncertainty in the history matched models and the extent of the CO_2 plume can be assessed through multiple realizations.

A synthetic example

First, the approach is demonstrated through a synthetic example where the flow path of CO_2 is deviated by large-scale permeability barriers. The objective was to infer the presence of those barriers by using injection well pressure to calibrate the initial models, and thereby to detect the deviation of the CO_2 plume from the initially predicted path. Figure 1 shows the reduction of the error in the estimated CO_2 saturation from the prior to the posterior set of realizations. In a second synthetic example, high permeability streaks diverted the CO_2 plume beyond the anticipated region. The presence of the streaks was captured by the final set of models. The results from these synthetic examples called the attention of the operators of In Salah, the largest on-shore CO_2 sequestration project, for the implementation of Pro-HMS to calibrate the available prior geologic model.

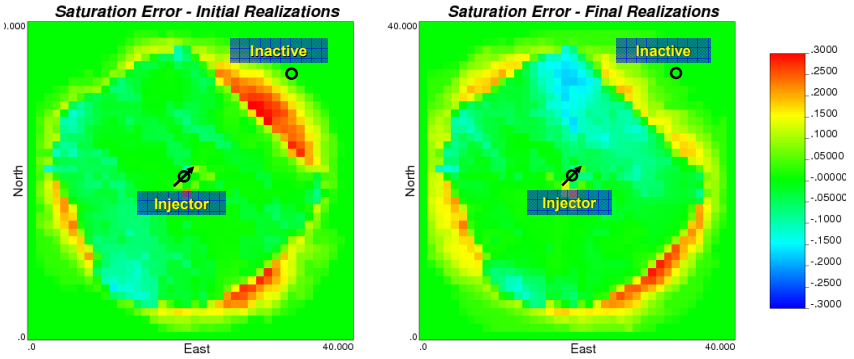


Figure 1: *Comparison of the mismatch in the predicted CO_2 saturation from the prior set of realizations to the posterior set of realizations in the synthetic example 1.*

A field example

The Krechba gas field is part of the In Salah project in Central Algeria. The structure is an anticline with natural gas at the top and saline aquifer down-dip. The injected CO_2 from one of the injectors showed a preferential migration which was contrary to the expected direction of migration and arrived at a suspended appraisal well much faster than was expected. This was attributed to some high permeability fracture pathways enabling the rapid migration of the CO_2 plume. Using the Pro-HMS algorithm in its original form failed to reproduce any high permeability features in the final model, since matrix permeability influences the plume migration differently from the directional fracture permeability. Hence, we applied a two-step approach to the history matching problem, creating a final model of only matrix permeability in the first step and overlaying this with high-permeability features in the second step, both steps being conditioned just to the injection well pressures. The final model successfully predicted the presence of the hypothesized heterogeneity features. The real application of this process however would lie in its ability to predict the preferential direction of migration

before there is any evidence for the same. So, we tested the method using injection data only till a time before the plume arrived at the appraisal well. We found that the final model created was still able to predict the presence of the high permeability feature between the injector and the appraisal well (as shown in Figure 2), and reproduce it over multiple realizations.

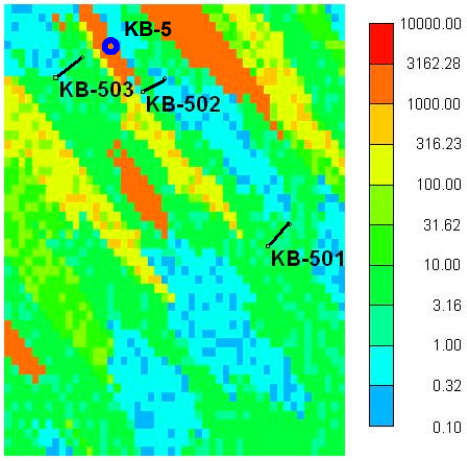


Figure 2: *Final realization of permeability at Krechba, predicting the presence of the high permeability feature between the injector (KB-502) and the appraisal well (KB-5)*

Conclusion

In conclusion, this methods allowed us to integrate dynamic and static data in the process of updating the conditional probability P{A|B,C}. As a result, the final model(s) yielded better predictions of the migration of the CO_2 plume, thus providing a cheap alternative for monitoring the fate of injected CO_2 in the subsurface aquifer and for initiating remedial action in the event of unexpected migration pathways.

References

Srinivasan, S. and Bryant, S. (2004) Integrating dynamic data in reservoir models using parallel computing approach. Paper SPE 89444 presented at the SPE/DOE 13th Symposium on Improved Oil Recovery, Tulsa, Oklahoma.

Journel, A. G. (2002) Combining knowledge from diverse sources: An alternative to traditional data independence hypothesis. Mathematical Geology, V. 34, No. 5, 573-596.

Caers, J. (1999) Markov Chain Theory for spatial stochastic simulation. Report No. 12, Volume 1, Stanford Center for Reservoir Forecasting, Stanford University

DETERMINING LOCAL SPATIAL VARIABILITY FROM A REGIONAL AIRBORNE RADIOMETRIC SURVEY

J. McKinley[1], A. Ruffell[1], M. Young[2]

[1] School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Elmwood Building, BT7 1 NN, Belfast, N. Ireland, j.mckinley@qub.ac.uk, +44 (0)28 9097 3827
[2] Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Malone Lower, BT9 5BF, Belfast, N. Ireland

Abstract

National and regional ground and remotely sensed surveys have become standard practice in generating large scale spatial databases and coverage maps of soil geochemistry, and regional geophysics. These are used for a wide range of applications comprising geological and soil mapping (e.g. organic soil carbon: see Rawlins et al. 2009), economic prospecting for mineral deposits, environmental studies (including national estimation of elemental soil concentrations for assessment of potentially harmful implications for health [Appleton et al. 2008]) and prediction of the soil provenance for forensic purposes (Lark et al. 2008). Different applications are based on the assumption that properties measured in the ground and airborne surveys are directly related to different aspects of the region such as underlying geology, superficial deposits and soil types. In all of these applications, assessing the baseline characteristics and spatial variability of the variables measured in the ground or airborne surveys is fundamental if accurate regional or national estimations are to be generated. The spatial variability or continuity depends on the distribution of the soil geochemistry or geophysical attribute. The assumption may be that spatial variability can be assumed to be similar within similar parent rock or soil types depending on the level of discontinuities such as faults within a certain area. Geostatistics provides a means through the variogram to investigate and quantify spatial correlation. This research investigates the use of variography to estimate local scale spatial variability in radiometric data generated by a regional remotely sensed geophysical survey flown in 2005 and 2006 as part of the Tellus project, GSNI. The survey collected information on naturally-occurring potassium (K), uranium (eU) and thorium (eTh), along with man-made radionuclides with an effective sample depth of 30cm (in bedrock). Survey lines were spaced 200 m apart and orientated NNW or SSE with a flying height of 56 m in rural areas and 240 m over urban areas. A subset of the radiometric data from the airborne geophysics survey was used to investigate local spatial variability and compositional differences in the Palaeogene acid intrusive rocks (granites and minor felsites) comprising the Mourne Mountains located in the SE of Northern Ireland. This area was selected because the rocks have relatively high K, U and Th contents and as result the effect of superficial peat deposits attenuating gamma rays may be minimised.

The granite intrusives of the Mourne Mountains cover an exposed area of approximately 150 sq km. The emplacement comprised successive stages of injections of acid intrusive creating five distinct granites: a feldspathic hornblende-granite, G1, a quartz-rich granite, G2, an aplitic fine-grained G3, a pink granite G4 and a late stage fine-grained micro-granite or granophyres G5. The results shown (Fig. 1) are for all radiometric data (total counts).

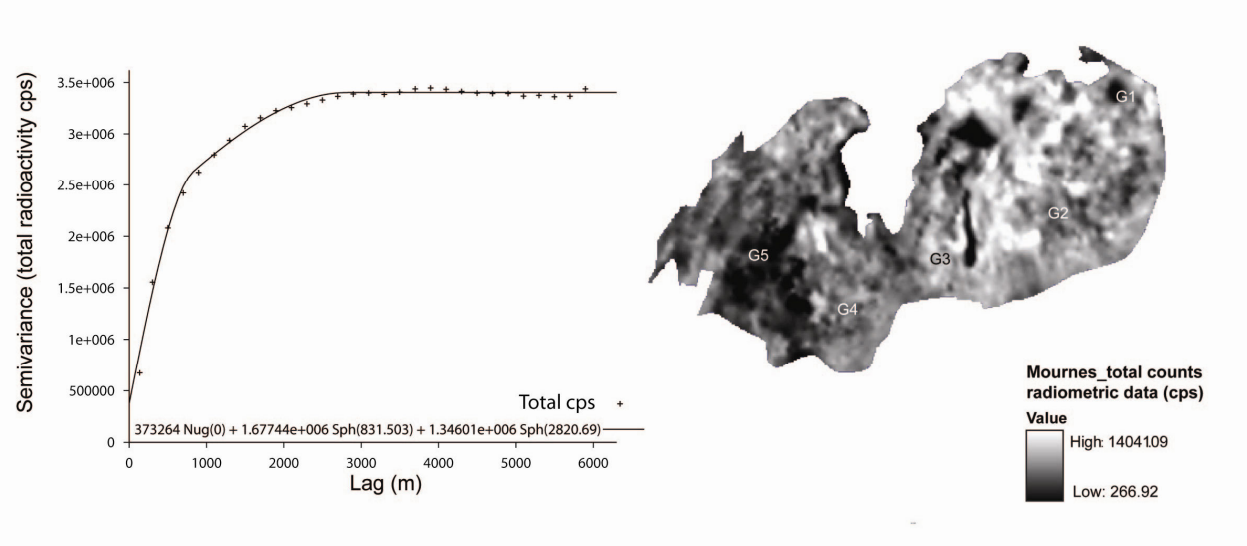


Figure 1: *Omnidirectional variogram modelled for total radiometric data. Ordinary kriged mapped output for all radiometric data using coefficients from modeled variogram. The successive stages of intrusions, G1 to G5 are identified.*

Estimation of the variogram was undertaken in order to investigate the spatial variation of radiometric signal across the granitic intrusives of the Mourne Mountains. The variograms were produced using Gstat (Pebesma and Wesseling 1998). A mathematical model was fitted to the experimental variograms using the weighted least squares functionality of Gstat. The mapped kriged map (Fig. 1) indicates spatial variability across the Mourne complex. Although some of the dark patches reflect peat cover and the sinuous water body of the Silent Valley Reservoir, the output suggests variability in the radiometric data also may be attributed to variability in chemical composition of the acid intrusives. The radiometric data were selected and analysed on the basis of chemical composition related to the successive stages of intrusion. The omnidirectional variograms for the quartz-rich granite, G2 and the aplitic fine-grained granite, G3 are shown in Figure 2. The modelled variogram (Fig. 2a) of the G2 granite data indicates periodicity which may be related to felsites intrusives. A nugget value indicates that there is spatial variability at a resolution finer than the scale of investigation of the airborne survey. The modelled variogram ranges indicate correlation distances from approximately 1 to 5 km. The modeled variogram for the fine-grained G3 granite suggests that the airborne survey has been successful in determining the spatial variability of the G3 granite. The range values indicate shorter correlation distances from approximately 0.5 to 1.7 km.

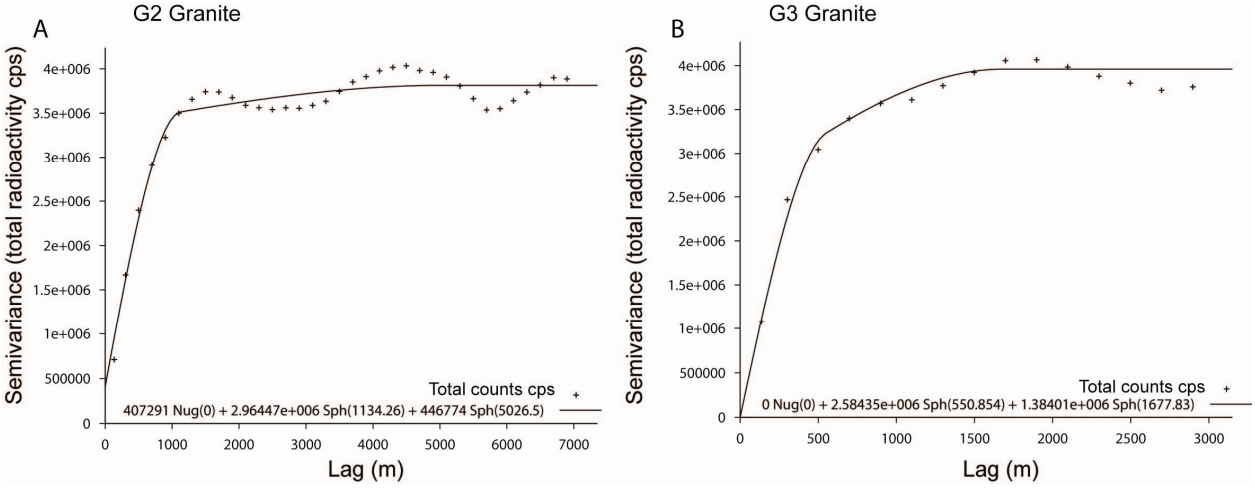


Figure 2: *Omnidirectional variograms modelled for total radiometric data for the instrusive stages G2 and G3.*

Further analysis involves the use of individual radioelements, eU, eTh and K to use the spatial variability of high eU and eTh relative to K to investigate if the compositional differences of the acid intrusive rocks can be identified from the airborne radiometric data.

Acknowledgements

The Tellus project was funded by the Northern Ireland Department of Enterprise Trade and Investment and by the Rural Development Programme through the Northern Ireland Programme for Building Sustainable Prosperity.

References

Appleton D, Rawlins B, Thornton I (2008) National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil : stream-sediment relationships. Applied Geochemistry 23(9): 2596-2611.

Lark, R.M, Rawlins, B (2008) Can we predict the provenance of a soil sample for forensic purposes by reference to a spatial database? European Journal of Soil Science 59(5): 1000-1006.

Pebesma E. J, Wesseling, C. G (1998) Gstat, a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24: 17–31.

Rawlins B.G, Marchant B.P, Smyth D, Scheib C, Lark R.M, Jordan C (2009) Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. European Journal of Soil Science, 60 (1): 44-54.

WAVELET-BASED POROSITY SIMULATION USING HIGH RESOLUTION ANALOGUE MODEL AND LOW-RESOLUTION TOMOGRAPHIC DATA

P. Tr. Simard[1], E. Gloaguen[1], D. Marcotte[2], and A. Boucher[3]

[1]Institut National de la Recherche Scientifique, Centre Eau-Terre & Environnement,
490 de la Couronne, Québec, QC, Canada, G1K 9A9, Patrick.Simard@ete.inrs.ca
[2]École Polytechnique de Montréal,
Dpt C.G.M., 6079 succ. Centre-ville, Montréal, EC, Canada, H3C 3A7
[3]Dpt of Environmental Earth System Science, Stanford University, Stanford, CA 94305, USA

Abstract

In geosciences, it is common to image a target property using indirect low-resolution geophysical imaging data. The task to down-scale low-resolution indirect data to the appropriate target scale is not trivial. Here, we propose to use a parametric texture model to obtain a statistical codification between high and low-resolution geophysical images. Wavelet coefficients obtained from both high and low-resolution tomographies of an analogue model are interpolated into a kernel. The wavelet coefficients of the high-resolution target porosity are generated using a Bayesian simulation of the wavelet coefficients from the measured velocity tomography and the interpolated kernel. High-frequency image structures are then generated to fit the statistical properties of the porosity analogue wavelet coefficients. The resulting inverse-transformed high-resolution porosity field shows high correlation with the known synthetic porosity model.

Introduction

Quantitative multi-scale porosity and hydraulic conductivity modeling is an important issue for groundwater characterization (Robinson et al., 2008; Srivastava, 1995). However, conventional data, i.e. well and pumping tests, does not permit to model the spatial continuity of the hydrogeological fields at the appropriate scale. In geosciences, it is common to image the spatial continuity of a target property using indirect low-resolution geophysical imaging systems. The task to down-scale low-resolution indirect data to the appropriate property target resolution is not trivial. In this study, we expand results from Portilla (2000) to obtain a statistical codification between high-resolution porosity analogue and low-resolution ground penetrating radar tomographic analogue, in order to simulate high-resolution porosity fields between two wells at the measured site.

Data and methodology

Two synthetic porosity fields were generated from two numerical representations of a sand deposit. From the two images, electrical property fields are calculated using known petro-physical relationships. Numerical simulation of a GPR tomography was done on both fields

leading to low-resolution electromagnetic velocity images of the corresponding porosity fields. One of the two fields will serve as a numerical analogue whereas the other is assumed to be the measured synthetic model. A steerable pyramidal wavelet-based parametric texture model (Simoncelli and Freeman, 1995) is applied to both analogue images (low-resolution GPR tomography and high-resolution porosity), as well as measured GPR tomography image. Figure 1 shows the kernel map from wavelet approximation coefficients of both analogue images.

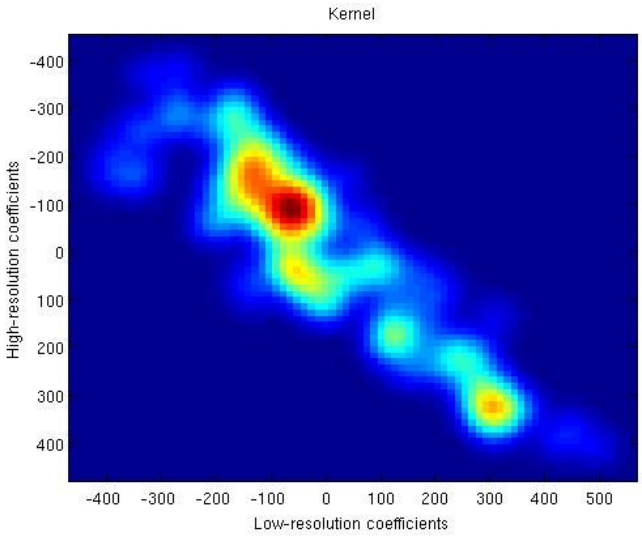


Figure 1: *Kernel of the analogue low-resolution GPR and the analogue high-resolution wavelet approximation coefficients.*

The kernel will serve to convert the measured low-resolution GPR wavelet coefficients in high-resolution porosity wavelet coefficients. More specifically, for each low-resolution GPR wavelet coefficients, a probability distribution function (*pdf*) of high-resolution coefficients is extracted from the kernel. The simulated high-resolution porosity wavelet coefficients are randomly generated from that *pdf*. Back-transform parametric texture synthesis of these random coefficients includes a simulated annealing which objective function honors both analogue wavelet coefficients and synthetic bore-hole data statistics. This method reproduces high-resolution textures of the target image as well as bore-hole data statistics (Figure 2). A loss of information can be seen with the low porosity lenses located on top left corner of the target porosity field and with the texture of lower zones.

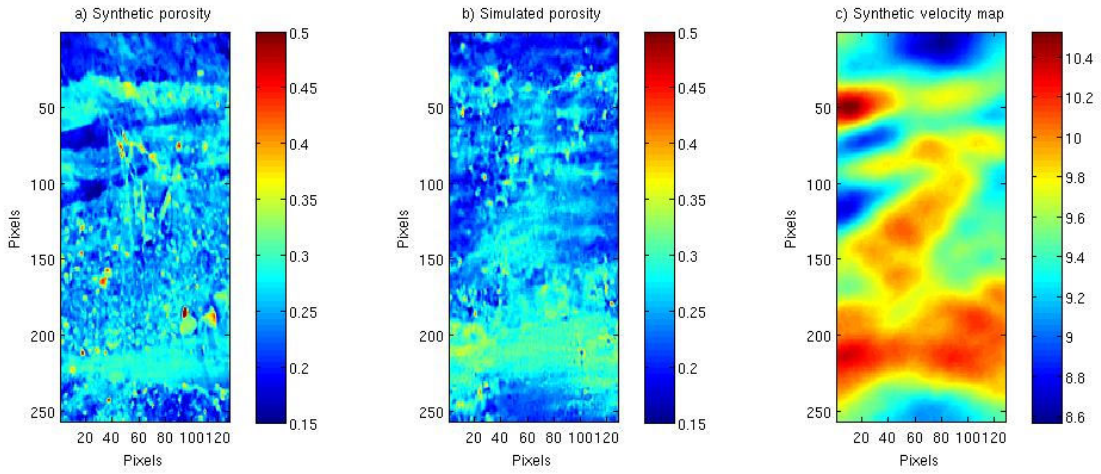


Figure 2: *a) normalized target porosity field, b) normalized simulated porosity field, and c) measured GPR tomography (m/ns).*

Conclusion

The proposed method enables high-resolution porosity field simulation from low-resolution information using a steerable pyramid. The main limitation of the method, inherent to training image-based algorithms, is in the construction of the analogue with similar wavelet statistics as the field under study. The method in not limited to geophysical radar data and can be extended to multivariate geophysical attributes.

Acknowledgments

This work was supported in part by Fonds Québécois de la Recherche sur la Nature et les Technologies, and Natural Sciences and Engineering Research Council.

References :

Portilla J. and Simoncelli E.P., (2000) A parametric texture model based on joint statistics of complex wavelet coefficients, International Journal of Computer Vision, Volume 40: 49-71

Robinson D.A., Binley A., Crook N., Day-Lewis F.D., Ferre T.P.A., Grauch V.J.S., Knight R., Knoll M., Lakshmi V., Miller R., Nyquist J., Pellerin L., Singha K. and Slater L., (2008) Advancing process-based watershed hydrological research using near-surface geophysics: a vision for, and review of, electrical and magnetic geophysical methods, Hydrological processes, Published online DOI: 10.1002/hyp.6963

Simoncelli E.P. and Freeman W.T., (1995) The steerable pyramid: a flexible architecture for multi-scale derivative computation, 2nd IEEE International Conference on Image Processing, Washington D.C., Volume III: 444-447

Srivastava R.M., (1995) An overview of stochastic methods for reservoir characterization, J. Yarus & R. Chambers, Eds., Stochastic modeling and geostatistics: Principles, methods and case studies, Volume 3, Chapter 1: AAPG computer applications in geology, Tulsa : 3-16

Walnut H., (2006) Fundamental Papers in Wavelet Theory, Princeton University Press: 878 pp.

TUESDAY 14 SEPTEMBER 2010

ORAL PRESENTATIONS

SPATIOTEMPORAL GEOSTATISTICS AND THE ENVIRONMENT: CROSSING PARMENIDES' GATES AND MORRISON'S DOORS.

G. Christakos

SDSU, San Diego, California, USA.

Those who are not mute in their souls could draw an intriguing analogy between Jim Morrison's simile that "There are things known and things unknown and in between are The Doors," and Parmenides' metaphor of "The Gates of Night and Day." In both cases, the role of the doors/gates is to offer a critical link between the unknown (world of ignorance) and the known (world of knowledge and truth). *Mutatis mutandis*, the goal of scientific inquiry is to crack the doors/gates connecting what is known with what is unknown about aspects of the human environment.

Viewed in the milieu of scientific inquiry, *spatiotemporal geostatistics* (STG) seeks mathematically rigorous and physically meaningful representations of in situ environmental systems under conditions of incomplete knowledge. The study of such systems is an increasingly complex matter that requires the collaboration of a variety of disciplines. Notions like random field, variogram, kriging, entropy, sysketogram and Bayes, *inter alia*, are suspended in STG's language. All these are formal and often powerful tools of problem-solving. Yet, no quantitative toolbox has any chance of real success, unless it is able to function in a broader context that integrates sound scientific methodology and substantive epistemological thinking and, on occasion, solid societal awareness. Accordingly, STG focuses not only on how a solution works (operational component), but also on why the solution works (substantive component), so that it can translate a general vision into an operational framework useful in real world problem-solving. STG is based on the general methodology of *Epibraimatics* that fuses epistemic principles of distinct fields to develop action-based mathematics for the solution of in situ problems. This methodology is schematically summarized by the following representation:

$$\left.\begin{matrix}\textit{Localized brain}\\ \textit{activities}\end{matrix}\right\} \rightarrow \left.\begin{matrix}\textit{Mental}\\ \textit{functions}\end{matrix}\right\} \rightarrow \left.\begin{matrix}\textit{Fundamental}\\ \textit{postulates}\end{matrix}\right\} \rightarrow \left.\begin{matrix}\textit{Mathematical}\\ \textit{operators}\end{matrix}\right\} \rightarrow \textit{Problem Solution.}$$

Mental functions emerge from brain activities in a way that reflects an adequate understanding of the dynamics of human nature. These functions subsequently lead to certain postulates that represent their essential features as close as possible. The postulates are finally translated into a set of mathematical operators used in integrative problem-solving. The representation above hints at a *symbiosis* of elements from brain science, neuropsychology, philosophy, linguistics, and mathematics within the contemplated solution space that promotes a twofold viewpoint: the concrete and the abstract. To understand both abstract concepts and concrete methods, an investigator must be able to consider them from different perspectives, and to interpret and

connect them to related concepts and methods.

The representation above forms the nucleus around which STG is developed. *Stochastic reasoning* underlines the conceptual and methodological framework of STG, and its formulation has a mathematical life of its own that (*a*) involves scientific inference in terms of physical law-based spatiotemporal dependence models, uncertainty functions, and stochastic logic conditionals; and (*b*) accounts for the multidisciplinarity of in situ problems, the multisourced uncertainties characterizing their solution, and the different thinking modes of the people involved. At any phase of its development, STG is expected to possess sufficient structure (theories and mathematical formalisms), with previous successes built into this structure, and guiding methodological principles that can encompass the different disciplines and generate new advances; and the solution of in situ problems will emerge by means of a *dialectic* between the problem space and the investigator (the dialectic may guide the investigator on how to organize and transform the problem space, revise the solution plan in light of new knowledge, discard elements that lead to dead ends and, when necessary, break out of the shackles of mainstream thinking and seek a fresh perspective). One implementation example of the methodological representation above in the case of space-time analysis and mapping is as follows:

$$\left.\begin{matrix}\textit{Stochastic}\\ \textit{Reasoning}\end{matrix}\right\} \rightarrow X_{\boldsymbol{p}} \sim \left.\begin{matrix} f_{KB} \\ \chi_p^{(\mathcal{R})}\end{matrix}\right\} \rightarrow \left.\begin{matrix} G \rightarrow \Omega_G \\ S \rightarrow \Omega_S \end{matrix}\right\} \rightarrow \left.\begin{matrix} \max_{f_G} \overline{I_G} \\ \text{s.t.} \quad \Omega_G \end{matrix}\right\} \left.\begin{matrix} \\ \\ \end{matrix}\right| \left.\begin{matrix} \\ \\ \end{matrix}\right| \rightarrow \boxed{\begin{matrix} \int d\boldsymbol{\chi}\,(\boldsymbol{g}-\bar{\boldsymbol{g}})\, f_G(\boldsymbol{\mu},\boldsymbol{g}) = 0 \\ \int d\boldsymbol{\chi}\,\boldsymbol{\xi}_S\, f_G(\boldsymbol{\mu},\boldsymbol{g}) - A[f_K(\boldsymbol{p})] = 0 \end{matrix}}$$
$$\left.\begin{matrix} \max_{f_K} \Psi_K \\ \text{s.t.} \quad f_G,\ \Omega_S \end{matrix}\right\} \left.\begin{matrix} \\ \\ \end{matrix}\right|$$

where $X_{\boldsymbol{p}}$ represents random fields, $\boldsymbol{p}$ denotes the space-time domain, Ω_G, Ω_S, $\boldsymbol{g}$, $\boldsymbol{\xi}_S$ are (core and specificatory) knowledge-based operators and vectors, f_{KB} are the associated probabilities, $\overline{I_G}$, Ψ_K express space-time information and adaptation measures, $\boldsymbol{\mu}$ are significance vectors, and A is a normalization functional. In real world applications STG modelling is conceived as a network of data bases, theories, beliefs, purposes and thinking modes in which any string in the net pulls and is pulled by the others in an interconnected way that can change the configuration of the whole. Applications involve a variety of fields such as environmental assessment, human exposure, health effects, risk analysis, TGIS, epidemiology, and medical geography.

The phase of *Decadence* characterizes every societal aspect (science, education, politics, economics, and culture). In higher education and research, e.g., one witnesses what is called an unholy alliance of financial corporatism and radical postmodernism (undermining tradition, knowledge, language and achievements of the past; promoting nihilism, and seeking to satisfy lower needs). In industry and institutionalized development the so-called corporate geostatistics has emerged in the tradition of Mr. Grandgrind's teachings in Charles Dickens' novel "Hard Times." Ubiquitous self-appointed expert practitioners, petulant and critical, protest against any kind of intellectually challenging theories, abstract thinking, contemplative analysis, and the like. Such peevish criticism is anything but practical, of course. By now it is widely accepted that

many of the simplistic and pseudo-practical techniques routinely used in corporate geostatistics lack methodological continuity and maturity, are not interrelated in a way that can offer a sound body of knowledge, and refer to situations with no scientific substance. Sometimes, it may be true that self-styled "bottom-line" and "no-nonsense practicality" methods are classic throwaway lines with a pleasant populist tinge that satisfy the "limited attention span" requirement. Yet, the shadow epistemology underlying such sound bites and slogans is deeply unsatisfactory and inefficient, ignoring the basic principles of space-time change and consistency (physical and logical) between the different information sources. The same epistemology de-emphasizes the quality of knowledge in favor of satisfying the need of the "quick and dirty" solution by which knowledge is encumbered and to which it is subordinate. In the end, just as the deceitful promises of "no-nonsense practicality" and "quick results" characterized the era of the financial corporatism bubble, the same promises have marked indelibly the fate of the corporate geostatistics *bubble.*

In light of the above, the consideration of Decadence and its effects is essential in the realistic study of environmental problems and their sustainable solution. Indeed, the *broad context* within which the problems emerge can affect their solution. The STG inquiry needs to integrate sophisticated technical tools and physical understanding with philosophical thinking and sociological analysis. Only by interpolating between the full range of disciplines (including mathematics, physics, neuropsychology, philosophy, history, and sociology) and the associated thinking modes the scientists can arrive at a satisfactory account of problem-solving, and be able to distinguish between a technically complete solution, and one that has social impact. To call a spade a spade, in a time of Decadence a thinker needs to be both *within* things and *outside* them; to develop the capacity to stand back from everyday experience and gain a broader problem-solving view. STG problem-solvers should not be isolated within the strict boundaries of technical expertise, but possess a wide educational background and constant awareness of the broad context (social, political, economic, and cultural) within which they operate. In this way they can be more valuable to their scientific fields, and, also, more useful to the public at large.

SPATIAL ASSESSMENT OF INFILTRATION CAPACITY OF SOILS FOR ARTIFICIAL RECHARGE PRACTICES USING GOOGLE EARTH IMAGES

D. Pedretti[1], M. Barahona-Palomo[1], D. Bolster[1-2], D. Fernàndez-Garcia[1], X. Sanchez-Vila[1]

[1] Department of Geotechnical Engineering and Geosciences, Technical University of Catalonia (UPC), Barcelona, Spain

[2] Department of Civil Engineering and Geological Sciences, University of Notre Dame, Indiana, USA

Amongst the many practices that exist for managed aquifer recharge, infiltration of water from the surface of artificial ponds is a viable option (e.g. Boewer, 2002; Jha et al., 2009). The design and long-term maintenance of these facilities depends on the proper assessment of the characteristic times of water residence within the ponds. These are essential to predict the quantity and the quality of the water volumes recharging the water. Characteristic times are directly related to the infiltration capacity of the soil in which the ponds are built. This property depends on the local geology, which is typically heterogeneous, and on kinetic physical, chemical and biological clogging processes on the run. The assessment of the initial infiltration capacity can be prohibitively expensive and time consuming, or even technologically unfeasible. Also, the exploration scales of these tests are limited (using traditional infiltrometers) to 10^{-1} – 10^{0} m^2. A spatial correlation between these primary values needs a sufficient number of initial sampling zones. In this sense, we explore an alternative way to assess the spatial variability of the infiltration capacities of soils using free satellite images. We assume that the color of the image pixels can be correlated to the hydrodynamic characteristic of the soil. Provided some premises are fulfilled (like the monotonic behavior of the interpolation model), we analyze the RGB raster to get exhaustive secondary information of the desired variable. In this sense, a correlation (variogram) model can be evaluated to model the spatial structure of the image. We tested our method in a 42 m x 100 m portion of an artificial aquifer recharge pond in the Llobregat delta, close to the Barcelona (Spain) (fig. 1). The rectangle in fig. 1 has a resolution of 326 x 730 pixels. Field measurements via double-ring infiltration tests in specific locations within the pond were satisfactorily correlated by a logarithmic trend to the pixel intensities of the Google Earth high resolution image available on the Nov. 15th, 2007. The correlation exists in all three color bands (Red, Green and Blue) of which the Google Earth capture is composed (Fig. 2 shows only the red band correlation). The spatial variability of the pixel values is fully described at the pilot area. We observed that the stationarity is reached within the test area; a bidirectional, multiple-structured model (fig. 3) fits the experimental variogram. We cokriged a map coupling the primary information from the infiltration datasets to the pixel values using a Markov model

(to avoid filtering effects of the density of the secondary variable with regards to the primary one).

The analysis shows that our predictions of the maximum infiltration rates (calculated mean of the cokriged map: 3.72 m/d) is close to the actual observations of the local water agencies and public administrations (about 3.6 m/d). This method could in principle be applied to other fields (i.e. agriculture, groundwater modeling, water management), as a way to quickly predict overall hydraulic properties spanning various scale in an inexpensive manner. Following the example of recharge ponds, for instance, the design of a network of new facilities can be performed developing stochastic simulations around the test area (fig.4), when the correlation is found to be stationary. The new ponds can be given specific expected characteristic times of water residence, depending on the pond positions; thus, specific management can be adopted for each of the new ponds (like inflow discharge rates or biological and chemical treatments of the water).

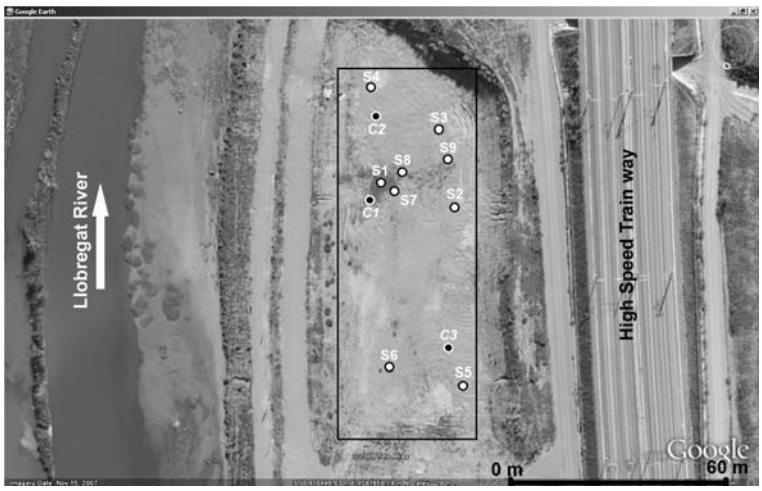


Figure 1: *Red band of the studied portion of the image. Infiltration test sites are shown (S1-S9); C1 to C3 indicate other excavated pits locations. The size of the rectangle is 42x100 m*^{2}

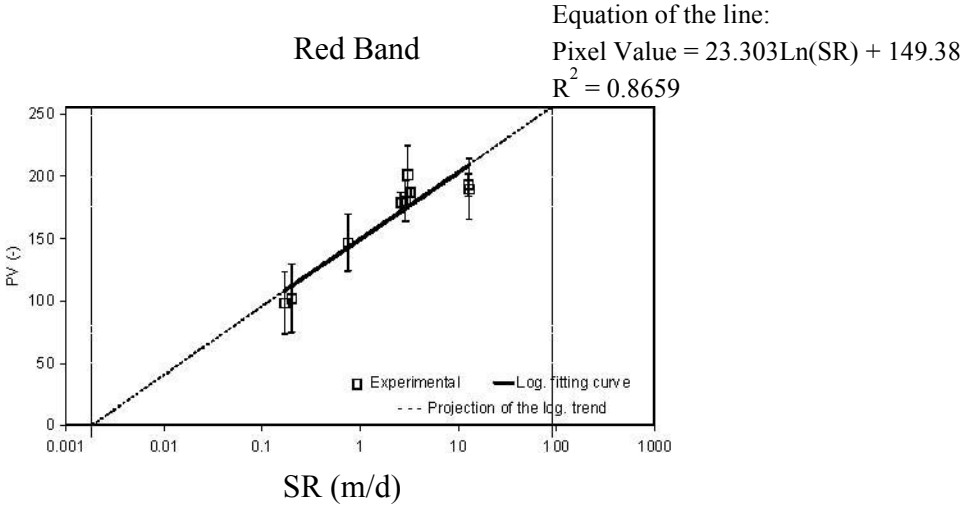


Figure 2: *Correlation between pixel values and seepage rates in the red band.*

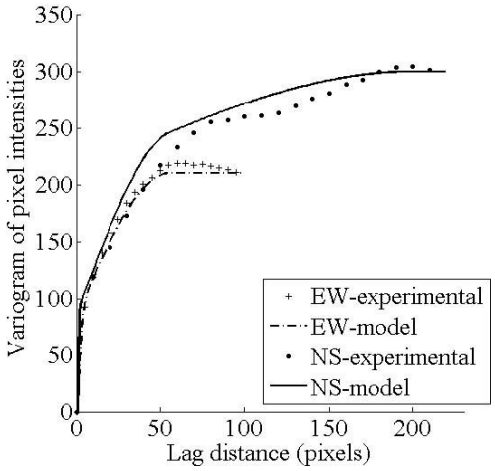


Figure 3*: N-S and E-W variograms of the studied image (experimental and model are depicted). The lag distance is expressed in pixels (1 m ~ 7 pixels)*

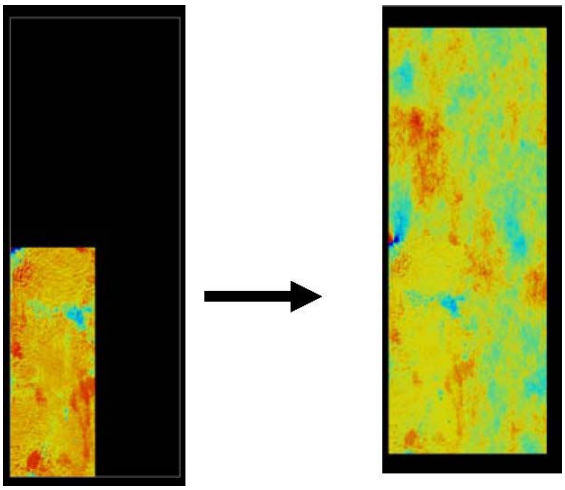


Figure 4: *A single realization of a new area (4 times bigger than the original testing area), conditioned to the initial map and using the spatial structure described by the variogram (see fig. 4)*

References

Bouwer H, (2002) Artificial recharge of groundwater: hydrogeology and engineering. Hydrogeology Journal, DOI 10.1007/s10040001-0182-4 Jha MK, Kamii Y, Chikamori K, (2009) Cost-effective Approaches for Sustainable Groundwater Management in Alluvial Aquifer Systems. Water Resources Management, 23(2), 219-233.

TRANSPORT UPSCALING USING MULTI-RATE MASS TRANSFER IN THREE-DIMENSIONAL HIGHLY HETEROGENEOUS POROUS MEDIA

L. Li, H. Zhou, J. J. Gómez-Hernández

Group of Hydrogeology, Universidad Politecnica de Valencia, Spain,
liali@upvnet.upv.es, Tel. (34) 963879615, Fax. (34) 963879492

Albeit upscaling flow and transport has been disregarded by some on the basis of increasing computer capabilities, there is always a discrepancy between the scale at which we can characterize the medium, and the scale at which we can run our numerical codes, rendering upscaling necessary to transfer the information collected at the fine measurement scale into a coarser computational scale well-suited for modeling.

In the current work, motivated by prior analysis, we follow the study of Fernandez-Garcia et al. (2009) and extend proposed transport upscaling method using multi-rate mass transfer into three-dimensional highly heterogeneous formulations, coupled with the elaborated interblock Laplacian with skin hydraulic conductivity upscaling method to more accurately preserve the memory in space compared with block simple Laplacian technique. Additionally, unlike previous studies that focused primarily on the realization-by-realization match, as is the intent of existing hydraulic conductivity upscaling techniques, the present study is characterized by an attempt to ensemble level upscaling to achieve the agreement of statistics of transport characteristics within two scales. Moreover, the propagation of solute transport associated with the uncertainty was further evaluated in the upscaling processes.

Consider a synthetic three-dimensional confined aquifer under a uniform, natural-gradient flow condition, serving as the reference to illustrate the effectiveness of upscaling scheme on flow and transport. The variance of natural logarithm hydraulic conductivity is set with value of 4.0.

In the fine scale, the five-point block-centered finite-difference groundwater flow model MODFLOW 2000 was employed to solve the fine scale flow. The velocity crossing the interface are calculated, and then were utilized in the random walk particle tracking code RW3D (Fernandez-Garcia et al., 2005), which was used to solve the fine scale transport equation since it is free from numerical dispersion and computationally efficient. In the coarse scale, the nineteen-point block-centered finite-difference groundwater model FLOWXYZ (Li et al., 2010) was employed to solve the coarse scale flow equation. Again, the RW3D was used to solve the coarse scale multi-rate transport equation based on the methodology presented by Salamon et al. (2006).

With regard to the flow upscaling, the results clearly show that the interblock Laplacian with skin approach is better to reproduce the fine scale flow compared with the block simple Laplacian method in highly heterogeneous media. For transport upscaling, we can see that the breakthrough curves are reproduced very well (Fig 1). We further evaluate the uncertainty by

calculating the spread in the ensemble of accumulative breakthrough curves at all the control planes. For the late time, it is evident that the use of double-rate mass transfer model outperforms the primitive model (Fig 2).

In this study, we have presented and demonstrated a new algorithm for the transport upscaling to avoid the computational burden in three-dimensional highly heterogeneous media. This work is an extension of the work by Fernandez-Garcia et al. (2009) in two dimensions. One of the critical features of this method is that, it uses an elaborated Lalpacian with skin approach to reproduce the flows instead of the simple Laplacian scheme, and the multi-rate mass transfer process is used to compensate the loss of information in the subgrid during upscaling. Not only do the hydraulic conductivities upscaling underestimate the late arrival time of breakthrough curves, it also severally underestimates its uncertainty. The loss of uncertainty can be compensated by the mass transfer model allowing solute mass exchange between high-permeability and low-permeability zones.

References:

Li, L. P., Zhou, H. Y., Gomez-Hernandez, J. J., 2010. in press. Steady-state groundwater flow modeling with full tensor conductivities using finite differences. Computers & Geosciences, doi:10.1016/j.cageo.2010.04.002.

Fernandez-Garcia, D., Llerar-Meza, G., Gomez-Hernandez, J. J., 2009. Upscaling transport with mass transfer models: Mean behavior and propagation of uncertainty. Water Resour. Res. 45, W10411.

Fernandez-Garcia, D., Illangasekare, T. H., Rajaram, H., 2005. Differences in the scale dependence of dispersivity and retardation factors estimated from forced-gradient and uniform flow tracer tests in three-dimensional physically and chemically heterogeneous porous media. Water Resour. Res. 41 (3), W03012.

Salamon, P., Fernandez-Garcia, D., Gomez-Hernandez, J. J., 2006. Modeling mass transfer processes using random walk particle tracking. Water Resources Research 42, W11417.

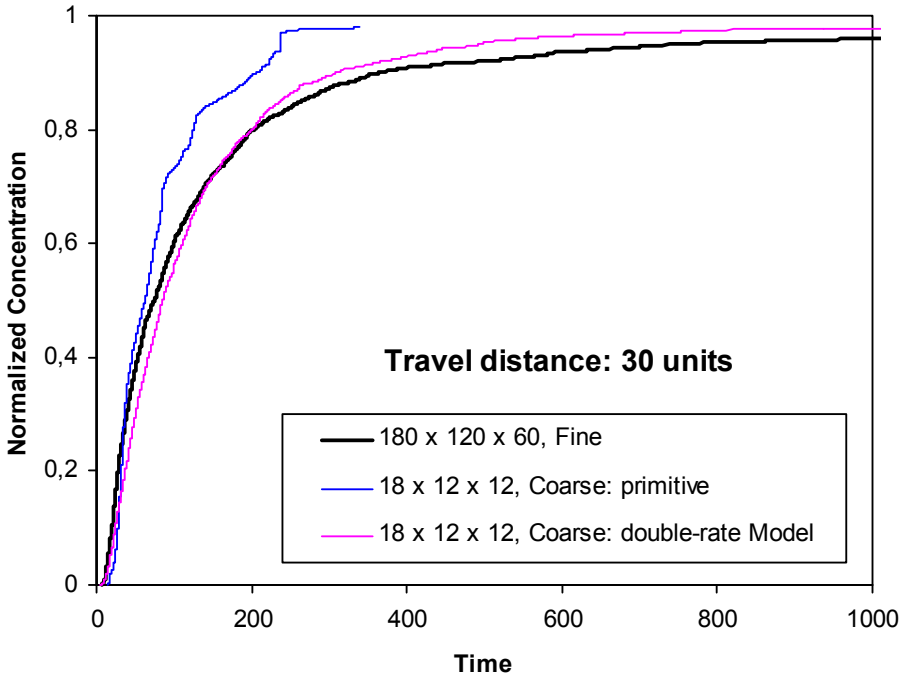


Figure 1: *Comparison of fine scale accumulative breakthrough curves with those obtained by the upscaled transport models. (primitive: only flow upscaling)*

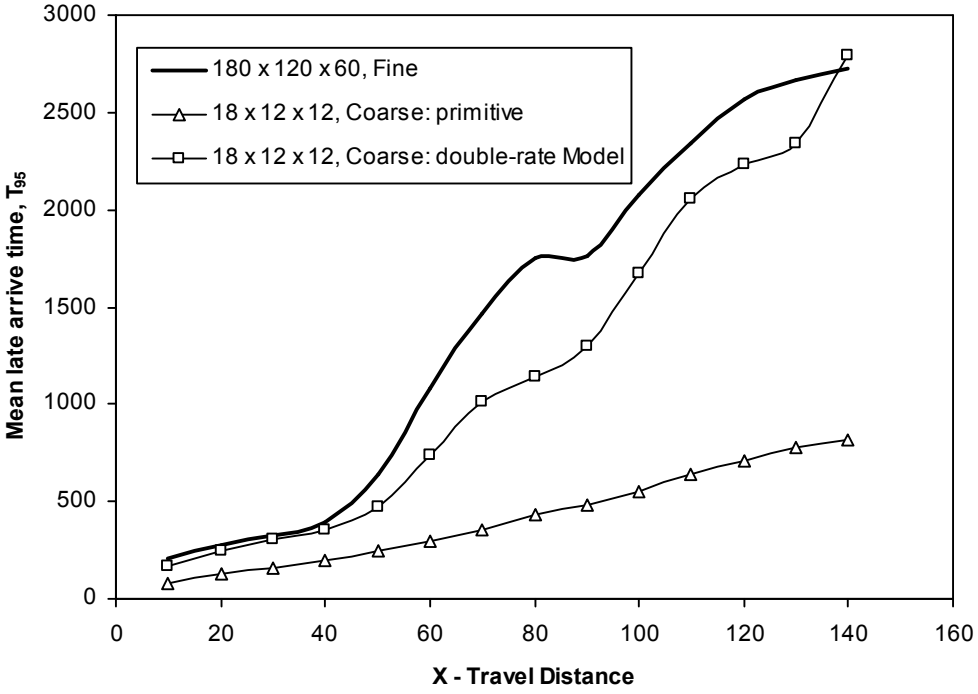


Figure 2: *Ensemble travel late times as a function of travel distance, and comparison of fine scale simulation with those corresponding to the upscaled models (primitive: only flow upscaling).*

STOCHASTIC SIMULATION OF TIME SERIES MODELS COMBINED WITH UNIVERSAL KRIGING: A FRAMEWORK TO PREDICT RISKS OF WATER TABLE DEPTHS IN TIME AND SPACE

R. Lilla Manzione[1,4], E. Wendland[2], M. Knotters[3]

[1]UNESP/São Paulo State University – Campus of Ourinhos. Av. Vitalina Marcusso, 1500, CEP: 19910-206 Ourinhos (SP), Brazil (manzione@ourinhos.unesp.br) Tel: +55 1433025716
[2]USP/University of São Paulo – EESC/Department of Hydraulics and Sanitary Engineering. PO Box 359, CEP: 13560-970 São Carlos (SP), Brazil (ew@sc.usp.br)
[3]ALTERRA/Soil Science Centre. Droevendaalsesteeg 4, 6708PB Wageningen, The Netherlands (Martin.Knotters@wur.nl)
[4]Project FAPESP#2009/05204-8

INTRODUCTION

Monitoring data of piezometric heads in a watershed gives information about the dynamic of the aquifer domain in time and space. Time series modelling is an elegant way to treat monitoring data without the complexity of physical mechanistic models. Also, the stochastic component of the model allows us to account for model uncertainty. Simple transfer function-noise models have been successfully applied to describe the dynamic relationship between precipitation surplus and water table depth in several studies (Tankersley and Graham 1994; Van Geer and Zuur 1997; Knotters and Van Walsum 1997; Yi and Lee 2003; Von Asmuth and Knotters 2004). The transfer function-noise models can be calibrated to a set of time series observed in various wells in an area. Next the time series model parameters or the model predictions can be interpolated spatially, using ancillary information related to the physical basis of these models (Knotters and Bierkens 2001). In this approach the spatial differences in water table dynamics are determined by the spatial variation in the system properties, while its temporal variation is driven by the dynamics of the input into the system. Aim of this work is to present a combined framework of time series modelling and geostatistics applied to groundwater monitoring data to predict risks of extreme water table depths in time and space.

PREDICTING RISKS OF EXTREME WATER TABLE DEPTHS IN TIME AND SPACE

The PIRFICT-model is an alternative to discrete-time transfer function-noise models where a block pulse of the input is transformed to an output series by a continuous-time transfer function (Von Asmuth et al. 2002). We consider this model because it not only enable us to account for model uncertainty, but also has physical based parameters given from the response function that models the dynamic relationship between precipitation surplus and water table depths. The simulation procedure of the PIRFICT-model is described in Manzione et al. (2008). Here, the 5^{th} and 95^{th} percentiles of the simulated water table depths are interpolated spatially using Universal Kriging (Matheron 1969; Pebesma 2004). These

maps present the water table depths that will be exceeded with 5 and 95% chance in some specific date, sampled from the simulated series. The digital elevation model is used as ancillary information in Universal Kriging. The aim of using elevation as a covariate is to improve the accuracy of the spatial predictions. Also, it can enhance the physical meaning of the predictions and yield more plausible spatial patterns. Elevation is physically related to water table depths, since it is related to local geomorphology. Areas with relatively low elevation and close to drainages present shallow water tables, whereas in areas with relatively high elevation and far from drainage devices the water table is deep (Furley 1999). The digital elevation model is incorporated as a drift in the spatial prediction model. Let the probability of exceedance be given as $z(x_1)$, $z(x_2)$, …, $z(x_n)$, where x_i is a two-dimensional well location and n is the number of observations. At a new, unvisited location x_0 in the area, $z(x_0)$ is predicted by summing the predicted drift and the interpolated residual (Odeh et al. 1994; Hengl et al. 2004):

$$\hat{z}(x_0) = \hat{m}(x_0) + \hat{e}(x_0) \tag{1}$$

where the drift m is fitted by linear regression analysis and the residuals e are interpolated using kriging:

$$\hat{z}(x_0) = \sum_{k=0}^{p} \hat{\beta}_k \cdot q_k(x_0) + \sum_{i=1}^{n} w_i(x_0) \cdot e(x_i); \quad q_0(x_0) = 1 \tag{2}$$

Here, the β_k are estimated drift model coefficients, $q_k(x_0)$ is the kth external explanatory variable (predictor) at location x_0, p is the number of predictors, $w_i(x_0)$ are the kriging weights and $e(x_i)$ are the zero-mean regression residuals. In this case, for water table depths h, the model estimator was formulated as follows:

$$\hat{h}(x_0) = \beta_0 + \beta_1 \cdot E(x_0) + e(x_0) \tag{3}$$

where E is the elevation value for each location and e is a zero-mean spatially correlated residual. Its spatial correlation structure is characterized by a semivariogram. The results of spatial interpolation are evaluated using cross-validation (Chilès and Delfiner 1999).

CASE STUDY: Monitoring water table depths in a Guarani aquifer recharge area

The study area is an outcrop zone of the Guarani aquifer system, in the municipality of Brotas, province of São Paulo, Brazil. The Ribeirão da Onça watershed is equipped with 23 wells installed in the watershed for groundwater level monitoring (Wendland et al., 2007). Water table depths were observed manually every 15 days, from April, 2004 until April, 2009, totalizing five years (1826 days) of continuous monitoring. These wells were selected purposively to cover the range of land uses and hydrogeological domains in the area, in a try to characterize the different responses of water table depths in the basin. The filter levels of the wells varied with soil depth. Series of precipitation and potential evapotranspiration data were available from CRHEA/USP (Center for Water Resources and Applied Ecology of the University of Sao Paulo), where climatologic data are collected continuously. These data were available from 1974 until 2009 (35 years) with a daily frequency. The watershed is

located approximately 1.5 km away from the station. The digital elevation model of the area is in a 1:10:000 scale, digitalized and interpolated from topographic maps.

The resulting maps can be used to evaluate possible risk areas of extreme low or shallow water table depths during any period of the year. These are valuable information for decision making and water management policies, in order to balance economical and ecological purposes in ground water exploitation. The method presented has demonstrated potential for unconfined and porous aquifers in Brazil.

REFERENCES

Chilès, JP, Delfiner, P (1999) Geostatistics: Modeling spatial uncertainty. Wiley-Interscience, New York

Furley, PA (1999) The nature and diversity of neotropical savannah vegetation with particular reference to the Brazilian Cerrados. Global Ecol Biogeogr 8: 223-241

Hengl, T, Heuvelink, GBM, Stein, A (2004) A generic framework for spatial prediction of soil properties based on regression-kriging. Geoderma 120: 75-93

Knotters, M, Van Walsum, PEV (1997) Estimating fluctuation quantities from time series of water-table depths using models with a stochastic component. J Hydrol 197: 25-46

Knotters, M, Bierkens, MFP (2001) Predicting water table depths in space and time using a regioinalised time series model. Geoderma 103: 51-77

Manzione, RL, Knotters, M, Heuvelink, GMB, Von Asmuth, JR, Câmara, G (2008) Predictive risk mapping of water table depths in a Brazilian Cerrado area. In: Stein A, Wenzhong, S, Bijker, W (eds) Quality Aspects in Spatial Data Mining. Boca Raton, CRC Press

Matheron, G (1969) Le krigeage universel. Cachiers du Centre de Morphologie Mathematique 1. Ecole des Mines de Paris, Fontainebleau

Odeh, I, McBratney, A, Chittleborough, D (1994) Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63: 197-214

Pebesma, EJ (2004) Multivariable geostatistics in S: the Gstat package. Comput Geosci 30: 683-691

Tankersley, CD, Graham, WD (1994) Development of an optimal control system for maintaining minimum groundwater levels. Water Resour Res 30: 3171-3181

Van Geer, FC, Zuur, AF (1997) An extension of Box-Jenkins transfer/noise models for spatial interpolation of groundwater head series. J Hydrol 192: 65-80

Von Asmuth, JR, Bierkens, MFP, Maas, C (2002) Transfer function noise modelling in continuous time using predefined impulse response functions. Water Resour Res 38: 23.1-23.12

Von Asmuth, JR, Knotters, M (2004) Characterising groundwater dynamics based on a system identification approach. J Hydrol 296: 118-134

Wendland, E, Barreto, CEAG, Gomes, LH (2007) Water balance in the Guarani Aquifer outcrop zone based on hydrogeologic monitoring. J Hydrol 342: 261-269

Yi, M, Lee, K (2003) Transfer function-noise modeling or irregular observed groundwater heads using precipitation data. J Hydrol 288: 272-287

PERFORMANCE OF MODIFIED ENSEMBLE KALMAN FILTER IN NON-GAUSSIAN HETEROGENEOUS MEDIA

H. Zhou , J. J. Gómez-Hernández and L. Li

Group of Hydrogeology, Department of hydraulics and environmental engineering, Universidad Politécnica de Valencia, 46022, Valencia, Spain. Email: haizh@upvnet.upv.es Tel: 34 963879615, Fax: 34 963879492

Model structure, boundary conditions, sources or sinks, and more importantly, parameters like conductivity and porosity are necessary input for groundwater flow and transport simulation. However, detailed description of the geologic is not available so that we usually have to resort to inverse methods. Inverse problem involves integrating the response of the groundwater system, such as hydraulic head, flux and concentration of contaminant, to infer the parameters of the aquifer. Ensemble Kalman Filter ("EnKF" hereinafter), as one of the Monte Carlo data assimilation algorithms, is first proposed by Evensen (1994) and further improved by Burgers et al. (1998), and now it is widely applied in such fields as oceanography, meteorology and hydrology.

EnKF mainly consists of two steps: forecast the state until the time when the measurements are available and then update it by assimilating the observation as follows:

$$X_t^f = F(X_{t-1}^a)$$
$$X_t^u = X_t^f + K_G(Z - HX_t^f)$$

where X_t^f is the state at time t forecasted from the state at time t-1 by the forward transition function F which could be the flow equations, K_G represents the Kalman gain calculated through the ensemble under the criterion of minimizing the error covariance, Z means the observations and H maps the forecast into the observations.

EnKF gains more and more attention as a data assimilation technique due to its ease of implementation and computation efficiency. However, it also suffers from some drawbacks, for instance, its intrinsic assumption of Gaussian distribution and risk of divergence. We focus on relaxing the Gaussian constraint by introducing a transformation into the regular EnKF, whose performance will be evaluated by a non-Gaussian synthetic experiment characterized by curvilinear channels (Fig 1).

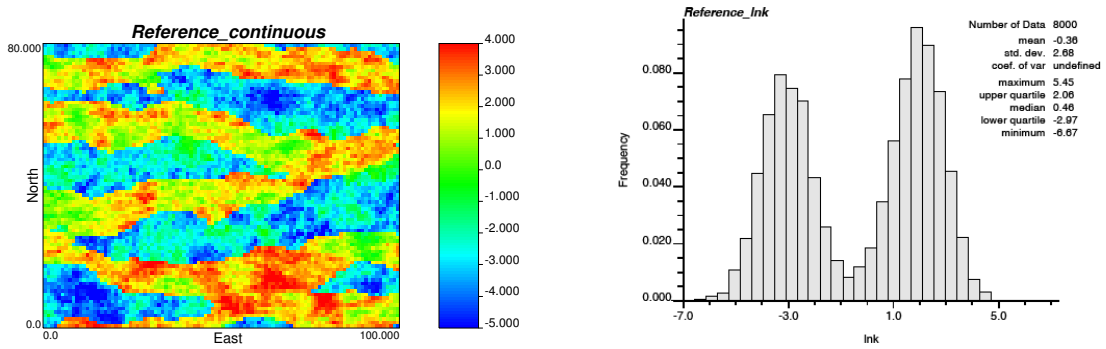


Figure 1: *Reference conductivity field*

An ensemble of 100 realizations is generated conditional on the conductivity measurements serving as the starting point of the inverse problem (Fig 2).

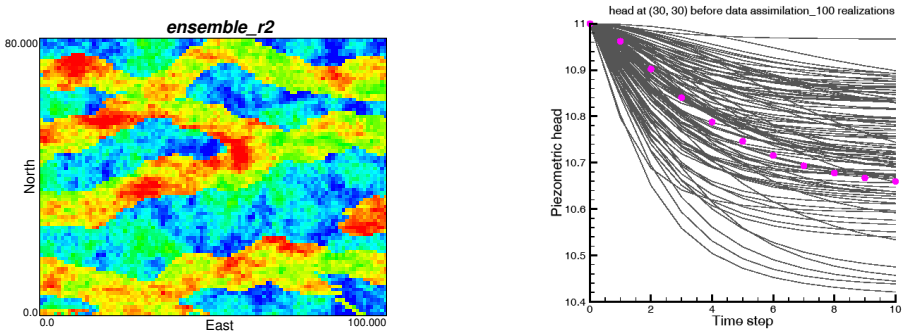


Figure 2: *One of the realizations of the initial ensemble and head evolution at node (30, 30) where the pink circles indicate the reference*

One of the realizations after assimilating head observations using the traditional EnKF is listed in Fig 3. We can find that the channel structure is distorted although the uncertainty of head calibration is reduced. In other words, EnKF does not preserve geologic structure, which is of significance in the subsequent transport simulation (Mariethoz, 2009).

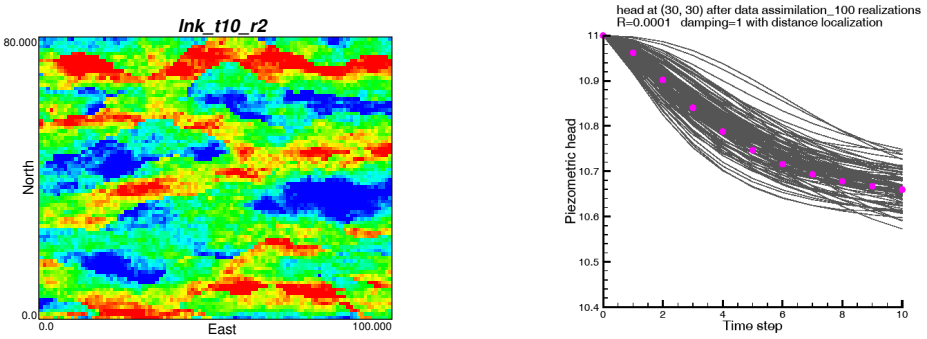


Figure 3: *One of the realizations of the updated ensemble of conductivity by traditional EnKF and head calibration at node (30, 30)*

By using a damping factor in the updating equation, we can attenuate the influence of the difference between the measurement and the forecast (Hendricks Franssen and Kinzelbach,

2008) which will give rise to a mild change to the conductivities. Alternatively, we can employ some type of transformation on the conductivities, for instance, a normal score transform. The aim of the transform is two-fold: relax the Gaussian assumption and constrain the ensemble to the initial structure as well. Updated realizations and the head evolution are plotted in Fig 4. It can be found that the channel structure seems more reasonable in comparison with the result of regular EnKF (Fig 3) and the uncertainty of flow calibration is reduced compared with the unconditional case (Fig 2), indicating that such transformation really improved the performance of EnKF.

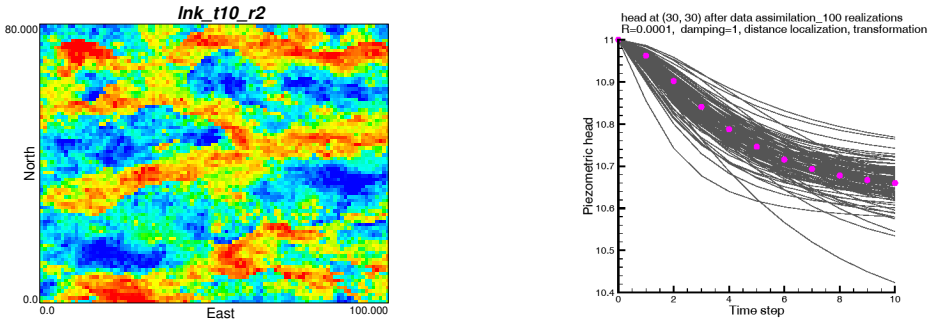


Figure 4: *One of the realizations of the updated ensemble of conductivity by EnKF with transformation and flow calibration at node (30, 30)*

Acknowledgements. The authors gratefully acknowledge the financial support by ENRESA (project 0078000067) and the European Commission (projects FUNMIG and PAMINA). The first author also acknowledges the financial aid through a fellowship from the China Scholarship Council (CSC). We thank Dr.Jianlin Fu in Stanford University for providing his code of EnKF.

References

Burgers G, van Leeuwen P J, Evensen G, 1998. Analysis scheme in the ensemble kalman filter. Mon Weather Rev 126: 1719-1724

Evensen G, 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using monte carlo methods to forecast error statistics. J Geophys Res 99: 10143-10162

Hendricks Franssen H.-J, Kinzelbach W, 2008. Real-time groundwater flow modelling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44, doi:10.1029/2007WR006505

Mariethoz G., 2009. Geological stochastic imaging for aquifer characterization. Ph.D.thesis, Universite de neuchatel, Neuchâtel

STOCHASTIC MODELING FOR LIQUID RADIOACTIVE STORAGE RESERVOIR

E. Savelyeva, S. Utkin

Nuclear Safety Institute (IBRAE) Russian Academy of Sciences, 52 B.Tulskaya, Moscow, Russia, esav@ibrae.ac.ru, tel. +7 495 9552231, fax +7 495 9581151

This work studies the dynamics of a technical reservoir with liquid radioactive wastes. The main characteristic presenting the dynamics of the reservoir is a level at the south end of it. It is a location where the reservoir is blocked from the open hydrological system by a dumb. The main problem is to estimate if a height of the dumb is enough to prevent an overflow above it. In the current work the probabilistic approach is used.

There is a set of factors influencing on the reservoir's level: 1) the technical liquid wastes. The reservoir is still in use for technical activities. The amount of such income is decreasing during last years, but still at a moment it is impossible to stop the release; 2) the weather conditions such as precipitation and evaporation at the reservoir surface; 3) exchange with the underground water system. It can also be supposed that there is the dependence between the amount of the filtering water and the precipitation/evaporation conditions at some territory. All these factors are classical components of a water balance equation for a not flowing through reservoir. But very high uncertainty in all influencing factors does not allow to construct and to solve an exact equation. It leads to probabilistic water balance equation [Ryazantsev & Trapeznikov, 2008].

Otherwise one can use the learning from data paradigm: the level can be considered as a main (primary) value accumulating all knowledge of the process. Thus all dynamics of the process can be found in the dynamics of the main value itself. All other available information and knowledge can be used as a secondary one.

The preliminary study of the dynamics of the main value by means of a phase portrait allows to indicate a type of a process under study. The process is presented by several stable and transfer periods. Application of additional information allows to formulate a purpose starting a transfer period, which appears to be a deterministic one – an increase of the yearly amount of the technical release. So, the deterministic release component is subtracted from the process.

Thus the process under study looks as a reduced water balance equation:

$$\varphi = \Delta H - \frac{Q}{S(H)}, \qquad (1)$$

where ΔH is a decrement of the level, Q is the technical release, $S(H)$ is the surface of the reservoir dependant on its current water level. The analytic relationship between surface S and the measured level H was presented in Mokrov et al, 2009.

The process φ can be considered as a stochastic process with a deterministic, for example a linear one, trend:

$$\varphi_t = \Psi_{\text{det}}(\varphi_{t-1},\ldots) + \psi_{t-1}.$$

$$\Psi_{\text{det}} = 0.457\varphi_t - 0.003117. \qquad (2)$$

The simplest way to deal with the residual component is to treat it as an independent normally distributed random values. The parameters of the normal distribution can be obtained from the real residual sequence, they are: mean = 0.000024, variance = 0.002. Using such approach one can obtain a set of stochastic realizations reproducing the dynamics of the water level and estimate the mean value and 95% confidence intervals (Fig. 1).

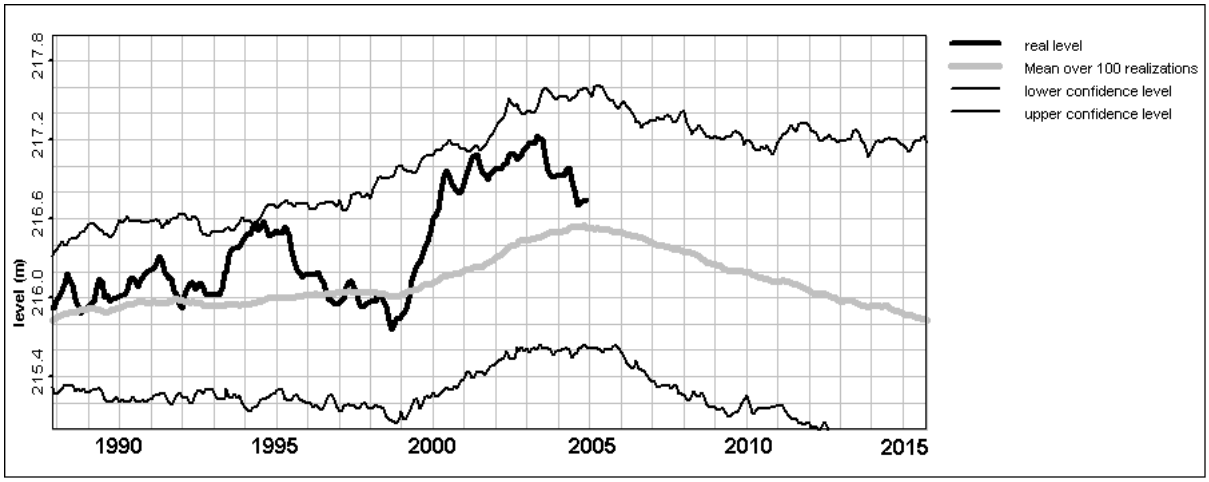


Figure 1: *Prediction of the water level in the reservoir modeling stochastic component as a independent normally distributed random values*

More likely is to consider the stochastic process φ as one with some hidden periodicity. Thus to the deterministic term (2) is added a periodic term containing 23 components:

$$\Psi_{per}(t) = \sum_{j=1}^{23}\left[A_j \cos(2\pi\nu_j t) + B_j \sin(2\pi\nu_j t)\right],$$

where ν_j are frequencies of the hidden periodicity. The significant frequencies are selected by means of a power spectrum after fast Fourier transform. Coefficients for a given frequency are estimated by solving linear systems constructed so as to minimize the variance of the estimate. The stochastic residuals again are treated as an independent normally distributed random values with distribution parameters: mean = -0.000037, variance = 0.00072.

The other possible approach to deal with the stochastic process φ is to take into the account weather parameters. It is known that such parameters as precipitation and evaporation are an important part of the standard water balance equation. Also precipitation affects the filtering component of the water balance. The influence of weather components is not always instantaneous (within the month), for example winter precipitation influences water level during the spring period.

The process φ presents periodic features, the same features are found in the auto-correlation and cross-correlation with a value characterizing an amount of water in the month (precipitation –

evaporation). But for close temporal distances this correlations can be modeled by spherical type of variogram (cross-variogram) with the following parameters: variogram sill = 0.0022, variogram range = 3.6; cross-variogram: sill = 0.0012, cross-variogram range = 4.5.

So, the stochastic component can be obtained as a geostatistical stochastic simulation with parameters for simulation estimated by co-kriging:

$$\varphi^*_i = \sum_{k=1}^{3} \lambda_k \varphi_{i-k} + \sum_{j=1}^{4} \alpha_j \left(P_{i-j} - E_{i-j}\right), \quad \sigma^*_i = \sum_{k=1}^{3} \lambda_k \gamma(k) + \sum_{j=1}^{4} \alpha_j g(j) + \mu_\varphi ,$$

where P means precipitation, E means evaporation, γ is a model of variogram and g is a model of cross-variogram. Coefficients ($\lambda_k : k = 1,2.3; \alpha_j : j = 1,2,3,4; \mu_\varphi$) are obtained by solving corresponding system of equations. Stochastic simulations are performed using direct approach proposed in Soares, 2001.

The geostatistically based approach looks more sophisticated as for predictions it requires additional information on unknown weather parameters. But from the other point of view it allows considering different weather scenario, for example a consequence of high water years.

Weather components can be introduced in the method with periodic component too. The residual ξ from formula (3) appears still to store some correlation with water amount. But here the geostatistical approach is not helpful because residuals $\xi(t)$ are not correlated. The correlation can be incorporated directly as a part of the deterministic component:

$$\varphi_t = \Psi_{\text{det}}(\varphi_{t-1}) + \Xi_{\text{det}}(P_t - E_t) + \Psi_{per}(t-1) + \zeta_{t-1} .$$

All proposed in this work approaches allow to obtain different kinds of estimates, including probabilistic estimates. The detailed comparison of obtained results will be discussed.

References

Mokrov Ju.G., Antonova T.A., Antropova E.V. (2009) Probabilistic predictions for dynamics of the level at the reservoir V-11 within different types of waterdecreasing techniques. The problems of Radioecological Safety, 1: 23-34, (in Russian)

Rumyantsev V.A., Trapeznikov Ju.A.(2008) Stochastic models of the hydrologic processes. Nauka, St.Petersburg, 152 p. (in Russian)

Soares A. (2001) Direct Sequential Simulation and Cosimulation. Mathematical Geology, 33: 911-926

CHARACTERIZATION OF NITRATE LONG TERM TRENDS AT REGIONAL SCALE USING STATISTICAL AND GEOSTATISTICAL TOOLS. EXAMPLE OF THE LOIRE-BRITTANY BASIN.

B. Bourgine[1], B. Lopez[1], N. Baran[1], D. Ratheau[2]

[1]Bureau de Recherche Géologique et Minière (BRGM), B.P. 6009 - 45060 Orléans Cedex 2, France Tél. : 33 2 38 64 32 02 - Fax : 33 2 38 64 33 99 – email : b.bourgine@brgm.fr
[2]Agence de l'Eau Loire Bretagne

Introduction

Europeans Framework Directives (2000/60/CE; 2006/118/CE; 91/676/EEC) require the Member States of European Union to take actions to achieve a "good" chemical and quantitative status of their groundwater resources by 2015. More specifically, Europeans countries have to determine a nitrate (NO_3) threshold and to identify ground waters where this threshold is exceeded or will be exceeded if no actions are undertaken (for instance France lowered down the NO_3 threshold to 40 mg/L). In case of failure to meet these requirements, the concerned countries will have to give explanations.

Together with the Loire-Brittany Water Agency, the BRGM conducted an exploratory study on the whole Loire-Brittany basin (157 000 km²). The main objective of this study was to collect information about the age of the water and to characterise past evolutions of the quality of the ground waters, in order to understand the current state of contamination and to predict (model) the possible evolution of the water quality.

This work involved (i) the determination of water resources age, (ii) collection of nitrate and water table level time series, the homogenisation and the verification of these data, (iii) the use of statistical and geostatistical tools to determine the state of water resources and the spatial or temporal evolutions, (iv) the definition of the typical behaviour and evolution, by comparison of piezometric and NO_3 signals with available data such as hydrogeological context and agricultural practices, (v) the comparison and the explanation of the whole set of information and data available. A quite similar work is currently conducted on the Seine-Normandy basin together with the Seine-Normandy Water Agency. This paper focuses on the statistical and geostatistical aspects, explaining the data analyses and the obtained results.

Used data

1,371,655 water table level data spread across 511 time series and 117,056 figures for NO3 distributed on 7,341 times series, collected between 1945 and 2007, were used in this study. Those data came from ADES and ONQES data bases managed by the BRGM or from the DDAS (local health agencies). Moreover, agricultural pressures of the year 2000 and 1/1,000,000 geological maps were used to achieve the study. A limitation of the study was that for a given

point it was not possible to have information for both the nitrate time series and the water table level time series.

Geostatistical analysis

From a geostatistical point of view, variograms of water table level and nitrate concentrations were calculated automatically on each time series, when the length of the time series was long enough. Regarding the piezometric data, variograms were calculated at 2 weeks time scale in order to reduce the sampling bias linked to heterogeneities in the measurement time interval. For the NO_3 concentrations, a 6 months time scale was applied because of the low frequency of the measurements. 1,483 experimental variograms (414 for piezometric data and 1,069 for the NO_3 data) were automatically fitted. Fitted models were composed of a combination of specific elementary components chosen for their ability to describe the behavior of the signal for the characteristic time scales: pepitic, short term and annual component of the signal as well as pluriannual trend. Consecutively, a hierarchical ascending classification (HAC) was conducted on the parameters of the fitted models. Thanks to this HAC, zones with a homogeneous nitrate and water table variation behaviour were defined. 90 zones resulted of the combination of the HAC with the agricultural pressure and the geological data. Correlations between NO_3 content and piezometric variations were studied for the spatially linked time series. We highlighted some positive correlations (with or without temporal gap) but also some not correlated signals.

Statistical tests to determine trends analyses

Trends were determined based on nitrate concentration time-series. Only waters sources for which there were at least 10 measurements and where there was not evidence of a natural denitrification (indication of a captive aquifer) were considered for these analyses. Therefore, 2,975 time series were selected. The non-parametric Mann-Kendall (MK) test, a robust statistical trend detection test, was used to estimate the trends of past evolutions in nitrate concentrations. The trend analyses were partitioned, since 1970, by decades in order to detect possible trend reversals along the studied period. A global trend was also determined for the 1970-2007 period. When the confidence level of the test was at least 95%, the slope of the trend was calculated (for this condition 645 points were studied). Following the Water Framework Directives, trends were also spatially represented using the Kendall Regional (KR) test. 31 zones (over 90) were so characterized. The KR test consists in the creation of a virtual well defined as the grouping of all the wells contained in the same zone. This exercise was conducted based on the annual averages of the Nitrate concentrations.

Preliminary results

Although the study is still ongoing, preliminary results indicate that trends evolution in nitrate concentration are directly depending upon the time and the geographical location. For some given zones, there is a clear trend reversal (decrease in nitrate concentration after a period of

increment) whereas in other zones there is a continuous increase. Globally, the highest nitrate concentrations are observed in relatively recent waters. Unfortunately, there are no very recent waters to establish if recent changes in agricultural practices have induced a decrease in nitrate concentration. In the case of trend evolutions showing seasonal variations, there is a correlation with the water table levels, associated or not with temporal gap. Given that there were no available maps of temporal evolution of the agricultural impact, it has not been possible to separate nitrate signal associated with the piezometric variations to the one associated with the agricultural practices. In this study a new methodology was assessed for the characterization of the trend evolution of the water resources at the scale of a large basin. This study also revealed several limitations: poor quality of some data, lack of accurate information on the aquifer and its state (captive or not), low frequency of nitrate measurements, bias induced by censoring (the fact that very often wells are closed and no more monitored when nitrate concentration exceeds a given threshold). The high volume of the data to be treated led us to develop several automatic procedures for statistical or geostatistical calculations. However, we still need to develop complementary tools to facilitate the work. For example the calculation and the automatic fitting of the cross covariance between nitrate and piezometric signals, or the automatic determination of the pairs of NO_3/piezometric wells that can be spatially linked. Beyond the characterization aspect, it remains to analyze and to explain the causes of the nitrate evolutions : a decrease of the nitrate concentrations observed in certain points can be explained either by a decrease of the water table level or by a previously decrease of the agricultural pressures or by a combination of these factors. The knowledge of the times and the modalities of the nitrate transfer from the unsaturated zone towards ground waters is a poorly known key mechanism which highly needs to be investigated and understood.

References

Aguilar J.B., Orban P., Dassargues A., Brouyère S., (2007) – Identification of groundwater quality trends in a chalk aquifer threatened by intensive agriculture in Belgium. Hydrogeology journal 15: 1615-1627.

Baran N., Gourcy L., Lopez B., Bourgine B., Mardhel V., (2009) – Transfert des nitrates à l'échelle du bassin Loire-Bretagne. Phase 1 : temps de transfert et typologie des aquifères. Rapport BRGM RP-54884-FR, 105 p.

Broers H.P., van der Grift B., (2004) – Regional monitoring of temporal changes in groundwater quality. Journal of hydrology 296: 192-220.

Frans L.M., Helsel D.R. (2005) – Evaluating regional trends in ground water nitrate concentrations of the Columbia Basin Ground Water management Area, Washington. U.S. Geological Survey Scientific Investigations, Report 2005-5078, 7p.

Stuart M.E., Chilton P.J., Kiniiburgh D.G., Cooper D.M., (2007) – Screening for long-term trends in groundwater nitrate monitoring data. Quaterly Journal of Engineering Geology and Hydrogeology, 40: 361-376.

CONTRIBUTION OF GEOLOGICAL HETEROGENEITY CHRACTERIZATION TO UNCERTAINTY IN PREDICTING UNSATURATED ZONE FLOW

C. Cronkite-Ratcliff[1], G. Phelps[2], A. Boucher[3]

[1]Dept. of Geological and Environmental Sciences, Stanford University, CA, USA,
collincr@stanford.edu
[2]US Geological Survey, Menlo Park, CA, USA
[3]Dept. of Environmental Earth System Science, Stanford University, CA, USA

The unsaturated zone is a major control on groundwater recharge and the input of contaminants to aquifer systems. Therefore, understanding flow in the unsaturated zone has important implications for determining the quality and availability of groundwater resources. This is particularly true in arid regions such as the southwestern United States, where the unsaturated zone may reach hundreds of meters in thickness.

Wherever the unsaturated zone is geologically heterogeneous, spatial variation in the hydraulic properties of constituent soil and rock may significantly affect the pathways of subsurface flow and contaminant transport (Wolfsberg et al., 2004). In the thick unsaturated zones characteristic of arid regions, large depositional features may form extensive, connected units with similar hydraulic properties, potentially influencing flow over long distances (hundreds to thousands of meters). In these regions, characterizing the geologic heterogeneity of the unsaturated zone can be a critical step in determining the location, rate, and spatial distribution of aquifer recharge, as well as the extent of effects from groundwater contamination.

Because of the limitations and incompleteness of geological and geophysical data used to characterize subsurface flow environments, modeling geological heterogeneity necessarily introduces uncertainty into predictions of subsurface flow and contaminant transport. Hydrologic modelers often adopt a single deterministic model of the subsurface geology, neglecting the uncertainty that results from incomplete geological data.

In our approach, we characterize the uncertainty in subsurface flow and contaminant transport predictions that result from the stochastic modeling of the subsurface geology. We generate multiple realizations of heterogeneous geology to represent the possible spatial variation of geology in the subsurface. We then simulate flow and transport for each realization, and measure uncertainty by tabulating the rates of recharge and contaminant mass flux to the water table.

We apply our approach in Yucca Flat, formerly a site for the underground testing of nuclear weapons. Yucca Flat is a large, geologically complex basin with a heterogeneous unsaturated zone ranging from 150 m to 580 m in thickness (Stoller-Navarro, 2006). Our study region is a 5km east-west, 7km north-south, 760m deep volume in the north-central portion of Yucca Flat. Using a procedure developed by Boucher and Phelps (2009), a series of 25 realizations of

heterogeneous subsurface geology are generated using training images developed from 2D geologic maps combined with 1D drillhole data (see Figure 1). The variability in the flow and transport predictions between each realization are a result of spatial variations in the permeability, which is assigned according to the sedimentary facies type generated in each realization (Figure 2).

Flow and transport are simulated using a 3D numerical model developed using the SUTRA software (Voss 1984). This model simulates both saturated and unsaturated flow. Contaminant transport is modeled using a conservative, non-reactive, non-sorbing tracer released continuously from 94 point sources, each corresponding to the working point of an underground detonation site. All 94 point sources are located above the water table (Hale et al 1995).

Flow and transport simulations consistently show that contaminants are initially transported downwards through the unsaturated zone and are then transported laterally once reaching the water table. However, the volumes, rates, and pathways of fluid flow and contaminant transport vary between each realization. Even over periods as short as one year, the predicted boundary of the modeled contaminant plume in the unsaturated zone can vary by hundreds of meters in both the horizontal and vertical dimensions between individual simulations (Figure 3).

Contaminant transport predictions appear to be sensitive to the spatial variation in the permeability assigned to specific sedimentary facies types. Over longer periods of time, the spatial variation in permeability and other hydraulic properties may magnify differences in contaminant transport between individual simulations as transport becomes affected by large depositional features over longer distances. These results suggest that estimations of the contaminant plume boundary over longer periods of time (hundreds to thousands of years) may be affected significantly by uncertainty in the geological model.

A natural next step would be to incorporate spatial variability in other transport parameters, such as dispersivity and porosity. Further research is also needed to investigate flow and transport over longer periods of time for long-term risk assessment. Such an analysis would provide an uncertainty measurement that can be used to guide risk analysis and decision-making for remediation design, groundwater monitoring, water resources management, or the construction of high-hazard facilities.

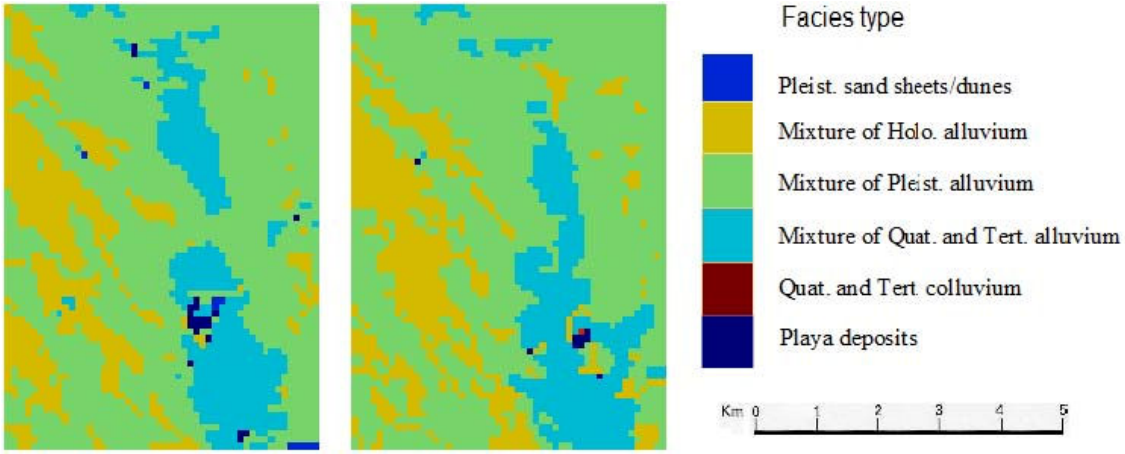


Figure 1: *Map of surface geology as generated in three separate realizations.*

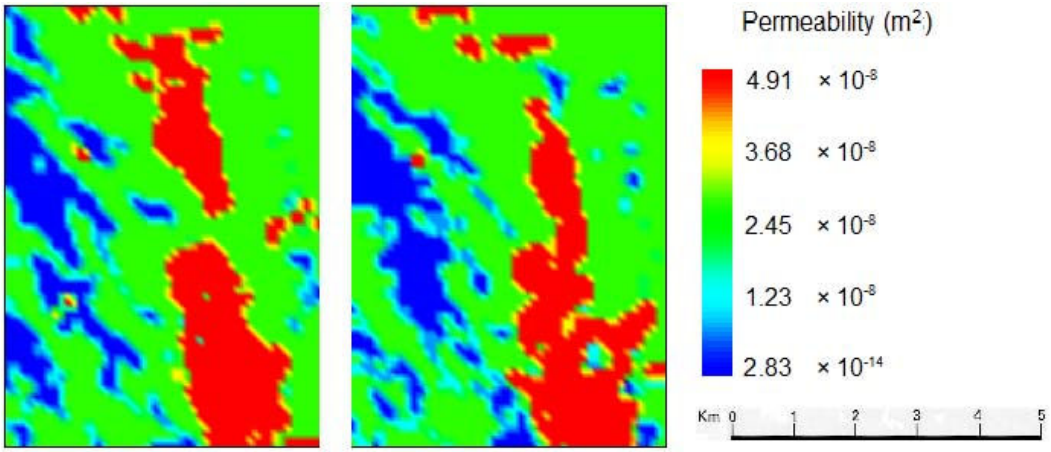


Figure 2: *Map of intrinsic permeability (Nimmo et al, 2009) at the surface assigned to the sedimentary facies type generated in three different realizations (corresponds with Figure 1).*

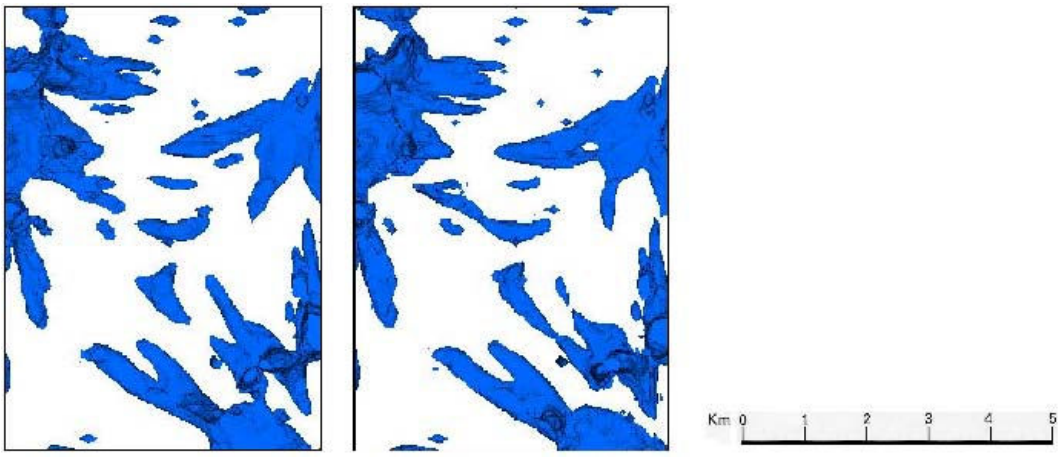


Figure 3: *View (from above) of modeled contaminant plumes from 94 continuously-releasing point sources within the study volume. Plumes (blue) represent regions where modeled contaminant concentration exceeds 4 ppb (concentration arbitrarily chosen to display the shape of the plumes).*

References

Boucher, A., and Phelps, G. Combining 2D geologic maps and 1D borehole information to simulate nonstationary 3D sedimentary basin. Conference on Computational Methods for the Earth, Energy and Environmental Sciences (Stanford, California, August 23-28, 2009), International Association of Mathematical Geology, 2009.

Hale, G.S., Trudeau, D.A., and Savard, C.S. Water-level data from wells and test holes through 1991 and potentiometric contours as of 1991 for Yucca Flat, Nevada Test Site, Nye County, Nevada. US Geological Survey Water Resources Investigation Report No. 95–4177, Carson City, Nevada, 1995.

Nimmo, J.R., Perkins, K.S., Schmidt, K.M., Miller, D.M., Stock, J.D., Singha, K., Andraski, B.J., and Munoz-Carpena, R. Hydrologic characterization of Desert soils with varying degrees of pedogenesis; 1, Field experiments evaluating plant-relevant soil water behavior: Vadose Zone Journal, v. 8, no. 2, p. 480-495, 2009.

Nimmo, J.R., Schmidt, K.M., Perkins, K.S., and Stock, J.D. Rapid measurement of field-saturated hydraulic conductivity for areal characterization: Vadose Zone Journal, v. 8, no. 1, p. 142-149, 2009.

Stoller-Navarro Joint Venture. Phase I Hydrologic Data for the Groundwater Flow and Contaminant Transport Model of Corrective Action Unit 97: Yucca Flat/Climax Mine, Nevada Test Site, Nye County, Nevada, Rev. 0, S-N/99205--077. Las Vegas, NV, 2006.

Voss, C.I. A finite-element simulation model for saturated-unsaturated, fluid-density-dependent ground-water flow with energy transport or chemically-reactive single-species solute transport. US Geological Survey Water Resources Investigation Report No. 84–4369, Reston, Virginia, 1984.

Wolfsberg, A.V., Tompson. A.F.B., Waddell, R.K., Hassan, A.E., Rehfeldt, K.R., and Russell, C.E. Heterogeneity and scaling: Technical basis document for Nevada Test Site Underground Test Area projects. Los Alamos National Laboratory report LA-14161, Los Alamos, New Mexico, 2004.

EFFECTS OF NON-GAUSSIAN SPATIAL DEPENDENCE OF HYDRAULIC CONDUCTIVITY ON HYDRODYNAMIC MACRODISPERSION

C. P. Haslauer [1], P. Guthke [1], A. Bárdossy [1], E. A. Sudicky [2]

[1]Universität Stuttgart, Institute of Hydraulic Engineering, Department of Hydrology and Geohydrology, Stuttgart, Germany, clau.haslauer@iws.uni-stuttgart.de

[2] University of Waterloo, Department of Earth and Environmental Sciences, Waterloo, Ontario, Canada

Hydraulic conductivity (K) is a fundamental parameter that influences groundwater flow and solute transport. The spatial distribution of K impacts the groundwater velocity field and hence directly influences the advective spreading of a solute migrating in the subsurface. Portions of the plume advance more rapidly than the average velocity while in other zones, migration rates are slower than the average velocity. This spreading phenomenon is commonly referred to as hydrodynamic macrodispersion.

The objective of the presented work is to use copulas as a novel non-Gaussian stochastic model to simulate spatially-correlated random fields of K that fit to real- world K data. The spatially distributed non-Gaussian K-fields are subsequently used to conduct a series of numerical tracer experiments using a high-resolution groundwater flow and contaminant transport model (HydroGeoSphere). Flow and transport characteristics are derived, particularly the rate of change of the spatial moments of the evolving contaminant plume. These characteristics are compared with those obtained under the assumption that the underlying spatial distribution of K has a Gaussian dependence structure, a commonly made assumption.

Both types of spatial K-fields were constrained to have the same marginal distribution, the same variogram, the same covariogram, and the same rank correlations (Figure 1). However, both types of K-fields do exhibit a different spatial dependence structure when modeled by copulas, and thus lead to a significantly different transport behavior. The dispersion tensor, which is proportional to the rate of change of the 2^{nd}-order spatial moments of the evolving contaminant plume, is different for the Gaussian and non-Gaussian descriptions of spatial dependence (Figure 2).

The outlined theory is applied to three-dimensional statistically anisotropic K-data obtained from two of the most extensively studied aquifer test-sites. Each site comprises ~1400 samples taken along two cross-sections. One site is the Borden, Ontario, aquifer, a modestly heterogeneous aquifer. The other site, located near North Bay, Ontario, is comprised of a highly heterogeneous glacial deposit.

In both of these differently deposited environments, non-Gaussian dependence structures of K and hence non-Gaussian transport characteristics have been found.

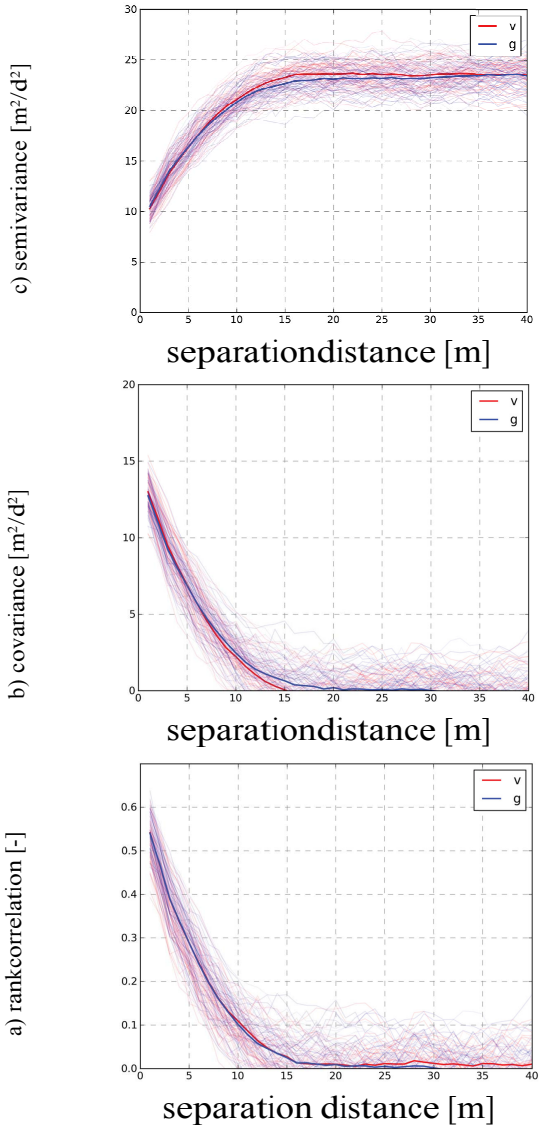


Figure 1: *Geostatistical measures for 100 realizations of simulated K- fields. Thick lines represent arithmetic averages. Blue indicates a gaussian model, red indicates a v- transformed Copula model.*

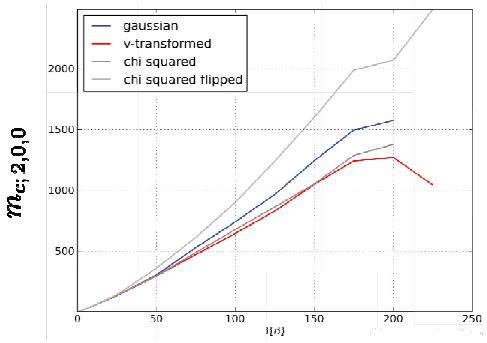


Figure 2: *Averages of central second spatial moments over time for four different types of spatially distributed K-fields based on 100 simulations.*

Affiliations

[1] Universität Stuttgart
Institute of Hydraulic Engineering
Department of Hydrology and Geohydrology
Pfaffenwaldring 61
70569 Stuttgart
Germany

corresponding author: Claus P. Haslauer,
PhD student
telephone: +49-711-68562473
email: clau.haslauer@iws.uni-stuttgart.de

Philipp Guthke
M.Sc. student
philipp_guthke@hotmail.com

András Bárdossy
Professor
Andra.Bardossy@iws.uni-stuttgart.de

[2] University of Waterloo
Department of Earth and Environmental Sciences
200 University Avenue W Waterloo, Ontario, N2L3G1
Canada

Edwin A. Sudicky Professor sudicky@sciborg.uwaterloo.ca

A COMPARATIVE STUDY OF THREE-DIMENSIONAL HYDRAULIC CONDUCTIVITY UPSCALING AT THE MACRODISPERSION EXPERIMENT (MADE) SITE, ON COLUMBUS AIR FORCE BASE IN MISSISSIPPI (USA)

L. Li, H. Zhou, J. J. Gómez-Hernández

Group of Hydrogeology, Universidad Politecnica de Valencia, Spain,
liali@upvnet.upv.es, Tel. (34) 963879615, Fax. (34) 963879492

Simple average, simple Laplacian, Laplacian with skin and nonuniform coarsening upscaling techniques are investigated using, as a reference, a fine scale realization of the hydraulic conductivities at the Macrodispersion Experiment site on Columbus Air Force Base in Mississippi (USA). This realization was generated using a hole effect variogram model and it was shown that flow and transport modeling in this realization is capable to reproduce the observed non-Gaussian spreading of tritium plume. The purpose of this work is twofold, first to compare the effectiveness of different upscaling techniques in yielding upscaled models able to reproduce the observed transport behavior, and second to demonstrate that careful upscaling of the flow model can provide a coarse model in which the standard advection dispersion equation can be used to model transport in seemingly non-Fickian scenarios. Specifically, the use of the ``Laplacian with skin" full tensor upscaling technique coupled with a nonuniform coarsening scheme yields the best results both in terms of flow and transport reproduction.

According to the previous study (Salamon et al., 2007), their realization #26 out of 40 realizations exhibits a similarly anomalous plume as the one observed on-site at 328 days. This realization is selected as the reference field in the following upscaling analysis.

Horizontal, depth integrated concentration distributions as well as vertical, laterally integrated concentration distributions of realization #26 are illustrated in Fig 2, 3 and 4 using different upscaling techniques. The concentration distribution of Fig 4 using nonuniform coarsening interblock Laplacian with skin approach is in qualitative good agreement with the fine scale in Fig 1, whereas the distribution of Fig 2 and Fig 3 using block and interblock simple Laplacian uniform coarsening display an underestimated tailing of solute spreading resulting from poor reproduction of the fine scale flow. Particularly, the use of block simple Laplacian method here significantly underestimates the contaminant tailing of solute plume.

References:

Salamon, P., Fernandez-Garcia, D., Gómez-Hernández, J. J., 2007. Modeling tracer transport at the MADE site: the importance of heterogeneity. Water Resour. Res. 30 (8).

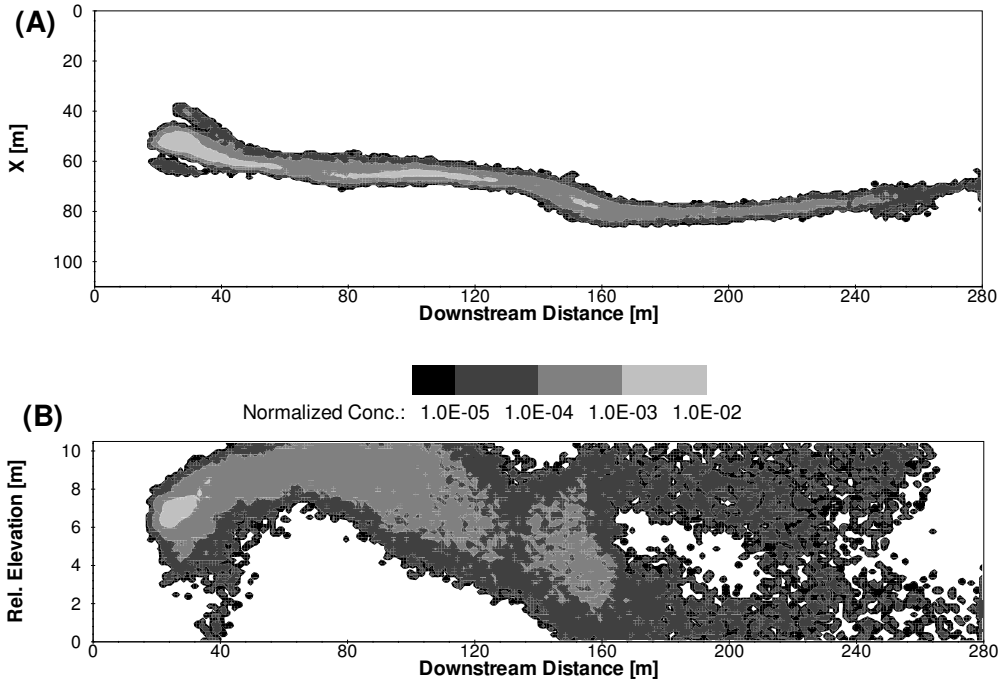


Figure 1: *(A) Depth integrated normalized concentration distribution after 328 days for realization #26 in the fine scale.(B) Laterally integrated normalized concentration distribution after 328 days for realization #26 in the fine scale.*

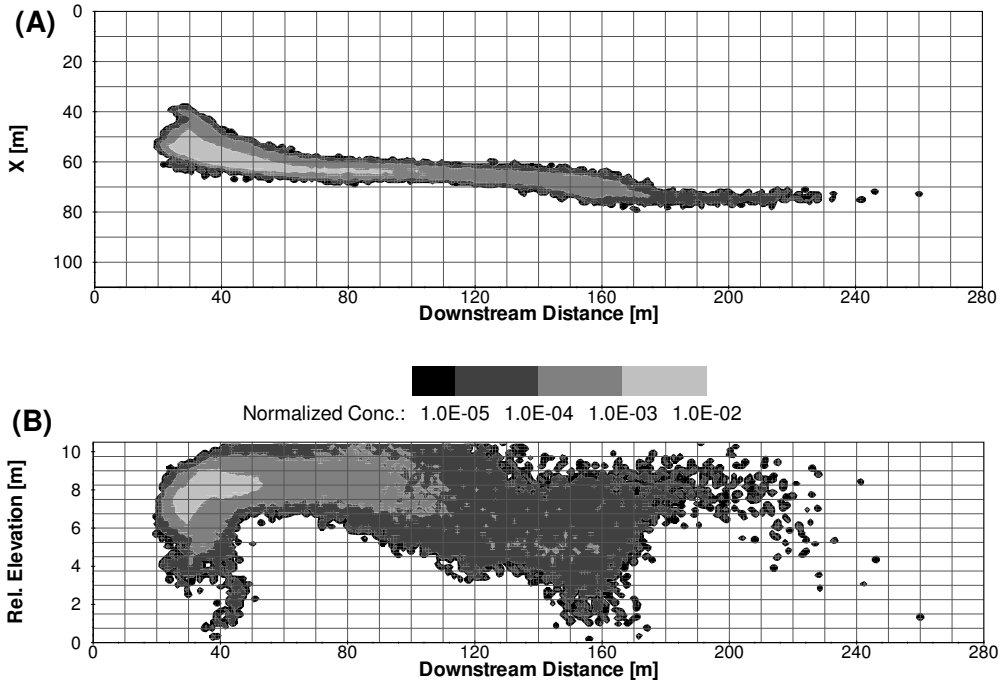


Figure 2: *(A) Depth integrated normalized concentration distribution after 328 days for realization #26 using uniform coarsening of block simple Laplacian upscaling.(B) Laterally integrated normalized concentration distribution after 328 days for realization #26 using uniform coarsening of simple Laplacian upscaling.*

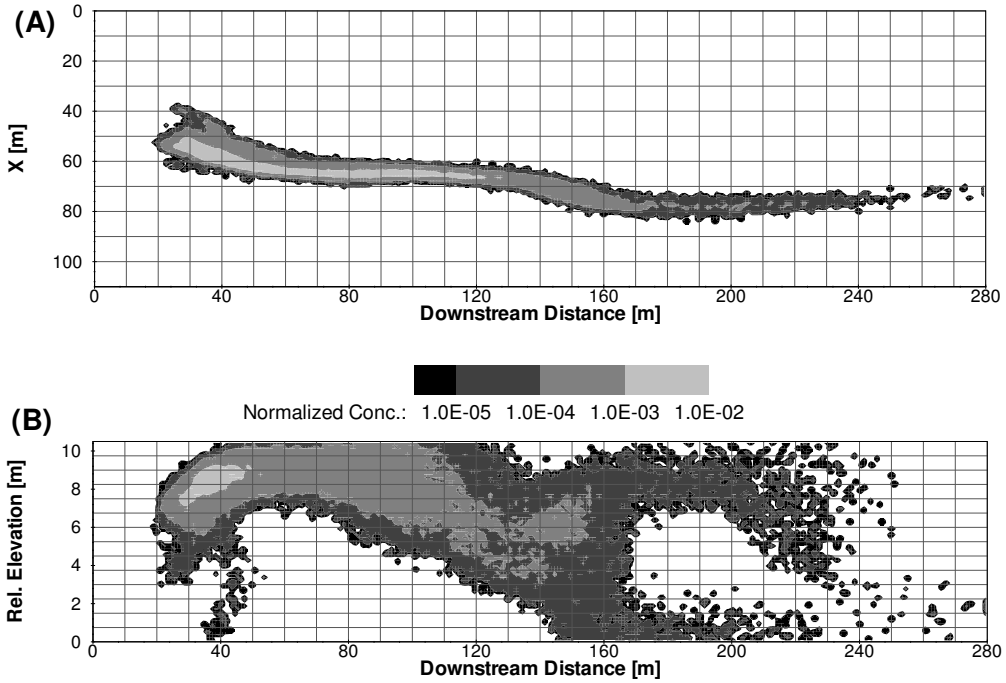


Figure 3: *(A) Depth integrated normalized concentration distribution after 328 days for realization #26 using uniform coarsening of interblock simple Laplacian upscaling. (B) Laterally integrated normalized concentration distribution after 328 days for realization #26 using uniform coarsening of simple Laplacian upscaling.*

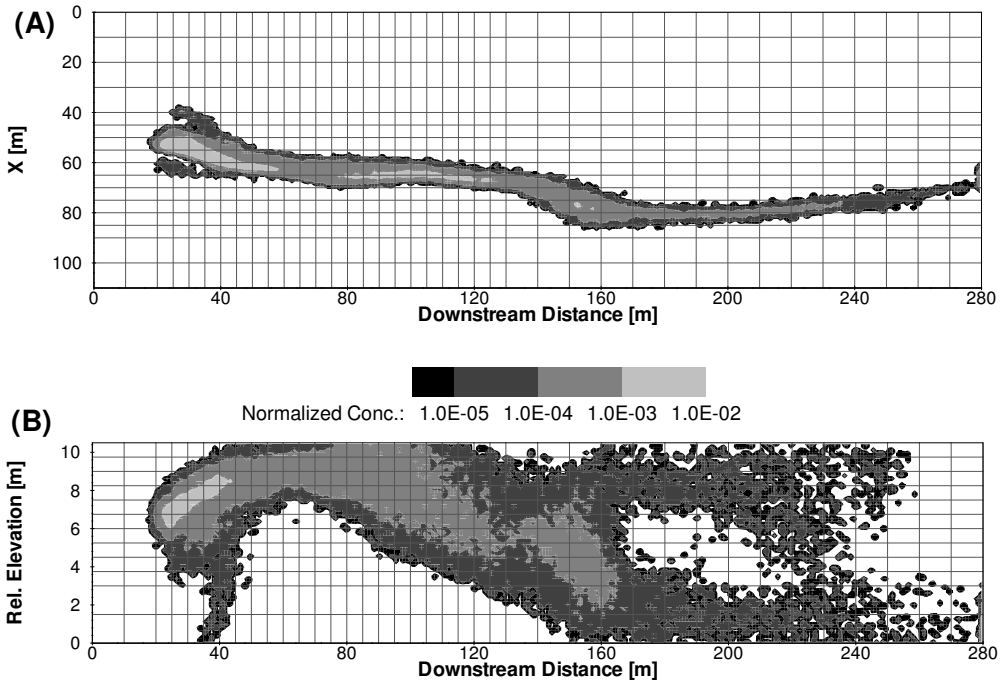


Figure 4*: (A) Depth integrated normalized concentration distribution after 328 days for realization #26 using nonuniform coarsening of Laplacian with skin upscaling.(B) Laterally integrated normalized concentration distribution after 328 days for realization #26 using nonuniform coarsening of Laplacian with skin upscaling.*

GEOSTATISTICAL ANALYSIS OF PRIMARY AND SECONDARY DATA IN A SANDY AQUIFER AT MOL/DESSEL, BELGIUM

B. Rogiers[1,2], D. Mallants[1], O. Batelaan[2,3],M. Gedeon[1], M. Huysmans[2] & A. Dassargues[2,4]

[1]Institute for Environment, Health and Safety, Belgian Nuclear Research Centre (SCK•CEN), Boeretang 200, 2400 Mol, Belgium, brogiers@sckcen.be.

[2]Dept. of Earth and Environmental Sciences, K.U.Leuven, Celestijnenlaan 200e - bus 2410, 3001 Heverlee, Belgium.

[3]Dept. of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium.

[4]Dept. of Architecture, Geology, Environment and Civil Engineering (ArGEnCo), Université de Liège, B.52/3 Sart-Tilman, 4000 Liège, Belgium.

Abstract

In the framework of the disposal of short-lived low- and intermediate-level radioactive waste in a near-surface disposal facility in Dessel, Belgium, additional extensive site characterization has been performed in 2008. The gathered data now enclose 388 hydraulic conductivity measurements on samples of 8 cored boreholes. Secondary information as grain size analysis, porosity, and borehole geophysical parameters was also gathered. In addition, the geology of the study area has also been thoroughly characterized by a set of 178 cone penetration tests (CPTs) to approximate 50 m depth. This dataset allowed to refine the hydrostratigraphical model of the region. The existing groundwater model, based on large-scale effective hydraulic properties, was updated accordingly. The next step is a small-scale probabilistic approach 1) to validate the current existing deterministic groundwater models and 2) to support design for a monitoring network. In preparation for stochastic realizations of the subsurface, a geostatistical analysis of the available primary and secondary data is performed.

Relationships between the borehole logs and the hydraulic conductivity measurements are examined. The considered logs are the caliper, gamma ray (GR), a focused electrolog (Res_FEL), and long- and short spacing resistivity logs (Res_64, Res_16). These logs are continuous, and allow to get some information about the hydraulic conductivity at the depths that were not sampled. This is especially interesting for studying the heterogeneity of the different hydrogeological units and calculating a representative hydraulic conductivity for a geological unit over its entire thickness. The correlation coefficients between the logs and the hydraulic conductivity depend strongly on the considered geological formation. For certain hydrogeological units, the caliper, gamma ray and resistivity logs show a correlation coefficient around 0.6 - 0.7 (Table 1).

Structural analysis was performed for all the hydrogeological units together. Figure 1 shows the variograms and cross-variograms that are calculated using Cressie's robust variogram estimation (Cressie, 1993) after standardization of each hydrogeological unit. Spherical models are fitted,

and a kriging and co-kriging cross-validation is performed. The method used is k-fold cross-validation with k ≈ n/2 (with n equal to the number of data points) in a way that the (almost) co-located horizontal and vertical K-measurements were always left out together (leave-two-out cross-validation). The correlation coefficient between observed and predicted values from the cross-validation increases from 0.68 for kriging to 0.77 for co-kriging with all the available secondary information.

Table 1: *Correlation coefficients between logarithmic hydraulic conductivity and secondary parameters for the different hydrogeological units. (r: correlation coefficient, n: number of points)*

Secondary parameters	Hydrogeological units								Total	
	Quaternary		Mol (Upper)		Mol (Lower)		Kasterlee			
	r	n	r	n	r	n	r	n	r	n
Caliper	*-0.27*	*12*	*0.77*	*39*	*0.22*	*68*	*0.44*	*25*	**0.42**	**144**
Res_FEL	*0.04*	*13*	*0.45*	*39*	*0.09*	*68*	*0.61*	*25*	**0.28**	**145**
GR_cps	*0.41*	*12*	*-0.59*	*37*	*-0.13*	*68*	*-0.47*	*25*	**-0.42**	**142**
Res_16	-	-	*0.67*	*28*	*0.05*	*68*	*0.44*	*25*	**0.34**	**123**
Res_64	-	-	*0.38*	*30*	*-0.33*	*68*	*-0.32*	*25*	**0.10**	**125**

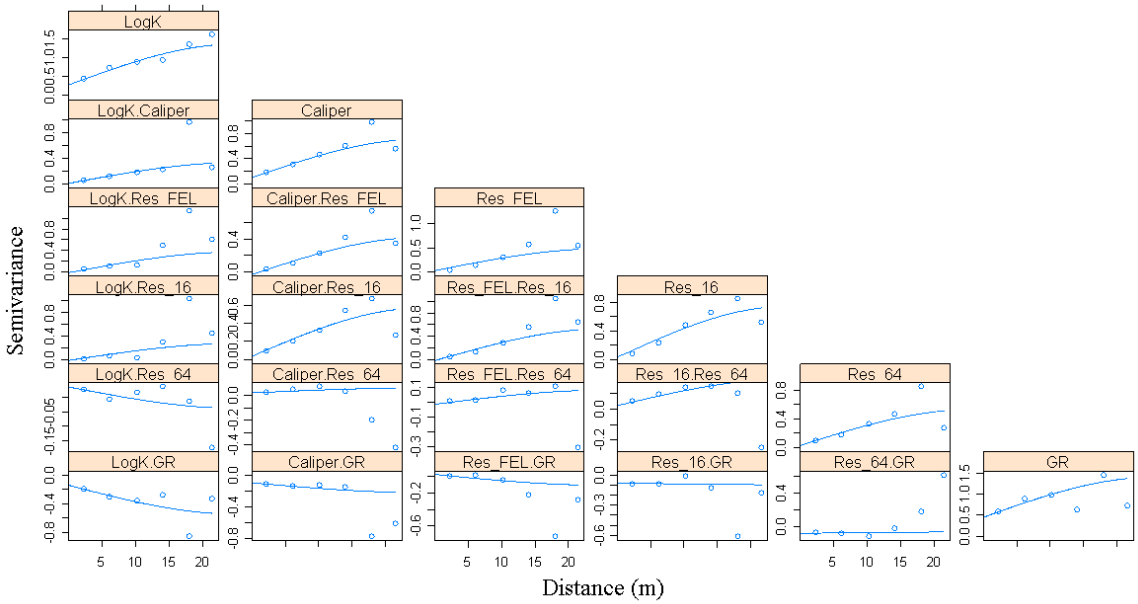


Figure 1: *Experimental variograms and cross-variograms of the hydraulic conductivity and the different borehole log parameters and their fitted spherical models*.

Subsequently, co-kriging is used to predict continuous vertical profiles of the hydraulic conductivity. This allows to detect small contrasting layers which might have been overlooked by the discrete sampling of the cores, and are hence not represented in the hydraulic conductivity measurements. The presence of contrasting layers is checked by visual inspection of the borehole

cores. Effective K-values are finally calculated for each hydrostratigraphical layer from these continuous profiles, and compared with the values obtained from the discrete sampling.

Another source of information is the CPT campaign. This data covers the entire area of interest, and would hence be a very usefull source of secondary information. The use of CPTs in predicting hydraulic conductivity has already been demonstrated for instance by Tillmann et al. (2008). They used several CPT parameters (including gamma activity, bulk and matrix density and water content) to predict grain size distributions and calculate the hydraulic conductivity from those. The parameters recorded in the current CPT campaign comprise the point resistance and the side resistance, which only allows to search for an empirical relationship with hydrogeologic parameters. The proportion of point to side resistance is called the friction ratio, and is commonly used to make a distinction between different lithological units. Because of the mechanical character of these measurements, one might indeed expect a relationship with the hydraulic conductivity. Unfortunately, the CPTs and boreholes are not co-located. The horizontal distance between the two sources of information is between 7 m and 350 m. Despite this drawback, it is however still possible to find a relationship. The point resistance shows the highest correlation with the logarithmic hydraulic conductivity. The fitted function has an R-squared of 0.73 (Figure 2, left). The demonstrated relationship is however only valid over a broad range of hydraulic conductivity values (-7.5 to -4.5; 3 orders of magnitude). This is probably due to the distance between the CPTs and boreholes. Additional CPTs close to the boreholes are necessary to be able to infer small-scale changes. A horizontal pseudo cross semivariogram (as defined in Goovaerts, 1997) is finally calculated for the boreholes and their neighbouring CPTs (Figure 2, right).

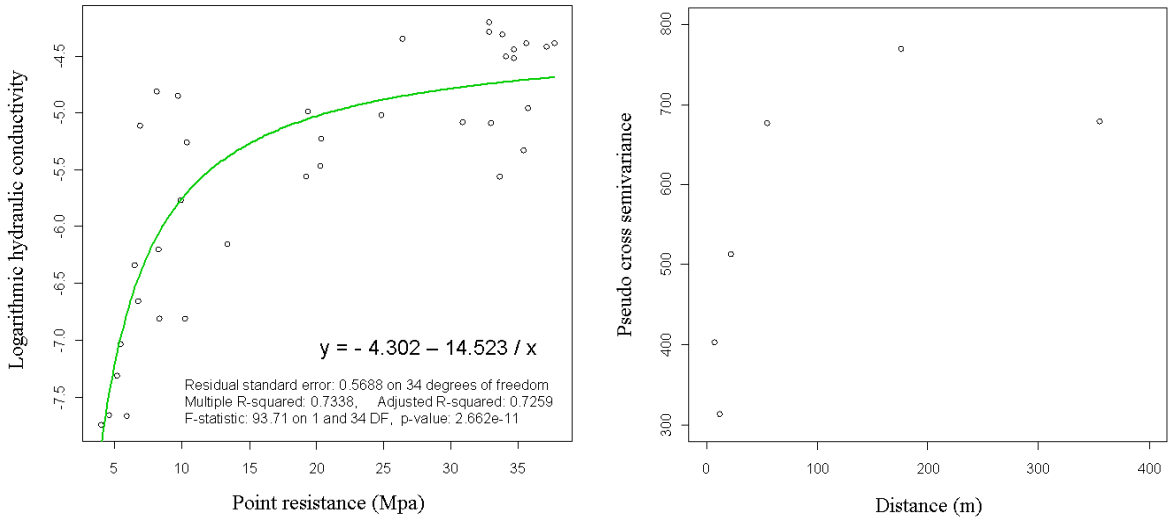


Figure 2: *CPT point resistance and logarithm of the hydraulic conductivity. Left: Scatter plot and fit. Right: Pseudo cross semivariogram*.

Currently, only vertical information is gathered based on the vertical data profiles. Additional sampling and measurements will be performed in the future to provide information of the horizontal spatial variance and to optimize the use of secondary data. Together with the current results, these will serve as the basis for conditional stochastic simulation of groundwater flow and contaminant transport.

References

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

Cressie N (1993) Statistics for Spatial Data. Wiley.

Tillmann A, Englert A, Nyari Z, Fejes I, Vanderborght J, Vereecken H (2008) Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the Krauthausen test site using Cone Penetration Test. Journal of contaminant hydrology 95(1-2): 57-75.

IMPROVED MAPPING OF DAILY PRECIPITATION OVER QUEBEC USING THE MOVING-GEOSTATISTICS APPROACH

N. Jeannée[1], D. Tapsoba[2]

[1] GEOVARIANCES, 49bis Av. Franklin Roosevelt, BP91, 77212 Avon, France, jeannee@geovariances.com, +33 (0)1 60 74 74 54.
[2] HYDRO-QUEBEC, Research Institute (IREQ), 1800 boul. Lionel-Boulet, Varennes, QC, J3X 1S1, Canada, tapsoba.dominique@ireq.ca, +1 450-652-8588.

Abstract

Daily precipitation is a key parameter for predicting hydropower generation in Canada. Several issues commonly make difficult the estimation of precipitation: scarce moniroting networks associated with high spatial variability. Though external factors such as a Digital Terrain Model may be useful when predicting average precipitation over a greater period of time (month, year), these factors become useless in the case of daily precipitation (Goovaerts, 2000). In the Quebec case, the presence of local anisotropies and clear large scale non stationarities are key issues when dealing with daily precipitation datasets. Applying classical geostatistical methods results in difficulties to capture the precipitation spatial continuity and its actual variability over the region. To overcome these issues, an attempt has been made to apply the Moving-GeoStatistics (M-GS) framework, which is dedicated to the local optimization of parameters involved in variogram-based models (Magneron et al., 2008). M-GS considers the structural and computational parameters as a set of parameters to be spatially optimized. Obtained structural parameters are then taken into account during kriging and stochastic simulations, the latter being required to obtain global precipitation distributions over watersheds. The results are compared with the ones obtained using a classical universal kriging approach.

Material & Methods

The daily precipitation dataset comes from the Réseau météorologique coopératif du Québec (RMCQ). The RMCQ is a partnership initiative from the managers of Quebec meteorological networks: Rio Tinto Alcan, Environment Canada, Hydro-Québec, the Quebec Department of Sustainable Development, Environment and Parks (MDDEP), the Quebec Department of Natural Resources (MRN) and the Society of Protection of Forests against Fire (SOPFEU).This paper is focused on a specific day, the 18th of January 2009 and on two watersheds (see Fig. 1): “Rivière Mouchalagane” (North) and “Rivière Rouge” (South).

The Moving-GeoStatistics (M-GS) framework offers several ways to determine local structural parameters to be used during kriging or simulations. Among them, a local cross-validation approach has been applied in the present case, aiming at determining the optimal structural parameters within adjacent windows. An overlapping of the windows is usually recommended to

ensure continuity in the variation of the structural parameters. Once determined, the parameters can be re-interpolated on the final grid and directly used in subsequent estimation / simulation procedures. The efficiency of M-GS is tested by comparison with a classical universal Kriging (UK) (Chilès and Delfiner, 1999). The comparison is based on: visual comparison of results, local ability to predict precipitation values on a distinct validation subset, global estimates of precipitation and related confidence interval over two watersheds.

Results

142 precipitation values are available for the 18th of January 2009. As illustrated by Fig. 1, the spatial distribution of precipitation values is highly non stationary with a global decrease of values towards the North.

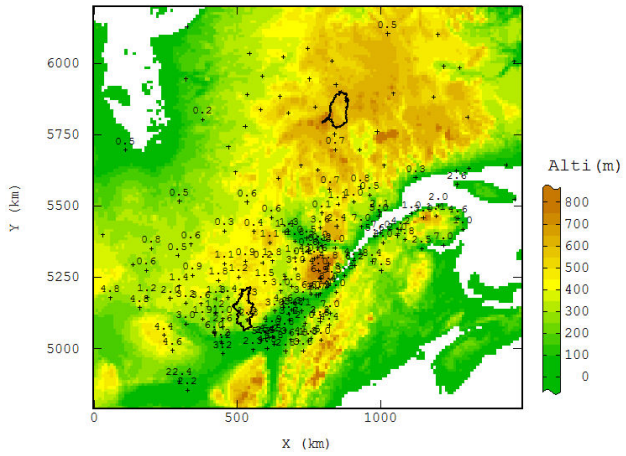


Figure **1***: Location and daily precipitation values for day 18/01/09. Crosses without values correspond to an absence of rain. Location of two watersheds of interest, display of Digital Terrain Model in background.*

A local determination of several parameters (M-parameters) has been performed (see Figure. 2): direction of anisotropy and ranges using the local cross-validation approach, sill through the computation of local variances within consistent windows. Kriging using these M-parameters is illustrated on Figure. 3 together with its associated standard deviation map; both maps provide more realistic estimates of the precipitation field compared to a classical UK approach, which is confirmed by statistical results obtained on a validation subset.

Average global estimates of precipitation over two watersheds are then estimated using stochastic simulations. Results derived from classical simulations provide unrealistic global precipitation estimates, particularly for the “Rivière Mouchalagane” watershed; this is much improved using M-parameters (see Table 2).

Perspectives

Applying Moving-Geostatistics leads to promising perspectives for mapping variables which present varying local anistropies and/or non stationarities. However, several theoretical aspects still need to be improved, including for instance the authorization conditions for making vary underlying M-parameters and the gaussian assumption underlying the application of stochastic simulations (word performed on raw values in the present cas).

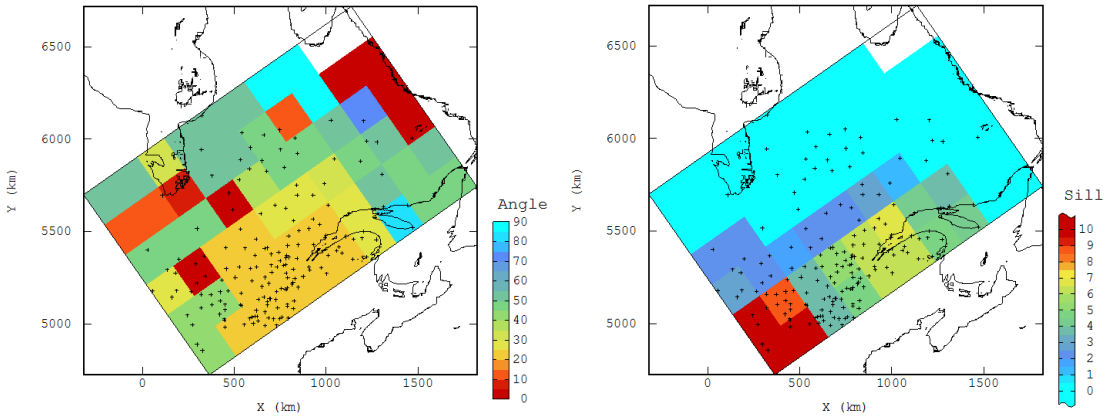


Figure 2: *Local determination of the direction of anisotropy (left) and the sill (right).*

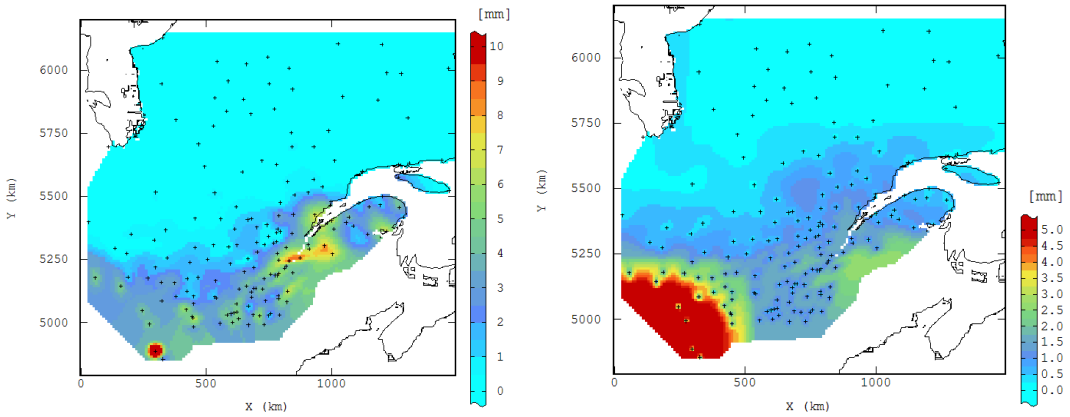


Figure 3: *Precipitation M-kriging (left) and standard deviation map (right) using M-parameters.*

Table 2: *Global precipitation estimates over two watersheds obtained using Turning Bands (with local trends) and SGS with M-parameters: bassin name, surface, accumulated precipitation estimates (Q5, average, Q95), standard deviation and coefficient of variation (CV).*

River bassin	Surface (km^2)	Simulation Approach	Accumulated precipitation (hm^3)				
			Q5	Average	Q95	Std Dev.	CV
Rivière Mouchalagane	5 945	TB	19	3 141	9 660	3 185	1.01
		M-GS	375	30 040	99 663	29 957	1.00
Rivière Rouge	5 619	TB	87 740	112 748	143 340	16 114	0.14
		M-GS	66 005	101 605	144 753	22 365	0.22

References

Chilès J.P. and Delfiner P. (1999) *Geostatistics: modelling spatial uncertainty*, Wiley Series in Probability and Mathematical Statistics, 695p.

Magneron C., Jeannée N., Le Moine O., Bourillet JF. (2008) Integrating prior knowledge and locally varying model parameters with M-GeoStatistics: methodology and application to bathymetry mapping. In *geoENV VII - Geostatistics for Environmental Applications*, eds. Renard, Demougeot-Renard & Froidevaux, Kluwer Academic Publishers, pp. 161-172.

Goovaerts, P. (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* 228, 113–129.

OBJECTIVE BAYESIAN ANALYSIS FOR THE CORRELATION PARAMETERS IN GAUSSIAN COPULA-BASED SPATIAL MODELS

H. Kazianka[1], J. Pilz[2]

[1] Technical University of Vienna, Austria, hannes.kazianka@tuwien.ac.at, Tel.: +43 1 58801 10525
[2] University of Klagenfurt, Austria, juergen.pilz@uni-klu.ac.at, Tel.: +43 463 2700 3113

In many works on Bayesian spatial analysis the prior for the correlation parameters is chosen to be vague (e.g. Banerjee et al. [1]), which does not reflec the absence of prior knowledge and may even be highly informative. Therefore, Berger et al. [2] developed default priors for unknown regression parameters, variance parameter and range parameter of the Gaussian random fiel model with isotropic covariance function. However, they did not take account of a possible nugget effect. De Oliveira [4] was the firs to calculate non-informative priors for the nugget effect but treated the range parameter as being known. In this paper we show that for the Bayesian spatial copula model (Kazianka and Pilz [5], Kazianka and Pilz [6]), the joint Jeffreys' prior for nugget and range given the marginal parameters $\boldsymbol{\eta}$ takes a simple form and serves as a default choice.

The correlation matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ of the Gaussian copula is parameterized by an isotropic correlation function model including a range parameter, here denoted by ϑ_1, and a nugget parameter, 1-ϑ_2. The following theorems state how Jeffreys' prior for $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2)$ given the parameters of the marginal distribution, $\boldsymbol{\eta}$, can be calculated and that it leads to a proper posterior distribution even though the propriety of the prior itself is not guaranteed. It turns out that the prior does not depend on $\boldsymbol{\eta}$ which supports the "approximate correlation invariance"-theory by De Oliveira [4]. Since the Gaussian spatial copula model includes the Gaussian random field the results are also valid in this case.

Theorem 1. *Denoting* $\boldsymbol{W}_i = \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{\theta}}}{\partial \vartheta_i} \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$*, i=1,2, Jeffreys' prior density for* $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2)$ *given* $\boldsymbol{\eta}$ *in the Gaussian spatial copula model can be written as*

$$p^J(\vartheta_1, \vartheta_2 \mid \boldsymbol{\eta}) \propto \left\{ \operatorname{tr}\left[\boldsymbol{W}_1^2\right] \operatorname{tr}\left[\boldsymbol{W}_2^2\right] - \left(\operatorname{tr}\left[\boldsymbol{W}_1 \boldsymbol{W}_2\right]\right)^2 \right\}^{\frac{1}{2}}. \tag{1}$$

Theorem 2. *If* $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (1-\vartheta_2)\,\boldsymbol{I} + \vartheta_2 \left(\mathbf{1}\mathbf{1}^T + \nu(\vartheta_1)\left(\boldsymbol{D} + o(1)\right)\right)$ *as* $\vartheta_1 \to \infty$*, where* $\nu(\vartheta_1) > 0$ *is differentiable and monotonically decreasing,* $\boldsymbol{D}$ *is a fixed nonsingular matrix and* $\mathbf{1}^T = (1, \ldots, 1)$*, then the posterior distribution of* $\boldsymbol{\theta} \mid \boldsymbol{\eta}$ *using Jeffreys' prior in the Gaussian spatial copula model is proper.*

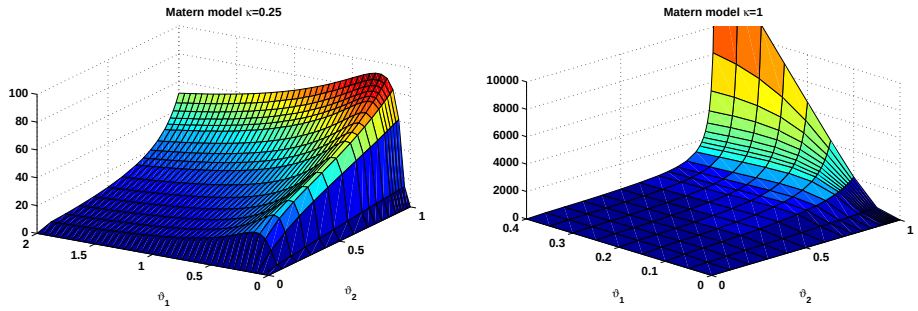
A joint prior for all the parameters can be determined by $p(\boldsymbol{\Theta}) = p^J(\boldsymbol{\theta} \mid \boldsymbol{\eta})\, p(\boldsymbol{\eta})$. If $p(\boldsymbol{\eta})$ is proper and $\boldsymbol{Z}$ denotes the data vector, the posterior $p(\boldsymbol{\Theta} \mid \boldsymbol{Z})$ is also proper.

Corollary 1. *Under the assumptions of Theorem 2* $p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta})$ *is not integrable.*

Corollary 2. *For the correlation function models discussed in Berger et al. [2] the behavior of* $p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta})$ *as* $\vartheta_1 \to \infty$ *is given by*

Model	$p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta})$	*Model*	$p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta})$
Matern:		*Rational quadratic*	$\mathcal{O}(\vartheta_1)$
$\kappa < 1$	$\mathcal{O}\left(\vartheta_1^{2\kappa-1}\right)$	*Power exponential*	$\mathcal{O}\left(\vartheta_1^{\nu-1}\right)$
$\kappa \geq 1$	$\mathcal{O}(\vartheta_1)$	*Spherical*	$\mathcal{O}(1)$

Generally, if $\Sigma_{\boldsymbol{\theta}}(h) = (1-\vartheta_2)\, I(h=0) + \vartheta_2 + c_1 \left(\frac{h}{\vartheta_1}\right)^2 + o(h^2)$ *as* $h \to 0$*, then* $\Sigma_{\boldsymbol{\theta}}$ *is at least two-times differentiable for* $\vartheta_2 = 1$ *and it holds that* $p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta}) \in \mathcal{O}(\vartheta_1)$.



(a) (b)

Figure 1: Jeffreys' prior $p^J(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ for a Matern model with (a) $\kappa = 0.25$ and (b) $\kappa = 1$. The 121 observation locations are placed on a regular grid inside the unit square.

Corollary 2 and Figure 1 clarify that the behavior of $p^J(\vartheta_1, \vartheta_2 = 1 \mid \boldsymbol{\eta})$ as $\vartheta_1 \to \infty$ strongly depends on the smoothness of the correlation function model. For differentiable random processes perfect correlation is a-priori most likely. For non-differentiable random processes, however, high values of ϑ_1 are equally likely a-priori (e.g. spherical or exponential model) or are even less likely than smaller values (e.g. Matern model with smoothness parameter $\kappa < 0.5$).
The Jeffreys' prior depends also on the sampling design. Priors for a regular, a random and a clustered sampling design with the same number of observations differ in their behavior towards $\vartheta_2 = 1$ and $\vartheta_1 = 0$. Therefore, care must be taken when $p^J(\vartheta_1, \vartheta_2 \mid \boldsymbol{\eta})$ is used for models that include geometric anisotropy. In this case the prior is conditional on the actual values of the anisotropy ratio a and the anisotropy angle φ that change the sampling design, $p^J(\vartheta_1, \vartheta_2 \mid \boldsymbol{\eta}, a, \varphi)$. This has to be considered at every step of the Metropolis-Hastings algorithm presented in Kazianka and Pilz [5].

To illustrate our methodology we analyze a data set which contains Cs137 values at 148 locations in the region of Gomel, Belarus. The data were observed in 1996, ten years after the Chernobyl accident. The observations are right-skewed which motivates us to use the log-normal distribution with parameters $\boldsymbol{\eta} = (\mu, \sigma^2)$ as a marginal distribution in the Gaussian copula-based spatial model. We treat μ and σ^2 as being independent a-priori and use a normal-inverse-gamma prior, $\mu \sim N(0.8, 1)$ and $\sigma^2 \sim IG(11, 30)$. Hyperparameters are chosen to provide conservative bounds for μ and σ^2. The prior for the exponential correlation function model is the Jeffreys' prior from Eq. (1). A uniform prior on $[1, 10] \times [0, \pi]$ is chosen for the

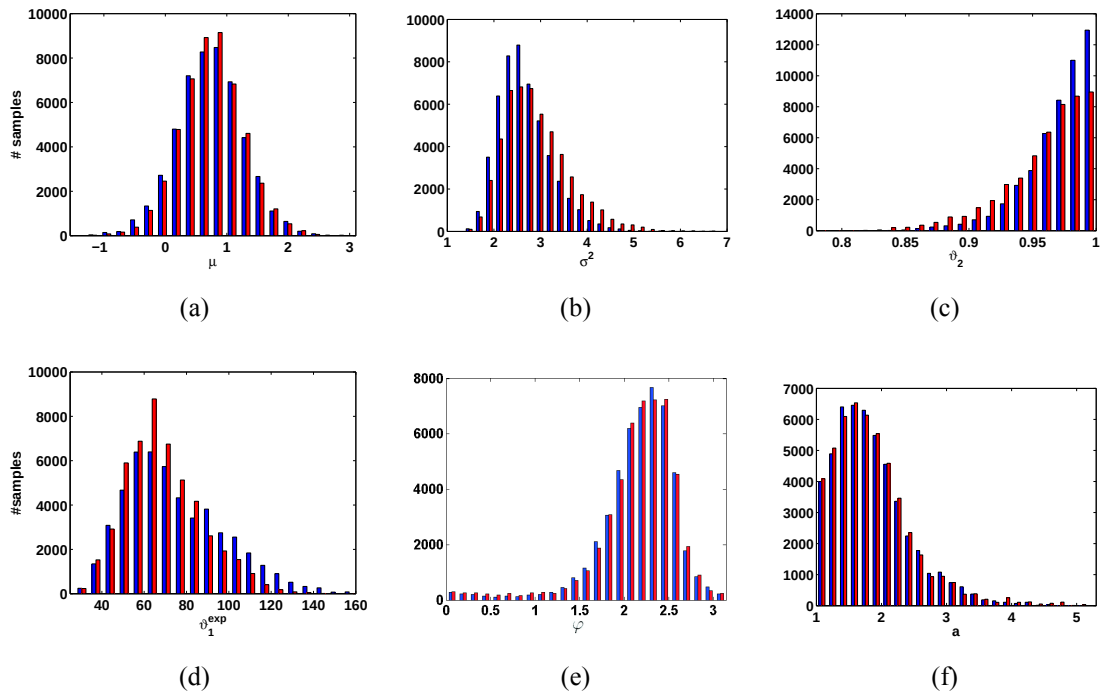


Figure 2: Histograms of 50000 posterior samples for the model parameters. Red bars: Uniform prior for $\boldsymbol{\theta}$. Blue bars: Jeffreys' prior for $\boldsymbol{\theta}$.

anisotropy parameters a and φ. Histograms of 50000 posterior samples are visualized in Figure 2. Compared with the uniform prior for $\boldsymbol{\theta}$, the Jeffreys' prior favors a smaller nugget and a larger range. To compare the two models we perform cross-validation using the means of the posterior predictive distributions as the estimators. We fin that the use of the Jeffreys' prior reduces the RMSE (19.12 vs. 19.38) the MAE (2.19 vs. 2.23) and the bias (0.17 vs. 0.26).

References

[1] Banerjee S, Gelfand A, Sirmans C (2003) Directional rates of change under spatial process models. J Am Stat Assoc 98:946-954

[2] Berger J, De Oliveira V, Sanso B (2001) Objective Bayesian analysis of spatially correlated data. J Am Stat Assoc 96:1361-1374

[3] De Oliveira V (2003) A note on the correlation structure of transformed Gaussian random fields Aust N Z J Stat 45:353-366

[4] De Oliveira V (2007) Objective Bayesian analysis of spatial data with measurement error. Canad J Stat 35:283-301

[5] Kazianka H, Pilz J (2010) Bayesian spatial modeling and interpolation using copulas. Comp Geosci (accepted)

[6] Kazianka H, Pilz J (2010) Copula-based geostatistical modeling of continuous and discrete data including covariates. Stoch Environ Res Risk Assess (in press). doi: 10.1007/s00477-009-0353-8

INCORPORATING NON STATIONARITY IN CATEGORICAL VARIABLES SIMULATIONS WITH MCS

D. D'Or[1], D. Allard[2] and R. Froidevaux[1]

[1] Ephesia Consult, 9 rue Boissonnas, CH-1227 Geneva, Switzerland. dimitri.dor@ephesia-consult.com, Tel. +41 22 300 4629.

[2] Biostatistique et Processus Spatiaux (BioSP), INRA, Site Agroparc, 84914 Avignon, France. allard@avignon.inra.fr. Tel. +33 4 32 72 21 71.

Abstract

The Multinomial Categorical Simulation (MCS) algorithm (D'Or *et al*., 2008) uses a maximum entropy approach to compute the local conditional distribution of a multinomial vector of variables conditionally to the categories observed in the neighbourhood. It has been shown in Allard *et al.* (accepted for publication) to be an efficient approximation of the BME approach for categorical variables proposed by Bogaert (2002).

The probability of occurrence of category i at location $\boldsymbol{x}_0$ given the categories observed in the neighbourhood can be computed using the equation

$$p^*_{i_0|i_1,\ldots,i_n} = \frac{p_{i_0}\prod_{k=1}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})}{\sum_{i_0=1}^{I} p_{i_0}\prod_{k=1}^{n} p_{i_k|i_0}(\mathbf{h}_{0k})} = \frac{p_{i_0}^{1-n}\prod_{k=1}^{n} p_{i_0,i_k}(\mathbf{h}_{0k})}{\sum_{i_0=1}^{I} p_{i_0}^{1-n}\prod_{k=1}^{n} p_{i_0,i_k}(\mathbf{h}_{0k})},$$

where p_{i_0} is the univariate probability (frequency) of category i at location $\boldsymbol{x}_0$, $p_{i_k|i_0}(\mathbf{h}_{0k})$ is the probability of occurrence of category i at location $\boldsymbol{x}_k$ conditionally to the occurrence of category i at location $\boldsymbol{x}_0$, and $p_{i_0,i_k}(\mathbf{h}_{0k})$ is the joint probability of occurrence of category i at location $\boldsymbol{x}_k$ and i at location $\boldsymbol{x}_0$. This equation is very simple and fast to compute since it only includes products and sums of first and second order probabilities.

Up till now, only stationary situations were considered. However, proportion trends are frequently observed in natural processes. For example, there can be a trend in proportion of geological facies, with some facies occurring more frequently in some part of the domain than in others.

In this paper, we extend the approach to the non-stationary cases. We discuss the various types of non-sationarity and we focus further developments on the most relevant one, i.e. when the univariate probabilities are non-stationary (First order non stationarity).

We show that the equation presented above can be computed using stationary transiograms $p_{i_k|i_0}(\mathbf{h}_{0k})$ (Carle and Fogg, 1996) reflecting the spatial correlation and non stationary univariate probabilities p_{i_0} as trend information.

Synthetic non conditional simulations using 3 categories are presented to illustrate the method. A first example considers the case of a North-South trend for the two first categories (Figure 1). Resulting maps (Figure 2) reflect the decreasing proportion of category 1 from South to North and the corresponding increasing of proportion for category 2.

In a second example, a more complex type of trend is considered (Figure 3), inducing a probability of occurrence near to zero for category 1 in the central South area and for category 2 in the central North area. Simulated maps (Figure 4) show that the desired features are reproduced, including the ordered sequences of categories into the North-East direction imposed by the model of transiograms.

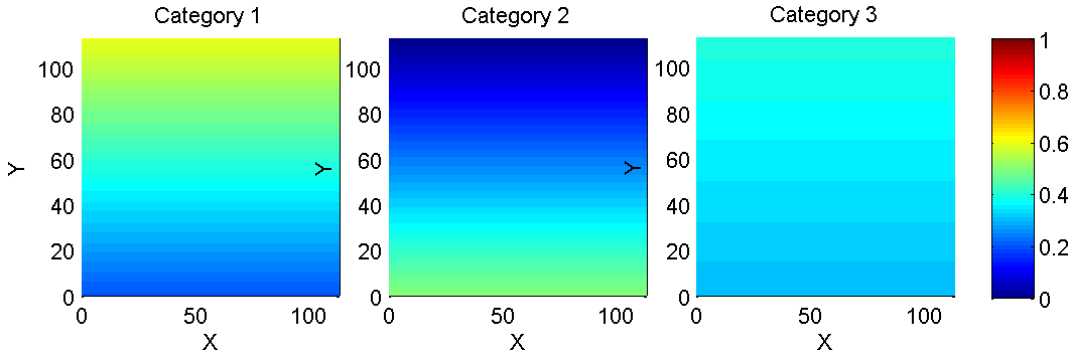


Figure 1: *Proportion trend maps for Example 1 : North-South trend in probabilities of occurrences for the 3 categories.*

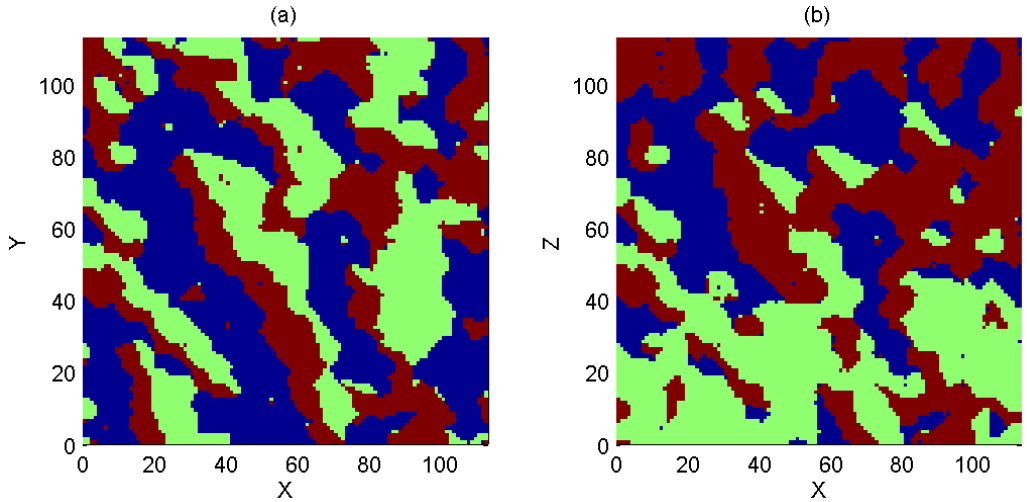


Figure 2: *Simulations using MCS (a) without incorporating the trend information, and (b) with the trend information.*

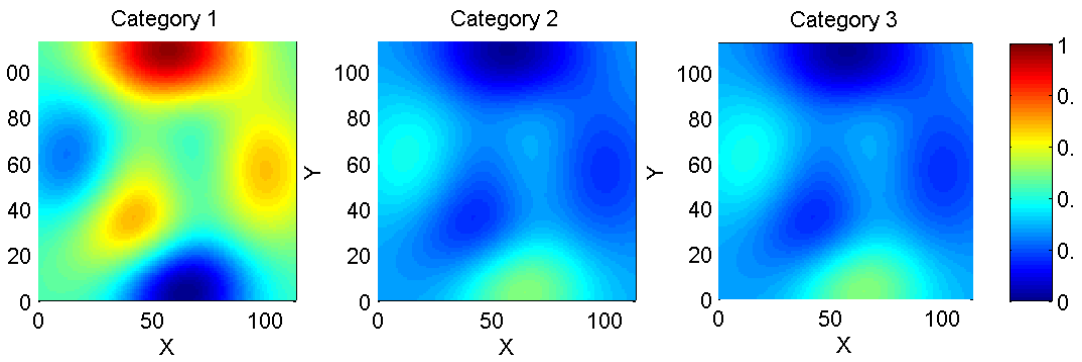


Figure 3: *Proportion trend maps for Example 2: Complex trend in probabilities of occurrences for the 3 categories.*

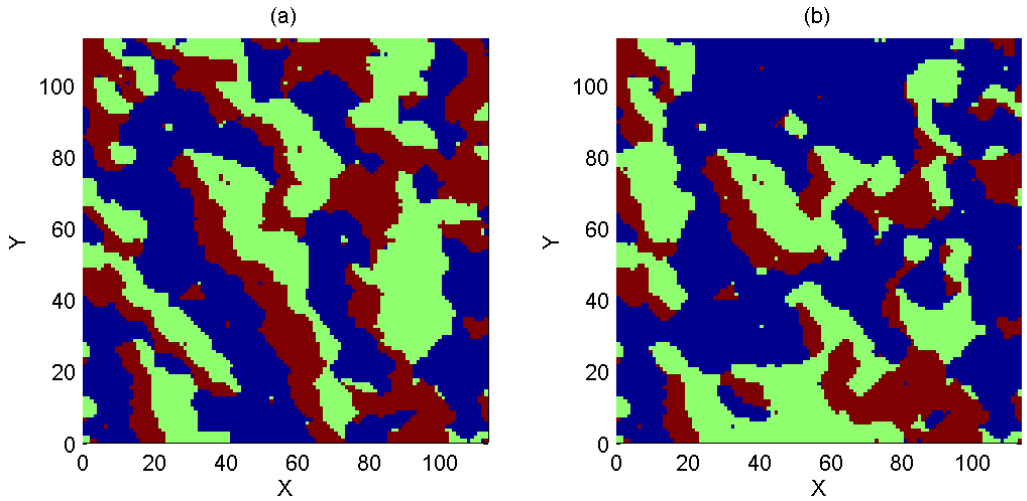


Figure 4: *Simulations using MCS (a) without incorporating the trend information, and (b) with the trend information.*

References

Allard D., D'Or D. and Froidevaux R. (accepted for publication) Estimating and Simulating Spatial Categorical Data Using an Efficient Maximum Entropy Approach. European Journal of Soil Science. To appear.

Bogaert, P. (2002) Spatial prediction of categorical variables: the Bayesian Maximum Entropy approach, Stochastic Environmental Research and Risk Assessment, 16, 425–448.

Carle, S. F. and Fogg, G. E. (1996) Transition probability-based indicator geostatistics, Mathematical Geology, 28, 453–476.

D'Or D., Allard D., Biver P., Froidevaux R. and Walgenwitz A.. (2008) Simulating categorical random fields using the multinomial regression approach. In: Proceedings of the Eighth International Geostatistics Congress. Eds. J. Ortiz and X. Emery. Santiago de Chile, pp. 59-68.

VISUALISATION TECHNIQUES FOR MOVING WINDOW KRIGING

P. Harris[1], C. Brunsdon[2], M. Charlton[1]

[1] National Centre for Geocomputation, NUIM, Maynooth, Co. Kildare, Ireland;
Paul.Harris@nuim.ie; Tel: +353 1 708 6204; Fax: +353 1 708 6456
[2] Department of Geography, University of Leicester, Leicester, UK;
cb179@leicester.ac.uk; Tel: +44 116 252 3843; Fax: +44 116 252 3854

A technique that has often proved worthy when modelling some heterogenic (or nonstationary) spatial process, is moving window kriging (MWK) (Haas 1990). Here, unlike most standard techniques, the variogram is allowed to vary across space. This interesting adaptation provides a MWK model with the potential to not only improve prediction accuracy over a standard kriging counterpart, but also improve estimates of prediction uncertainty. In this study, we investigate the MWK model in some detail, where we demonstrate ways to explore and visualise this model, both to help in its calibration and to help in the interpretation of its output.

For MWK, a local variogram is determined at every target location in this fully-automatic, continuous and locally-adaptive technique. At each location, the variogram parameters are used to calibrate a kriging algorithm to provide a prediction and its variance. Crucial to MWK is the spatial scale at which the local variography takes place. If the window is too small, then the variography tends to be unreliable as information is limited. If the window is too large, the output of MWK will tend to that of its standard kriging counterpart, and thus offer no benefit. The decision of whether or not to apply MWK over a stationary counterpart often depends on a trade-off between: (a) many ill-fitted local variograms, but with potentially more accurate model outputs and (b) a well-fitted (and understood) global variogram with possibly less accurate model outputs. As MWK is inherently more complex and approximate, there must be good reason to favour it. MWK will be of little worth if local variograms are only marginally different from each other and from the global one.

MWK models can extend most standard (stationary variogram) kriging algorithms. For example: Haas (1990; 1996) constructs lognormal, regression and co- kriging versions; Lloyd and Atkinson (2002) and Pardo-Igùzquiza et al. (2005) present universal kriging versions; and Cattle et al. (2002), an indicator kriging version. There are also MWK-nonparametric hybrids, where the local variogram parameters found from MWK are kernel smoothed (i.e. KS-MWK) (Haas 2002; Pardo-Igùzquiza et al. 2005). This use of a kernel function allows a *global* model of spatial dependence to be defined that represents the spatial process as a whole (Sampson et al. 2001). Smoothed parameters can also be robust to outlying data that can affect the local variography. A different hybrid is possible when a kernel function is not used to smooth local variogram parameters (as in KS-MWK), but is used instead to smooth the individual (estimated) semivariances of each lag interval according to the distance of these paired values from a target

location. Here a geographically weighted variogram (GWV) is found at any location, creating an alternative MWK hybrid (Harris et al. 2010). The expected benefit of a GWV is that it does not rely on limited local information to be local and reliable, as that does a local classic variogram estimator of a standard MWK model.

Rudimentary MWK models *globally-specify* (i.e. the specification is fixed): a classic variogram estimator, a weighted least squares (LS) variogram model fit and one variogram model-type (e.g. Matérn) using the raw sample data. Local and global issues of trend, nonlinearity and data clustering are commonly ignored. However, supplementing the extensions and hybrids above, more elaborate MWK models are possible. For example, models can be *globally-specified* with: a different variogram estimator (e.g. a weighted variogram for clustered data); an alternative LS variogram fitting technique; WSS or AIC to choose the best LS variogram model fit; etc. Models can also be *globally-specified* with: maximum likelihood parameter estimation (Pardo-Igùzquiza et al. 2005); nonparametric variography; techniques to incorporate any anisotropic effects; etc. Furthermore, models can *locally-specify* (i.e. the specification varies across space): the variogram estimator (e.g. use a robust variogram in areas of outliers); the variogram model-type; the data detrending (Haas 1990; 1996; Pardo-Igùzquiza et al. 2005); and the data transform (Haas 2002).

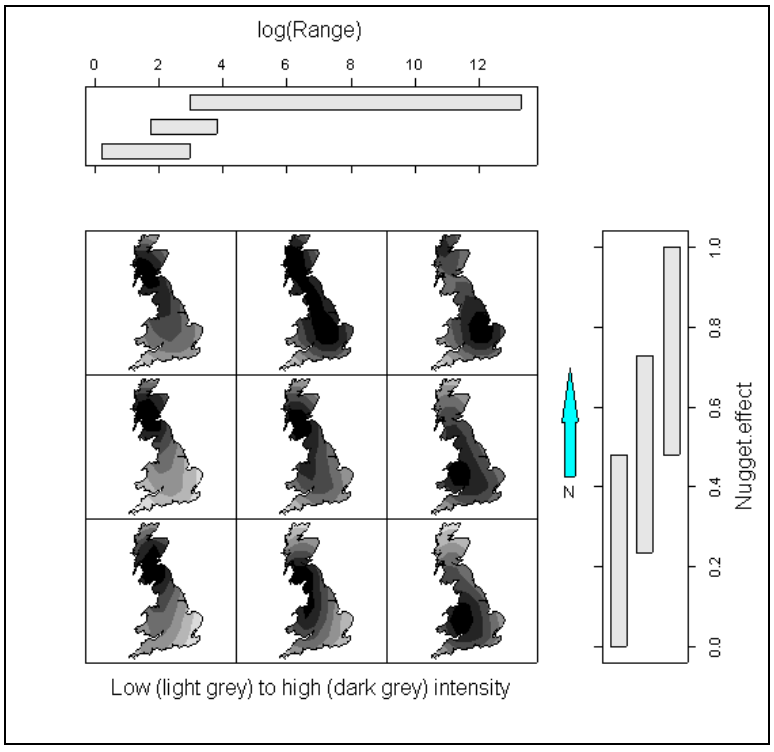


Figure 1: *Comap (specified using kernel density estimation) for the local nugget effect versus the local correlation range (in logs) found using a basic MWK model of the critical load data.*

Therefore MWK models in basic, elaborate and hybrid form are all possible; most of which are constructed and visually explored in this study. We apply MWK models to a freshwater acidification critical load dataset covering Great Britain (CLAG Freshwaters 1995). The size and scale of this dataset suits a nonstationary approach where EDA provided strong evidence of variogram nonstationarity. This EDA included a visual deconstruction of the global variogram using $\mathbf{h}$-scatterplots, the variogram cloud and the interactive methods of Glatzer and Müller (2004). It also included an assessment of the proportional effect and the construction of local variograms within eight (geographical) partitions. This study's graphical outputs from any MWK calibration should confirm these exploratory findings.

Thus at every target location of a MWK model, a wealth of information (or model parameter data) is possible depending on the complexity of the MWK model adopted. For example, not only can the local variogram parameters be mapped, but also: the local window size; the number of pairs at the first lag; the local trend parameters; the local transform parameter; local model fit statistics; etc. Furthermore, indicator data can be used to map the types of: transform; detrending; variogram model; and variogram estimator used. It is also possible to compare the lag-specific semivariances of each local variogram estimator to those of the global estimator.

To move on from a simple mapping of these rich MWK spatial datasets, we instead calculate geographically weighted (GW) summary statistics (Brunsdon et al. 2002) to visualise this data. For example, GW correlations can be found that spatially relate the local nugget effect (the ratio of nugget variance to total variance) to the local range. This not only provides an assessment of variogram nonstationarity but also identifies regions most suited to kriging (kriging performs well in a relative sense when the nugget effect is small and the range is large). Alternatively, a spatial variant of the co-plot, the comap (Brunsdon 2001), can be constructed to assess the same relationship. Such a comap is given in Fig. 1, where there is clear evidence of variogram nonstationarity, as both the nugget effect and the range vary across space. Furthermore, critical load prediction accuracy should be relatively strong in areas that traverse the borders of Wales and England (bottom-right map), but relatively weak in areas of W Scotland and NW England (top-left map). In this study, other MWK relationships are similarly visualised with the comap and for relationships in higher dimensions, a spatial visualisation is not so easy, so here we present parallel coordinate plots. Other visualisation techniques are currently under consideration.

References

Brunsdon C (2001) The Comap: exploring spatial pattern via conditional distributions. Comput Environ Urban Syst 25:53-68

Brunsdon C, Fotheringham AS, Charlton ME (2002) Geographically weighted summary statistics - a framework for localised exploratory data analysis. Comput Environ Urban Syst 26:501-524

Cattle JA, McBratney AB, Minasny B (2002) Kriging methods evaluation for assessing the spatial distribution of urban soil lead contamination. J Environ Qual 31:1576-1588

CLAG-Freshwaters (1995) Critical Loads of Acid Deposition for United Kingdom Freshwaters. ITE, Penicuik: Critical Loads Advisory Group, Sub-report on Freshwaters

Glatzer E, Müller WG (2004) Residual diagnostics for variogram fitting. Comp Geosci 30:859-866

Haas TC (1990) Lognormal and moving window methods of estimating acid deposition. J Am Stat Assoc 85:950-963

Haas TC (1996) Multivariate spatial prediction in the presence of non-linear trend and covariance non-stationarity. Environmetrics 7:145-165

Haas TC (2002) New systems for modelling, estimating, and predicting a multivariate spatio-temporal process. Environmetrics 13:311-332

Harris P, Charlton ME, Fotheringham AS (2010) Moving Window Kriging with Geographically Weighted Variograms. To appear in Stoch Environ Res Risk Assess

Lloyd CD, Atkinson PM (2002) Nonstationary approaches for mapping terrain and assessing prediction uncertainty. Transactions in GIS 6:17-30

Pardo-Igùzquiza E, Dowd P, Grimes D (2005) An automatic moving window approach for mapping meteorological data. Int J Climatol 26:665-678

Sampson PD, Damian D, Guttorp P (2001) Advances in modeling and inference for environmental processes with nonstationary spatial covariance. NRCSE-TRS No 61, Nat. Res. Centre for Statistics and the Environment Technical Report Series, 2001

LOCAL POLYNOMIALS FOR DATA DETRENDING AND INTERPOLATION IN THE PRESENCE OF BARRIERS

A. Gribov[1] and K. Krivoruchko[1]

[1]Environmental System Research Institute, 380 New York St, Redlands, CA, USA, 92373
agribov@esri.com, kkrivoruchko@esri.com, 1-909-793-2853.

Summary

In this paper, we discuss features of local polynomial interpolation (LPI), focusing on the problem of unstable solution of the LPI system of linear equations. First, a diagnostic based on condition numbers is discussed. Then, a variant of Tikhonov regularization is proposed, which allows producing continuous predictions and prediction standard errors nearly everywhere in the data domain. This variant of interpolation can be applied in the presence of barriers defined by polylines. We illustrate the use of LPI with both simulated data and with the water nutrient data collected in the Chesapeake Bay.

Local polynomial interpolation and kriging

LPI is used in meteorology since 1949. An excellent review of this model's features can be found in the beginning of the first book on objective spatial interpolation written in 1963 by Lev Gandin [1]. Several disadvantages of LPI motivated Gandin to develop the optimal interpolation model now known as kriging [2, 1]. Recent investigations of LPI's features can be found in the statistical literature on Geographically Weighted Regression, a generalization of LPI, see for example [3].

We discuss disadvantages of LPI in detail. However, LPI has several useful features:

It is fast: the computational time is proportional to the number of the data in the searching neighborhood. By comparison, kriging's computational time is proportional to the number of data raised to the third power.

It is successfully used for data detrending in various geostatistical models.

It is equivalent to the universal kriging model when all spatial variation is described by the trend and the semivariogram model is a nugget effect, which consists of the measurement error only (i.e. there is no microstructure component of the nugget, see also [4]). In this case, LPI can be used as an optimal interpolator.

When LPI is used to describe large-scale variation in conventional kriging models, its prediction uncertainty can be ignored because it is small when the moving window kernel is large and, therefore a sufficient number of observations is used for smooth surface creation. This is not the case, however, when the model attempts to completely describe the spatial data variation, because in this case, optimal kernel size is usually small and the solution of the LPI system becomes unstable. We show how this instability can be described by a variant of condition

number diagnostic, which we call spatial condition number. Prediction and prediction standard error should not be used in the areas with large spatial condition numbers. This requirement leads to a problem with finding optimal parameters of the model using cross-validation diagnostic. We show how cross-validation can be used when predictions cannot be made at some data locations.

Tikhonov regularization and interpolation in the presence of barriers

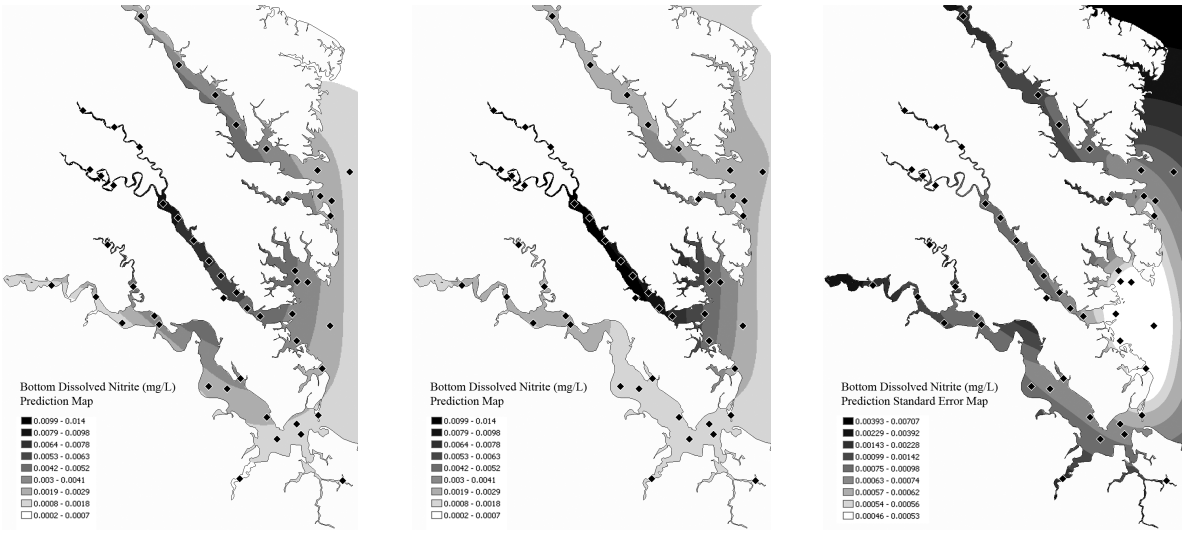
Inability to make predictions in all locations in the data extent makes LPI model less attractive for GIS applications. This problem can be solved using Tikhonov regularization, also known as ridge regression [5].

We discuss a variant of Tikhonov regularization for spatial interpolation with first order local polynomials. The resulting model has a Bayesian interpretation: the more unstable the linear system, the larger the correction required. This correction is varying spatially and it can be displayed as a map. The prediction standard error cannot be calculated exactly because it requires knowledge on the true regression coefficients, which are unknown. We show how a pessimistic estimate of the prediction standard error can be made.

Since LPI with Tikhonov regularization model makes continuous predictions in the data extent, it can be used for interpolation in the presence of barriers defined by a series of polylines (instability of standard LPI can be very large in this situation). In the case of non-transparent barriers, the distance between points is calculated as the shortest sum of the series of lines which do not intersect the polylines.

Figure 1 shows an example of local polynomial interpolation using the water nutrient data collected in the Chesapeake Bay (symbols ♦ show the locations of the observations.) The map on the left shows predictions calculated based on Euclidean distance metric, ignoring the border of the water bed. The map in the center show predictions in the presence of barriers and a map at right show the prediction standard errors in the presence of barriers.

As expected, the predictions and the prediction standard errors (not shown for standard LPI in this abstract) are clearly different for Euclidean and non-Euclidean distance metrics. It can be shown that the predictions are more accurate when the barriers for water flow are used.



a) b) c)

Figure 1: *a) Predictions made using Euclidean distance metric; b) Predictions in the presence of barriers; c) Prediction standard errors in the presence of barriers*.

References

[1] Gandin LS (1963) Objective Analysis of Meteorological Fields. Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965)

[2] Gandin LS (1959) The problem of optimal interpolation. Trudy GGO 99: 67-75 (In Russian)

[3] Waller LA, Zhu L, Gotway CA, Gorman DM, Gruenewald PJ (2007) Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. Stochastic Environmental Research and Risk Assessment 21 (5), pp. 573-588

[4] Krivoruchko K, Gribov A, and Ver Hoef JM (2006) A new method for handling the nugget effect in kriging. In T.C. Coburn, J.M. Yarus, and R.L. Chambers, eds., Stochastic modeling and geostatistics: Principles, methods, and Case Studies, Volume II: AAPG Computer Applications in Geology 5, p. 81 – 89

[5] Tikhonov AN (1943) On the stability of inverse problems. Doklady Akademii Nauk SSSR **39** (5): 195-198

MULTI-WAY SENSITIVITY ANALYSIS USING CLUSTERING TECHNIQUES

G. Mariethoz[1], J. Caers[1], C. Scheidt[1]

[1]Stanford University, ERE Department, 367 Panama st, rm 65, Stanford, CA 94305, USA.
Email: gregoire.mariethoz@minds.ch

When applying geostatistical algorithms to build geological models, one has to define parameters describing geological continuity, such as the choice of a training image, the variogram model or the Boolean model parameters and rules. In most cases, spatial variation is represented by a combination of parameters (for example mean, variance, variogram type and range) in the presence of constraining data (hard, soft, dynamic data). We use sensitivity analysis to determine the influence of such combinations of parameters on a given target. By figuring out what matters and what doesn't, such analyses can lead to a more focused modeling of the subsurface uncertainty.

Experimental designs are well suited for assessing the sensitivity of parameters which have a smooth relationship with the response of interest such as the mean or the variogram range. However, traditional methods for sensitivity analysis can be challenged when the relationship between parameters and models response cannot be defined analytically, and when different parameters are related to each other in a non-linear manner. Nor do experimental designs easily allow for inclusion of data, such as obtained from wells, geophysics or dynamic responses that may constrain these parameters.

This is illustrated in figure 1, where we show the one-way correlation between a given decision variable (the target) and the parameters used to generate 1000 realizations with an object-based method that has 5 parameters. The first parameter determines the type of objects, and the remaining parameters define various objects attributes (objects anisotropy, size, number). If it exists, any relation to the target is not visible when considering parameters individually. Indeed, no correlation can be deduced from figure 1.

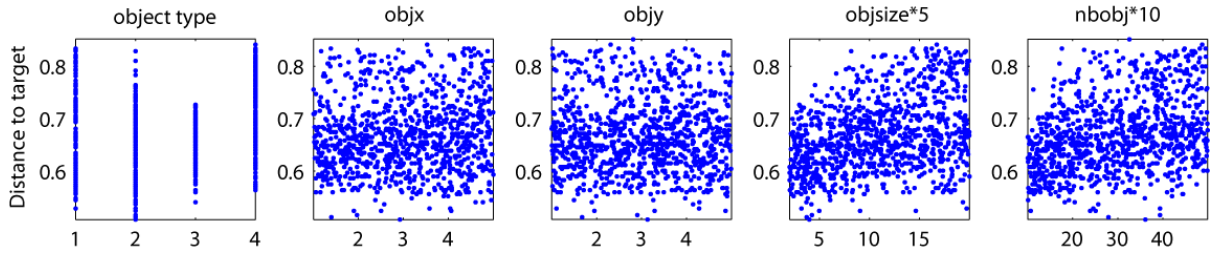


Figure 1: *One-way sensitivity analysis*

To tackle this problem, we propose to use distance-based techniques (Suzuki and Caers, 2008; Scheidt and Caers, 2009; , 2010). The approach is based on the notion of a distance (a single

scalar value) that is defined with respect to a specific goal. The purpose of this distance is to bring structure to the large variability of geological models.

Using a distance, one can create a map of a set of complex models. This tool is known in statistics as Multi-Dimensional Scaling (Borg and Groenen, 1997). Given a distance matrix **D**, such a representation displays an ensemble of models $\mathbf{m}_i$ as a set of points in a possibly high-dimensional space, each point corresponding to a model. Points are arranged in such a way that their respective distances are preserved as much as possible, in a least-squares sense. **D** can be computed using any appropriate measure of distance.

Once the models are represented as a set of points, clustering techniques can be used to define groups of models that share similar characteristics. Interestingly, different distances will produce different groupings.

In the framework of sensitivity analysis, we propose to use distances and clustering techniques to 1) identify the combinations of parameters that influence a given target and 2) to assess the information contained in the data with respect to this target. To this end, we distinguish three possible distances between models:

- The target distance, defined by the difference in the prediction given by two models.
- The parameters distance that defines the similarity between the sets of parameters used to generate models.
- If data are present the data distance is considered. It is the difference in the reproduction of data by any two geological models.

The data distance can be used to assess the worth of matching the data. Perfectly informative data would mean maximum correlation between data and target distances. In a realistic setting, these two distances may be poorly correlated, indicating that the exercise of matching the data through an inverse problem is irrelevant. Figure 2 displays the correlation between data distance and target distance for the same 1000 object-based models as before. It shows that in this case, the data are informative on the target.

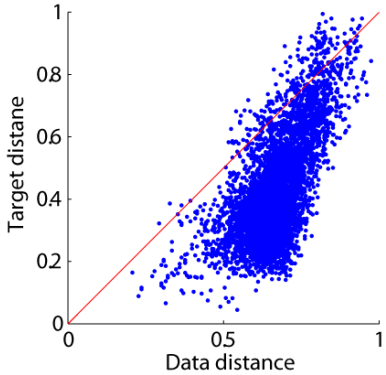


Figure 2: *Correlation between data distance and target distance*

The parameters distance is best suited when one-way sensitivity analysis is not able to identify the influence of individual parameters, such as in figure 1. It consists in the distance between two

vectors of parameters. We use the weighted linear combination of the distances between each individual parameter in the parameters vectors. The weights for the linear combination are chosen such that the parameters distance is as much as possible correlated to the data distance. In other words, models that have similar parameters will yield similar predictions.

Then, we perform clustering using the parameters distance. Each resulting cluster is made of models that have been produced with similar parameters sets. Finding sensitive combinations of parameters (or sensitivity envelopes) is then a matter of identifying those clusters that are in accordance with the data and that are influential on the target. Figure 3 displays the combinations of parameters related to two such clusters. One can see that specific relations between parameters should be considered. For example, it seems that one should use either numerous small objects of type 3, or many large anisotropic objects of type 2.

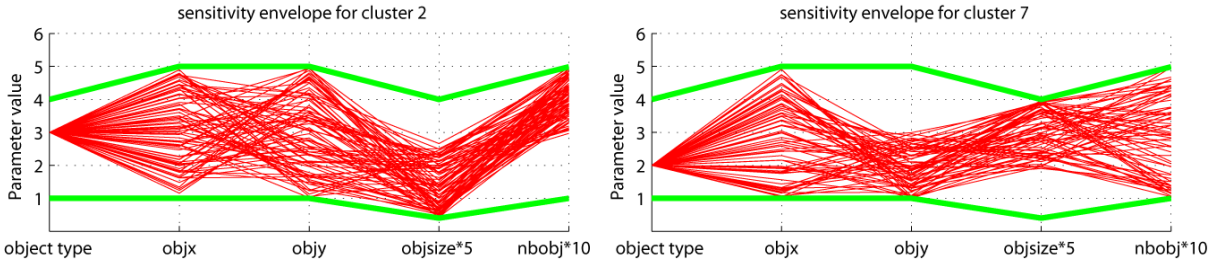


Figure 3: *Clusters of parameters that are in accordance with the data and influential on the target.*

Once favorable combinations of parameters have been identified, one has basically updated the prior or initial conceptual uncertainty into a new (posterior) uncertainty that is constrained by the data. Note that geological models are not matched to data. Instead, the distance methodology is used to identify the combinations of parameters that are consistent with the data and relevant for the target.

References

Borg, I., and Groenen, P. (1997), Modern multidimensional scaling: theory and applications, Springer.

Scheidt, C., and Caers, J. (2009), Representing Spatial Uncertainty Using Distances and Kernels, Mathematical Geosciences, 41(2009): 397-419.

Scheidt, C., and Caers, J. (2010), Bootstrap confidence intervals for reservoir model selection techniques, Computational Geosciences, 14: 369-382.

Suzuki, S., and Caers, J. (2008), A Distance-based Prior Model Parameterization for Constraining Solutions of Spatial Inverse Problems, Mathematical Geosciences, 40(2008): 445-469.

PATTERN RECOGNITION IN ENVIRONMENTAL DATA USING GENERAL REGRESSION NEURAL NETWORKS

M. Kanevski, V. Timonin, S. Robert, L. Foresti

Institute of Geomatics and Analysis of Risk, University of Lausanne, Amphipole, 1015 Lausanne, Switzerland, Mikhail.Kanevski@unil.ch, tel. +41 21 6923531; Fax. +41 692 3535

Introduction

Traditionally environmental data are considered in low dimensional geographical spaces (2 or 3) using geostatistical models (interpolations or simulations) and variography is used to characterize spatial (correlation) anisotropic structures (patterns). In reality, many real environmental problems are described in high dimensional input spaces, which consist of geographical coordinates and geo-features, such as satellite images, physical models, or generated, for example, from digital elevation models (slope, curvature, etc). For example, the nonlinear problems of topo-climatic modelling in mountainous regions, assessment of natural hazards (landslides, avalanches) are considered in spaces of dimension more than 10 (typically 30-50). The number and combination of active features contributing to the solution can vary depending, for example, on season or phenomena under study. This is a typical problem of feature selection/extraction. Therefore, development and adaptation of efficient tools recognizing and modelling patterns in high dimensional spaces and capable to automatically select relevant geo-features is an important issue.

The present methodological research studies the application of adaptive General Regression Neural Networks (GRNN) to solve the following tasks: 1) pattern detection, or discrimination between structured data and random noise; 2) automatic selection of relevant features; 3) modelling of patterns (spatial predictions); 4) analysis of the residuals. The study is carried out using real data sets of monthly wind measurements in Switzerland.

Analysis of Patterns Using General Regression Neural Networks

GRNN is a development of classical Nadaraya-Watson regression estimator (Specht 1991; Kanevski et al 2009). In the general case of advanced adaptive (anisotropic) GRNN model, when kernel bandwidths depend on the geo-features, the GRNN prediction (conditional mean value) is defined by the following formula Specht (1991):

$$\hat{f}(\mathbf{x}) = \frac{\sum_{n=1}^{N} f_n \exp\left\{-\sum_{k=1}^{m} \frac{D^2_{x,n,k}}{2\sigma_k^2}\right\}}{\sum_{n=1}^{N} \exp\left\{-\sum_{k=1}^{m} \frac{D^2_{x,n,k}}{2\sigma_k^2}\right\}}$$

where, m – dimension of the input space (number of geo-features), f_n - measurements; $D^2_{x,n,k}$ - k^{th} component of the distance between prediction point and n^{th} data measurement; σ_k - kernel

bandwidth corresponding to the k^{th} geo-feature. Usually Euclidean measure is used to compute the distances. In a more general setting Mahalanobis distance can be applied. The model can also be generalized to the case when bandwidth is variable in space.

The only unknown parameters are kernel bandwidths and it is a task of GRNN training. There are several possibilities to tune bandwidths. The most frequently used are based on cross-validation error and its variants (leave-one-out; k-Fold) or splitting of data into training and testing subsets. Optimal bandwidths have minimum testing/cross-validation errors.

GRNN has important properties: when bandwidths are very small GRNN is equivalent to nearest neighbour estimator (Voronoï polygons); when bandwidths are very large – larger than the region of the study, the solution is a mean value which corresponds to unstructured data (no patterns). Such situation is encountered when error curves (surfaces) do not show a clear minimum. This can be used as an important diagnostic tool for the analysis of the residuals. When bandwidths are larger than maximal distances between points in some directions, than these features are automatically ignored. Therefore anisotropic GRNN automatically performs feature selection by scaling input features differently. Another possibility to select the most appropriate features using GRNN is to consider all possible combinations of features (2^m-*1*) and to select models with the minimum cross-validation or testing errors. When number of features is less than 20 this approach is quite feasible but for more features other optimization techniques can be used to find approximate good solutions.

Real case studies of wind speed mapping support the ideas and hypotheses considered above. The seven dimensional input space is composed of geographical coordinates, elevation and four geo-features generated from digital elevation model (terrain slope, directional derivative and convexity at different scales). Output is monthly wind speed. Number of measurements is 132. In order to avoid problems with scaling, all inputs and outputs were normalized (Z-score transformed). Therefore we can perform both the analysis with isotropic model (all bandwidths are the same in the Z-score space) as well as complete anisotropic analysis (bandwidths are feature dependent).

Figure 1 presents an analysis of patterns in a geographical space. Cross-validation error curve shows no minimum, depicting the absence of spatial structure, which is confirmed by exploratory variography (pure nugget effect at the measurement scale).

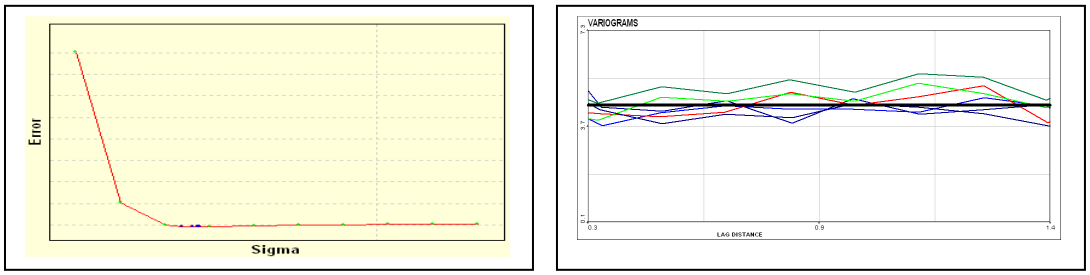


Figure 1: *GRNN training in a geographical space. Left – kernel bandwidth (X-axis) vs. cross-validation error (Y-axis); right – lag step (X-axis) vs. directional variograms (Y-axis)*

Exploratory analysis in the geo-feature space is shown in Figure 2. In this space cross-validation error curve has a clear minimum, i.e. there is a structure. In Figure 2 (right) a cross-validation curve for shuffled data (spatial structure destroyed) is presented, which extends the hypotheses about absence of structure to a high dimensional space.

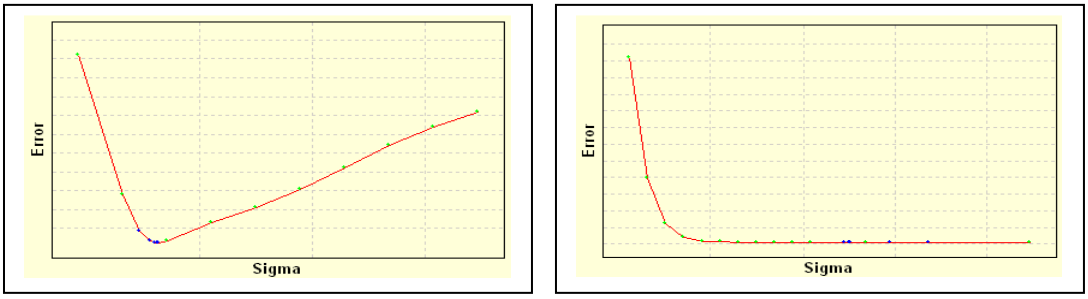


Figure 2: *GRNN training in a geo-feature (7D) space. Left – raw data wind measurements; right – shuffled data (destroyed spatial structure)*

Now, let us consider the application of GRNN as a feature selection tool. In this case (7D) all possible models (all possible combinations of inputs) were evaluated using cross-validation and then they were ordered according to the cross-validation error (Figure 3). Most relevant inputs of best models (red rectangle in Figure 3) are the ones derived from topography and not the original XY coordinates.

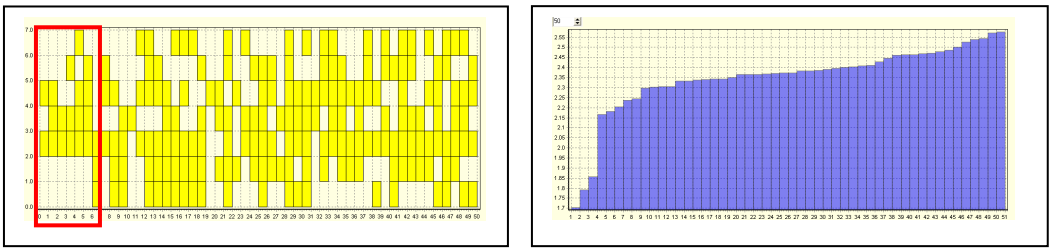


Figure 3: *Wind data. "Feature" selection map (left) and ordered cross-validation errors of the first 50 models (right). X-axes correspond to the ordered solution (best at the origin) and Y axes correspond to feature's order (left figure, from bottom: X,Y,Z,terrain convexity1+2,directional derivative, slope) and cross-validation error (right figure)*

Conclusions

In the present study an efficient application of General Regression Neural Networks for the detection, analysis and modelling of high dimensional geo-spatial data is proposed. It was demonstrated that adaptive GRNN can be used as an efficient feature selection algorithm. These properties can be important in developing automatic mapping systems and as exploratory

modelling tools. Future research will consider other real case studies and extension of GRNN in order to estimate the variances and uncertainties of the predictions.

Acknowledgements

The research was partly supported by Swiss NSF projects N200021-126505 (www.geokernels.org) and N 200020-121835 (www.kernelcd.org).

References

Kanevski M., A. Pozdnoukhov, and V. Timonin (2009) Machine learning for spatial environmental data. Theory, applications and software. EPFL and CRC Press, Lausanne.

Specht D. (1991) A General Regression Neural Network. IEEE Transaction on Neural Networks., 2 pp. 568-576.

MODELING LOCALLY VARYING ANISOTROPY OF CO2 EMMISSIONS

J. B. Boisvert[1], C. V. Deutsch[1]

[1]Department of Civil and Environmental Engineering, 3-133 Markin/CNRL Natural Resources Facility, University of Alberta, Edmonton, Alberta, Canada, T6G 2w2.

In a property modeling context, the variables of interest to be modeled often display complex non-linear features. Techniques to incorporate these nonlinear features, such as multiple point statistics or cummulants, are often complex with input parameters that are difficult to infer. The methodology proposed in this paper uses a classical vector based definition of locally varying anisotropy to characterize non-linear features and incorporate locally varying anisotropy into numerical property models. The required input is an exhaustive field of anisotropy orientation and magnitude. The methodology consists of (1) using the shortest path distance between locations to define the covariance between points in space (2) multidimensional scaling of the domain to ensure positive definite kriging equations and (3) estimation or simulation with kriging or sequential Gaussian simulation respectively. The methodology is demonstrated on a CO_2 emissions data set for the United States in 2002.

Introduction

Incorporating anisotropy into numerical models enhances the predictive power of the resulting models. Consider a simple example of air pollution, adding a slight northerly wind (Figure 1 left), the pollution is more continuous in a single direction (constant anisotropy). If we are given the knowledge that the pollution source is located in a mountainous valley (Figure 1 right), the wind is now more erratic and the pollution spread is complex, thus, the direction of anisotropy varies and the pollution is said to be continuous in locally varying directions (locally varying anisotropy). Knowing the anisotropy of the variable of interest increases predictive power. At the unsampled location A, if there is a constant northerly anisotropy in the form of wind we would predict no pollution. Adding non-stationary anisotropy in the form of erratic winds alters our estimate of the pollution level (Figure 1). Knowing the direction and magnitude of anisotropy helps in making better predictions at unsampled locations. This can be applied to modeling any variable of interest that displays locally varying anisotropy (LVA) in the domain of interest.

Past Work

Some methodologies exist for incorporating LVA because variables such as pollution spread, rain fall patterns and animal migration can be extremely non-linear. Sampson and Guttorp's (1992) early work in this area has since been expanded upon. Typically, applications using

stream distances involve pollution or fish migration studies; small 1D grids are sufficient. Generating 3D models containing millions of cells with these methods would be infeasible.

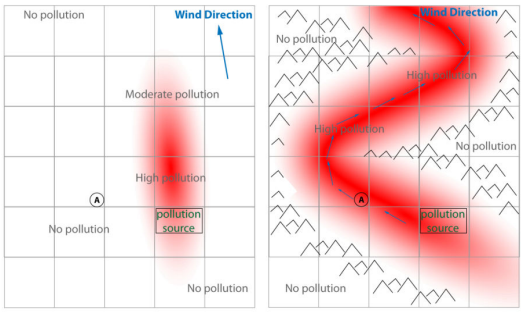


Figure 1: *Cloud indicates increased concentration. Left: North wind, constant anisotropy. Right: Erratic wind, LVA.*

A similar approach is considered here with the water distance analogous to the shortest path distance (SPD). The current work with stream distances attempts to incorporate physical boundaries and barriers (i.e. rivers) rather than locally varying anisotropies as presented in this paper.

Methodology

At the heart of geostatistics is the spatial prediction of variables from sparse sample data. From these spatial predications of concentrations, grades or porosities, volumetric calculations can be made and site classifications determined. LVA is incorporated into the spatial predictions of variables by using the optimized SPD between locations. Kriging and sequential Gaussian simulation (SGS) are geostatistical tools often applied in geostatistical studies as an alternative to IDW and are modified to incorporate LVA. Multidimensional scaling (MDS) techniques are implemented to guarantee the positive definiteness of this system of equations.

Step 1: Generate the LVA field: The LVA field is parameterized by local directions and magnitudes of anisotropy which can be specified differently for each cell of the model. In 2D two parameters are required (1) a strike and (2) an anisotropy ratio (magnitude in strike direction/magnitude in orthogonal direction). In 3D, three angles are required to fully define the orientation of the anisotropy and two ratios are used to define the magnitude (as per GSLIB convention). The LVA field is an integral portion of this methodology. A number of data sources could be used to infer the LVA field, including hard/soft data, seismic, knowledge of the underling physical laws acting on the variable of interest, direct measurement of the anisotropy parameters at discreet locations or expert interpretation. In the proposed methodology it is assumed that an exhaustive LVA field exists for the variable of interest. We are currently working on a manuscript that addresses LVA generation.

Step 2: Calculate the SPD: With the definition of the LVA field the distance between any two locations in space can be determined by formulating the problem as a graph (Boisvert and

Deutsch, 2010). The distance between points is determined by the path that results in the shortest anisotropic distance between locations. In the case of second order stationarity, this shortest path is the straightline path; however, when LVA is introduced, nonlinear paths that follow the major directions of anisotropy can return shorter distances.

Step 3: Incorporating LVA into modeling: The SPD metric can be used with an inverse distance estimator by changing the distance metric in the traditional inverse distance equation to the SPD. Incorporation of the SPD into kriging and SGS is more difficult as the SPD metric does not ensure positive definite kriging equations (Curriero 2005). ISOMAP-L (Tenenbaum et al. 2000), a multidimensional scaling technique, is used to guarantee positive definiteness (Boisvert and Deutsch, 2010). Kriging and SGS are applied as per a typical implementation but with the distance between locations obtained from the coordinates of the grid cells in the *q*-dimensional Euclidian space after ISOMAP-L.

Case Study

The methodology summarized above and given in detail in Boisvert and Deutsch (2010) is applied to the 2002 total CO_2 output of the United States (
Figure **2**). Much of the LVA in this data set is due to high population centers and major traffic ways. The data is available on a 10km square grid. The LVA field is shown in
Figure **2** and was generated with the moment of inertia method (Mohammadhassanpour and Deutsch, 2008). In general the LVA field follows the major traffic ways. In areas where there is no anisotropy (all high values or all low values) the orientation of the LVA field is erratic, but the anisotropy ratio is ~1.0 (isotropic), thus has little effect in these areas. The exhaustive grid is resampled to 100km resolution and kriging with LVA is compared to traditional kriging (Figure 3). Local prediction is improved, measured by an increase of 9% in the covariance between the truth and estimates. Moreover, local features due to the traffic (highlights on Figure 3) are seen in the map using LVA (Figure 3). For convenience, comparisons are made in normal score units; gains in covariance and local features are similar in original units.

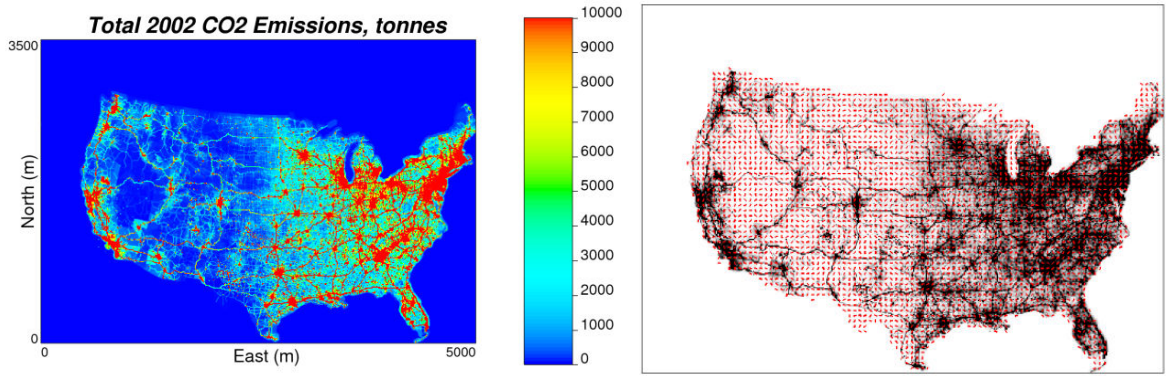


Figure 2: *Tonnes of CO_2 emissions in the United States, 2002 (source: Vulcan Project, Gurney 2010, data available online). Right: Grayscale image with LVA field overlain, plot dimensions are identical.*

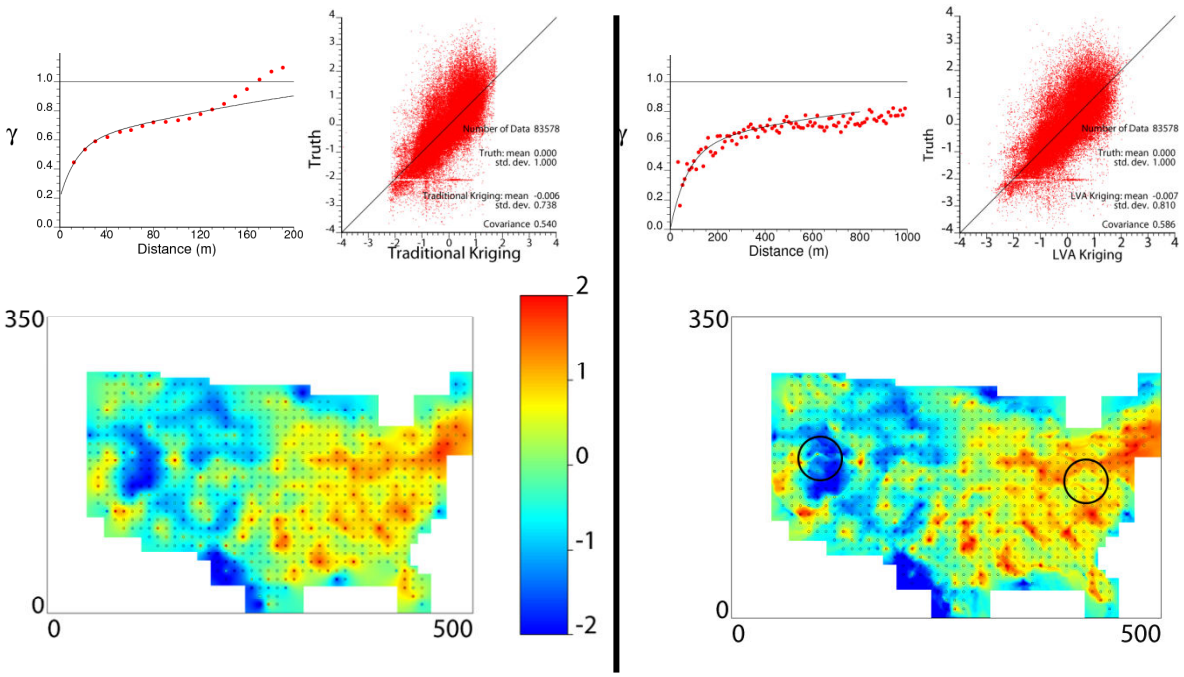


Figure 3: *Left: Traditional kriging modeled variogram, estimate mean map (showing locations of the 834 samples used) and cross-validation with the truth (scatterplot). Right: Kriging with LVA.*

References

Boisvert J, Deutsch C (2010) Programs for Kriging and Sequential Gaussian Simulation with Locally Varying Anisotropy Using Non-Euclidean Distances. Accepted to Computers & Geosciences.

Curriero F (2005) On the use of non-Euclidean distance measures in geostatistics. Mathematical Geology. 38:907-926

Gurney, K, 2010. Vulcan Project. Accessed March 2010. http://www.purdue.edu/eas/carbon/vulcan/index.php

Mohammadhassanpour R and Deutsch C, 2008. Fitting local anisotropy with mass moments of inertia. Center for Computational Geostatistics Annual Report 10. University of Alberta. 9p.

Sampson P, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association. 87:108-119

Tenenbaum J, Silva V, Langford J (2000) A global geometric framework for nonlinear dimensionality reduction. Science. 290:2319–2323

EFFICIENT PARAMETRIC VARIOGRAM ESTIMATION FOR REAL-TIME INTERPOLATION OF ENVIRONMENTAL MONITORING DATA

K. Henneböhl[1], E. Pebesma[1,2], W.G. Müller[3]

[1] Institute for Geoinformatics, University of Münster, Weseler Str. 253, 48151 Münster, Germany, katharina.henneboehl@uni-muenster.de

[2] 52° North Initiative for Geospatial Open Source Software GmbH, Münster, Germany

[3] Department of Applied Statistics, Johannes-Kepler-University of Linz, Austria

Abstract

Increasingly, environmental monitoring data become available within a few hours after data capture. This has motivated the development of automatic real-time mapping systems that use geostatistical interpolation techniques at their core. Unless one assumes that autocorrelation does not change over time, a central problem for such automated geostatistical interpolation is the automated fitting of a model for the spatial or spatio-temporal correlation of the monitoring data. This paper will look into efficient strategies for doing this, based on generalized least squares estimation.

One common, classical kriging approach for real-time mapping algorithms involves the computation of the empirical semivariogram by binning the $n(n-1)/2$ semivariogram cloud values $0.5(Z(si) - Z(sj))^2$ into distance classes, and fitting a parametric model to the binned values typically using ordinary least squares (OLS) or weighted least squares (WLS; Cressie 1985). The advantage of this method is that it is simple and fast, and it allows visual checking using an easy graphical diagnostic tool, the obligatory variogram plot. When applied in an automated algorithm, the procedure however needs to know the number and widths of the bins ahead of computing them, as no interactive exploration is possible. Furthermore, the $N = n(n-1)/2$ semivariogram cloud values are obtained from n observations, and thus exhibit strong correlations, due to sharing common points or due to spatial vicinity, and these correlations are typically ignored when fitting with OLS or WLS. An alternative that does not have these two disadvantages is to fit a variogram model directly to the variogram cloud, using the generalized least squares (GLS) method (Müller 1999). Other alternatives are maximum likelihood (ML) or restricted maximum likelihood (REML) fitting of variogram models, but these methods make strong assumptions about multivariate normality that are hard to verify.

The GLS estimation method calls for an iterative solution where each iteration needs to solve several $N \times N$ systems, which has rendered this method only possible for very small data sets. As N is proportional to n^2, and solving a system of N linear equations is $O(N^3)$, the algorithm has complexity of $O(n^6)$, which is not particularly attractive for larger n. However, induced by the

rapid technological development of today's high performance processors, grid and cloud computing infrastructures, and the availability of large amounts of cheap memory, computational limits have lost significance, and this method now becomes relevant for the fitting of variograms to moderately large data sets as well. Major questions that remain to be answered are (i) how much does this efficient GLS estimator improve estimation, compared to OLS and WLS approaches? (ii) how much advantage does it give for the case of automated variogram fitting, also e.g. compared to ML and REML methods, and (iii) how can efficient computer infrastructures be used to obtain estimates fast enough for near real-time application. A fourth question is whether a GLS estimator based on reduced information (e.g. by discarding large distances, or discarding the most redundant variogram cloud points) is still more efficient than OLS or WLS methods that ignore correlations.

This paper will look into these questions by using simulation experiments for cases where the (generating) variogram model is known to compare efficiencies between the various methods. In addition, an application to a real case study (near real-time interpolation of hourly air quality measurements) will be demonstrated and evaluated. Finally, efficiency in terms of computational efficiency will be adressed, i.e., this paper will explore the possibilities to re-formulate the algorithm for GLS variogram estimation in a parallelized manner such that it makes appropriate use of the computational resources within real-time mapping systems. These may be grid infrastructures or highly performant graphics processing units (GPU).

References

Cressie N (1985) Fitting variogram models by weighted least squares. Mathematical Geology 17: 563-586 Müller WG (1999) Least-squares fitting from the variogram cloud. Statistics & Probability Letters 43: 93-98

AUTOMATIC SEMIVARIOGRAM MODELING BY WEIGHTED LEAST SQUARES AND BY CROSS-VALIDATION

A. Govaerts[1], A. Vervoort[2]

[1] Research unit mining, K.U.Leuven, Kasteelpark Arenberg 40 bus 2448, B-3001 Leuven, Belgium, annelies.govaerts@bwk.kuleuven.be, tel: +32(0)16321774, fax:+32 (0)16 321988

[2] Research unit mining, K.U.Leuven, Kasteelpark Arenberg 40 bus 2448, B-3001 Leuven, Belgium, andre.vervoort@bwk.kuleuven.be, tel: +32(0)16321171, fax:+32 (0)16 321988

In a geostatistical study the correct modeling of the semivariogram is a decisive step. The fitting of a semivariogram to field data and the evaluation of this fit are usually done by manual fitting by trial-and-error of a model selected by the user, usually based on the visual appearance of the experimental semivariogram. According to Lamorey and Jacobson (1995) more quantitative methods are needed to estimate and evaluate semivariogram parameters, especially when assessing small changes in a variable's spatial characteristics or if limited field data are available.

Therefore different possible automatic methods are described in this paper. The first two are weighted least squares methods, with different weight functions. Also two cross-validation methods are developed. For this methods a procedure based on the mean square error and d_i ($=\sigma_i/\varepsilon_i$) is developed. The results of this procedure are two parameters with a value between 0 and 1 for every semivariogram model j. The 'correct' range and nugget effect are then chosen as the range and nugget effect where S, the sum of these two parameters, is the smallest (see Figure 1 for an example). The sill of the model can be calculated based on d_i. In that case the procedure is called the fully-automatic cross-validation method. But one can also determine the sill in such a way that the model, with the calculated range and nugget effect, fits the experimental semivariograms visually the best. This is called the semi-automatic cross-validation method.

The four different modeling procedures are used in a case study, whereby the models of the different modeling techniques are compared with the visually determined model. The original dataset existed of 35 pile base resistance values at a certain depth in a tertiary sand layer in Belgium; resulting in a spherical semivariogram with a range of 50 m, no nugget and a sill of 16.9 MPa^2. In order to achieve a larger dataset with smaller separation distances this semivariogram is used for a conditional simulation of 2350 points with a separation of 1 m. 15 subsets are chosen out of this dataset. These subsets have different separation distances between the data points, different starting points, and different total lengths:

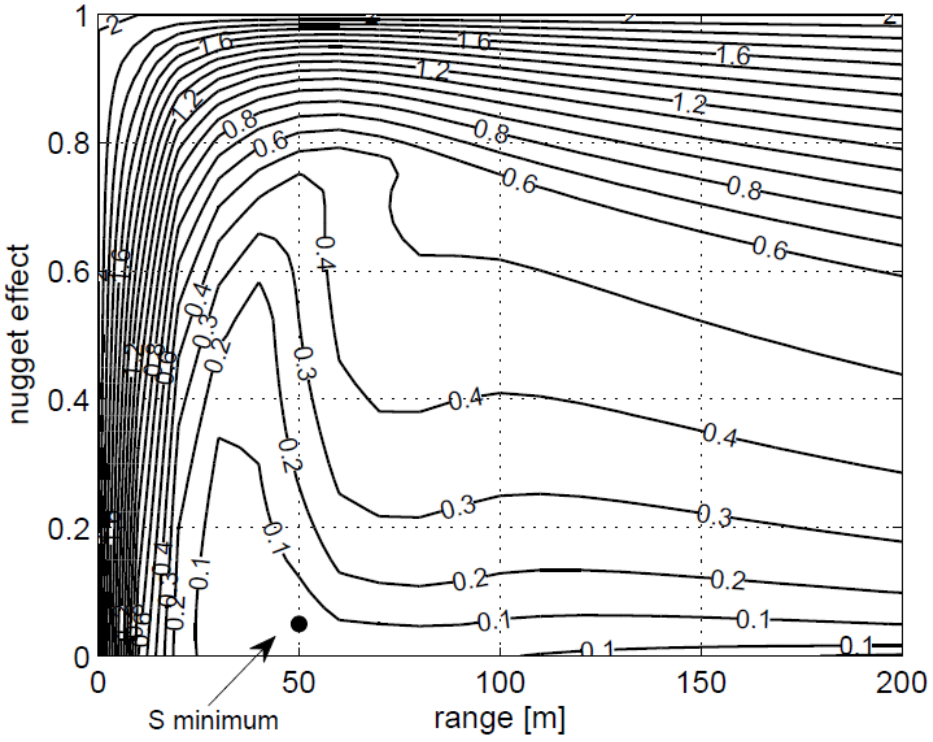


Figure 1: *Contourplot of parameter S. The range and nugget effect are determined based on the minimum of S (used data are from subset 8 of the case study).*

As an example Figure 2 shows the semivariogram of a dataset covering a total length of only 50m. This is low compared to the expected range of 50m. Furthermore there are only 50 datapoints available. For this subset, and other subsets with a low total length, the range and nugget effect are best determined with the cross-validation technique (see Figure 2).

The case-study shows that generally all methods can be used to determine the semivariogram model. If there are enough data available and the semivariogram is easily modeled visually, the automatic methods do not show better results than the visual method. Of course, if one wants to model a lot of semivariograms, using an automatic method can provide a gain in time. Furthermore in a lot of casers, small datasets are available in a real study. In cases where it is more difficult to see a clear structure the weighted least square methods and the semi-automatic cross-validation method are good methods to determine the semivariogram. In terms of the nugget effect, the results of the cross-validation methods are better than the weighted least squares methods.

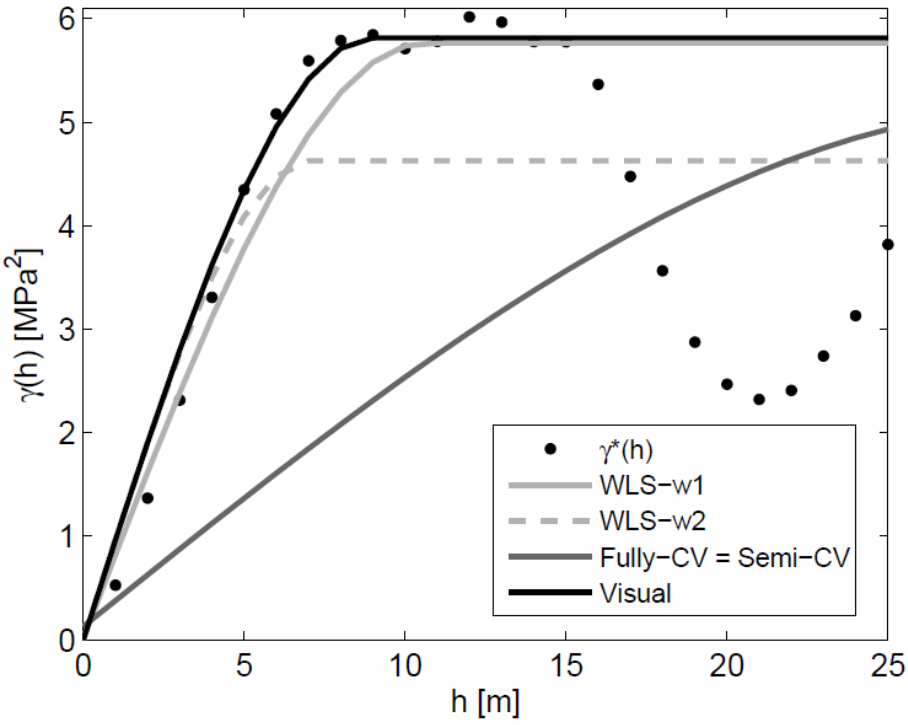


Figure 2: *Experimental semivariogram values and semivariogram models of subset 2. The models are determined with the weighted least squares method using weight function w1 (WLS-w1) and weight function w2 (WLS-w2), with the fully automatic (Fully-CV) and semi-automatic cross-validation method (Semi-CV) and visually.*

Lamorey G, Jacobson E (1995) Estimation of semivariogram parameters and evaluation of the effects of data sparsity. Mathematical Geology 27(3): 327-358

PROBABILISTIC SPATIAL PREDICTION OF INDOOR RN, FROM SOIL AIR CONCENTRATION

P. Bossew

Bundesamt für Strahlenschutz / German Federal Radioprotection Authority; Köpenicker Allee 120-130, 10318 Berlin. e-mail: pbossew@bfs.de , ph. +49-30-18333-4231.

Abstract

An important task in radioprotection is estimating if at a location the radon (Rn) concentration in indoor air exceeds a reference level. Spatial estimation and modelling of Rn related variables are known to be delicate due to high spatial variabilities and strongly left-skewed (right-tailed) data distributions. For regulatory purposes one is however more interested in local probabilities that a level is exceeded, than in actual levels whose estimates are, inevitably, very uncertain given the physical nature or the variable, apart from statistical complications.

In this contribution I propose modelling the probabilities on normal- (nscore-) transformed data. It is shown that exceedance probabilities for soil and indoor concentrations correlate well, and that there are optimal combinations of thresholds of soil and indoor concentration for achieving high correlations. Thus, if a reference dataset of Rn in soil air is available as indicator of the geogenic Rn potential, predictions of local Rn hazard can be made.

Data

We used the data set Rn concentrations in of soil and indoor air as included in the German "BURG" Rn database. The data are restricted to the region Oberpfalz of Bavaria, known for high Rn concentrations in some areas (due to geological peculiarities; not to be further discussed here). Indoor values are taken from ground floor measurements only.

Table 1: basic statistics of the used Rn data

	soil (kBq/m³)	indoor (Bq/m³)
no. data	500	891
median	54	57
range	5 … 1134	2 … 3043

A uniformly distributed random noise component of 2% (below measurement uncertainty) was added to the concentrations to avoid tiers in rank statistics.

Nscore transform

If one succeeds to transform the data $\{z_i\}$, i=1…N, into normal, $\{y_i\}$, the wanted local exceedance probability at x* equals $p(x^*)=[prob(Z>T)](x^*) = \Phi((y^*-\tau)/s_Y)$, with Φ = standard normal, τ the threshold in Y-space, y* the OK estimate and s_Y the kriging SD in Y-space, at x*. We assume here multi-Gaussian distribution of Y.

In order to achieve univariate normality the data, distributed F, were nscore transformed to univariate normal by $y := \Phi^{-1}(F(z))$. The problem is how to estimate the quantiles $q:=F(z)$ out of the empirical data $\{z_i\}$. For the "bulk" of data, $q^b_i := (-.5+\text{rank}(z_i))/N$ may be a reasonable choice, if enough data are available altogether, but in general not so for the "tail" data. For modelling the upper tail (the lower tail is of minor importance and can be treated in analogy) I propose the following procedure: cut the empirical distribution at q_G or $z_G=F^{-1}(q_G)$, located quite high. To the data above q_G, until a visually chosen upper limit, fit a power distribution G to the graph (z_i, q^b_i), which is reasonable for many environmental datasets, including Rn; $1-G(z)=\beta\ z^{-\alpha}$; from $F(z_G)=G(z_G)$ we find β, and $G(z)=1-(1-F(z_G))(z/z_G)^{-\alpha}$. The estimated quantile is then, $q_i = q^b_i$, for $z_i \le z^G$; $q_i = G(z_i)$, for $z_i \ge z^G$. The new data are thus built $y := \Phi^{-1}(q)$. The estimated q_i for high z_i depends then on the choice of q^G or z^G, and also the upper limit. Of course, since the estimated G depends on possibly "wrong" estimates q^b, it is still an imperfect estimate of the true F.

Spatial estimation

The such transformed data y_i(soil) and y_i(indoor) are subjected to point-OK independently. The variograms are well structured. A grid of 20 km x 20 km was chosen as reasonably adapted to the data spacing. The grids were blanked with deliberately chosen (after visual inspection) cut-off values of the kriging SDs, generating a total 974 grid points for further analysis. The estimated transformed soil and indoor Rn concentrations, (y(soil))* and (y(indoor))* at each non-blank grid node are well correlated, Figure 1.

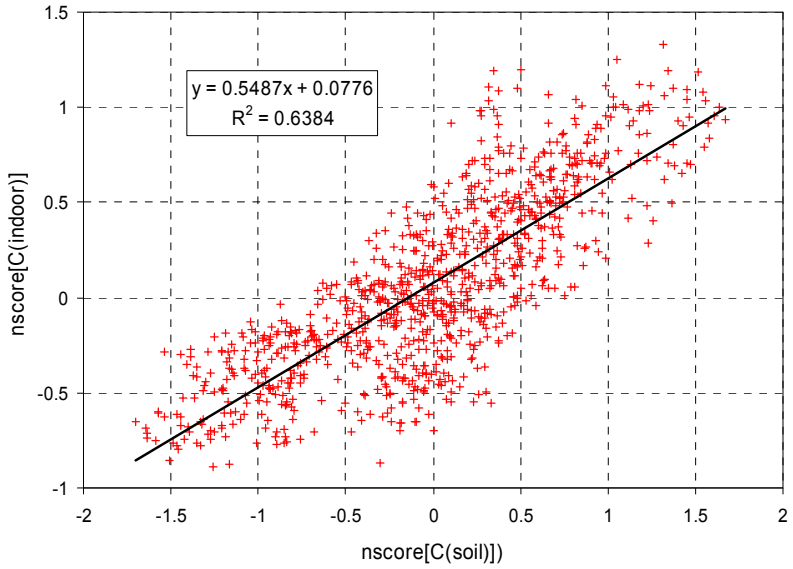


Figure 1: *Correlation of nscore-transformed Rn concentrations in soil and indoor air estimated at grid nodes***.**

For a number of thresholds T(soil) and T(indoor) the exceedance probabilities $p_i:=p(T_i;$ soil) and $p_j:=p(T_j;$ indoor) are computed at each grid node. Next the Pearson correlation coefficient (r^2), together with linear regression, is calculated for p_i vs. p_j. The result is shown graphically in Figure 2.

Conclusions

We observe that local exceedance probabilities of Rn concentrations in soil and indoor air are well correlated, which allows, to some extent, modelling or prediction of the latter from the former. As Figure 2 shows, however, that given a threshold of indoor Rn whose exceedance probability shall be estimated, the threshold of soil Rn should be chosen carefully to enable optimal prediction (i.e. minimize uncertainty). For example, if one wants the model the probability that the indoor concentration exceeds 400 Bq/m³ (a common reference value), from $\log_{10}(400)=2.6$ we find an optimal $\log_{10}(T(soil))$ = about 2.2 or T(soil)=150 kBq/m³. In other words, optimal prediction of the indoor concentration to exceed 400 Bq/m³ is made from the probability that soil air concentration exceeds 150 kBq/m³.

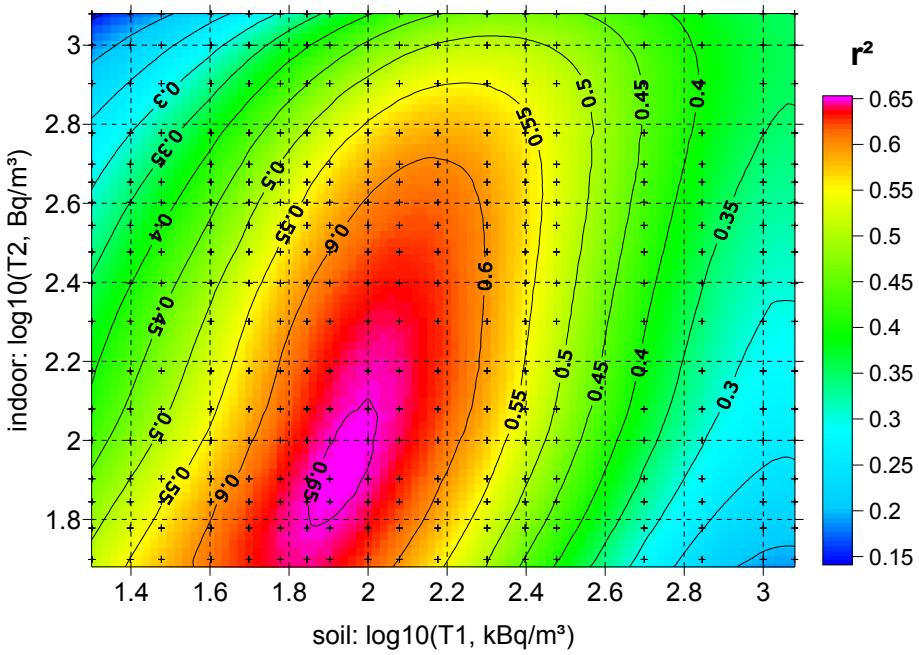


Figure 2: *Correlation coefficients r^2 between prob[C(soil)>T_1] and prob[C(indoor)>T_2], in dependence of T_1 and T_2. Crosses: calculated r^2 values in (T_1, T_2) space. Surface interpolated with local polynomials (Surfer 8 software).*

As next steps we plan cross-validating the method by partitioning the data into training and validation sets. On the side of physics, we want to investigate if the correlations can be improved by considering geological control and possibly anthropogenic factors related to house construction styles which are known to affect indoor Rn.

As to technicalities, so far the variables soil and indoor concentration were estimated (in Y-space) independently. If we have the technical means, we want to co-estimate them since they are not independent variables, as partly being controlled by the same physical source. Then we want to investigate the robustness of OK estimates of nscore transformed data in view of uncertainties in modelling the distribution tails. Further multi- (or at least approximate bi-) Gaussian distribution of the $\{y_i\}$ has to be checked, and the regressions between the two probabilities have to be examined in depth for linearity, influence of leverage, etc., and finally if there is an optimal choice of grid spacing for prediction purposes.

GEOSTATISTICS FOR RADIOLOGICAL EVALUATION: METHODOLOGICAL DEVELOPMENTS AND OUTCOMES

Y. Desnoyers[1], J.-P. Chilès[2], D. Dubot[3], N. Jeannée[1], J.-M. Idasiak[4]

[1] GEOVARIANCES, 49bis Av. Franklin Roosevelt, BP91, 77212 Avon, France
desnoyers@geovariances.com/jeannee@geovariances.com, +33 (0)1.60.74.90.90.
[2] MINES ParisTech, Centre of Geosciences, 35 Rue St-Honoré, 77300 Fontainebleau, France
jean-paul.chiles@mines-paristech.fr , +33 (0)1.64.69.47.86
[3] CEA DSV/FAR/USLT/SPRE/SAS, 18 route du panorama, BP6, 92265 Fontenay-aux-Roses, France, didier.dubot@cea.fr, +33 (0)1.46.54.82.94
[4] CEA DEN/VRH/DTEC/SDTC/LTM, Centre de VALRHO, BP 17171, 30207 Bagnols-sur-Cèze, France, jean-marc.idasiak@cea.fr, +33 (0)4.66.79.63.67

Introduction

For more than a century, the development of the French nuclear industry has led to the construction and the exploitation of hundreds of facilities to produce nuclear fuel, to burn it through experimental reactors or nuclear power plants, and to recycle it. Dozens of them are now under decommissioning.

At the end of process equipment dismantling, the complete decontamination of nuclear facilities requires the radiological assessment of residual activity levels of building structures. As stated by the International Atomic Energy Agency (2001): "Segregation and characterization of contaminated materials are the key elements of waste minimization".

In this framework, the relevance of the geostatistical methodology relies on the presence of a spatial continuity for radiological contamination, characterized through the variographic analysis (Chilès and Delfiner 1999). Geostatistics then provides reliable methods for activity estimation, uncertainty quantification and risk analysis, which are essential decision-making tools for decommissioning and dismantling projects of nuclear installations.

However the use of the geostatistical methodology points out several theoretical issues which are to be taken into account like the combined integration of non-destructive surface measures and in-depth samples or the use of spatialized migration profiles, etc.

The paper first presents the geostatistical methodology applied to a former nuclear facility and its added value as regards sampling optimisation and waste categorisation. Then it discusses one of the theoretical issues which is the implementation of an isofactorial model to deal with the observed destructuring of high values.

Geostatistics for nuclear waste characterisation

The geostatistical approach is currently applied to several former nuclear facilities of the "Commissariat à l'Energie Atomique" in France (Desnoyers 2009). The added value of the geostatistical methodology lies in (i) the optimisation of the sampling strategy according to the evaluation objectives, (ii) the reliable mapping of the contaminated areas using a cokriging estimation integrating the radiation measures as auxiliary information, and (iii) the

estimation of the corresponding waste volumes through risk analyses on conditional simulations.

The ATUE (enriched uranium workshops) facility, located in Cadarache, is a case in point (Lisbonne and Seisson 2009). The complete dataset is constituted by more than 1,600 surface measures combined with 56 destructive sampling points.

The remediation support constraint is taken into account during the calculation of the probability to exceed a given activity level, here each workstation area is considered as an effective remediation support as presented on **Figure 1**. Thus waste segregation can be carried out on these relevant decision-making tools, which make the radiological characterisation easier and more reliable.

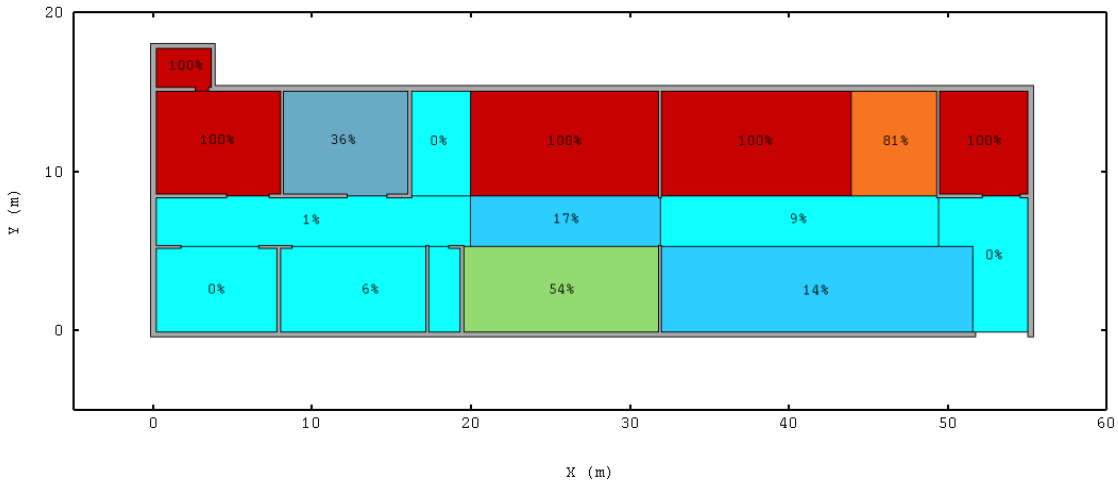


Figure 1: *Decision-making tool for waste segregation: risk map to exceed a given activity level.*

Destructuring of high values

Radiological characterisation within the geostatistical framework has to cope with more theoretical issues. In particular, the use of nonlinear geostatistical methods (gaussian anamorphosis, isofactorial model, disjunctive kriging, etc.) implies the assumption of several distribution hypotheses that have to be checked. In our case, the Gaussian-transformed radiological data do not conform to a multigaussian nor to a bigaussian assumption. Fortunately the Hermitian model provides a family of spatial behaviours from the pure diffusive one (bigaussian model) to the mosaic one. This family can be ruled by a positive parameter β which can be practically fitted on a scatterplot comparing the variogram of order 1 (madogram) with the variogram of order 2 (classical variogram); see, e.g., Chilès and Delfiner (1999). A shown on **Figure 2**, our dataset perfectly fits with a β parameter equal to 2.5, between the pure diffusive case ($\beta \to \infty$) and the mosaic one ($\beta = 0$).

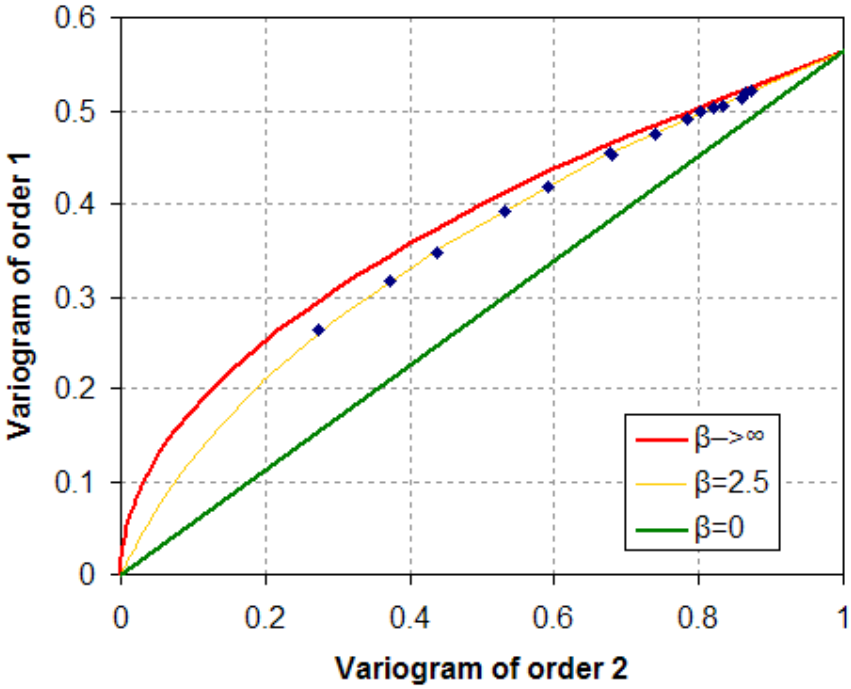


Figure 2: *Relationship between the ordinary variogram (order 2) and the variogram of order 1 of the Gaussian-transformed data. The sample points can be fitted to a Hermitian isofactorial model with $\beta = 2.5$.*

Different approaches are then compared, whose pros and cons are discussed with regard to their capability in adequately modelling the spatial structure of high values: at the scale of each workstation the bigaussian model can be used; at the scale of a whole workshop, discontinuities due to wall separations can be modelled either as faults or through a Hermitian isofactorial model.

References

J.-P. CHILÈS, P. DELFINER, Geostatistics: Modeling Spatial Uncertainty, Wiley, New-York, 695 p, 1999.

Y. DESNOYERS, J.-P. CHILÈS, N. JEANNÉE, D. DUBOT, F. LAMADIE, "Geostatistical Sampling Optimization and Waste Characterization of Contaminated Premises". Int. Conf. Global 2009, Paris, France.

INTERNATIONAL ATOMIC ENERGY AGENCY, "Methods for the Minimization of Radioactive Waste from Decontamination and Decommissioning of Nuclear Facilities", Technical Reports Series No. 401, IAEA, Vienna, 2001.

P. LISBONNE, D. SEISSON, "Enriched Uranium Workshops towards Decommissioning Cadarache Nuclear Research Centre". Int. Conf. on Decommissioning Challenges: An Industrial Reality?, Avignon, France, 2008.

SHOP OPENING HOURS AND POPULATION EXPOSURE TO NO_2 ASSESSED WITH AN ACTIVITY-BASED TRANSPORTATION MODEL

E. Dons[1,2], C. Beckx[1], T. Arentze[3], G. Wets[2], L. Int Panis[1,2]

[1] VITO (Flemish Institute for Technological Research), Boeretang 200, 2400 Mol, Belgium, evi.dons@vito.be

[2] Transportation Research Institute, Hasselt University, Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium

[3] Urban Planning Group, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

Traditionally studies of exposure have linked the concentrations of air pollutants with static data on population density. The spatial and temporal variation in population densities which is a result of traffic is often simply ignored even though traffic is a major source of urban emissions. A new class of traffic models, activity-based transport demand models (Davidson et al 2007; McNally 2000), can predict where people will perform specific activities during the day. These models can also be used to provide the traffic data (associated with the trips to these activities) needed to model ambient pollutant concentrations with a dispersion model (Shiftan 2000). In this way a truly dynamic exposure analysis can be made by geographically matching hourly concentrations and hourly population densities. Further, more detailed exposure analyses can also be performed, focusing on the differences in exposure between different activities and different subpopulations. These results have the potential to provide useful information for policy purposes.

In this paper we demonstrate for the first time that this activity-based approach can be used to evaluate the impact of general policy measures, not specifically aimed at reducing the exposure of the population, on people's exposure (Int Panis 2010). Our hypothesis is that policies that seem unrelated to either the transport or air quality domain may nevertheless have an unintended secondary impact on exposure to traffic emissions. The modeling framework will be summarized in a nutshell since this is not the main focus of this paper. An extensive description of the framework can be found in Beckx et al. (Beckx et al 2009a; Beckx et al 2009b; Beckx et al 2009c).

In a first phase, the activity-based transport model ALBATROSS was used to predict activity-travel patterns for the Dutch population (Arentze and Timmermans 2000; Arentze and Timmermans 2004). The model started by developing a synthetic population for The Netherlands and activity-travel schedules were simulated for all the individuals within this population. Next, the predicted trips for the entire population were assigned to a road network and traffic flows were converted into vehicle emissions by applying the emission factor approach from the MIMOSA emission model.

In a second phase, the activity-based approach was further extended by converting the emissions into pollutant concentrations. For this purpose, the AURORA model was applied to simulate the dispersion and conversion of the emissions into concentrations.

And finally, in the last phase, population exposure analyses were performed. The predicted hourly concentration fields from the ALBATROSS-MIMOSA-AURORA modelling chain were combined with hourly information on people's location to calculate the exposure. By using the population information from the activity-based simulation, hourly population maps were simulated and dynamic exposure values could be estimated.

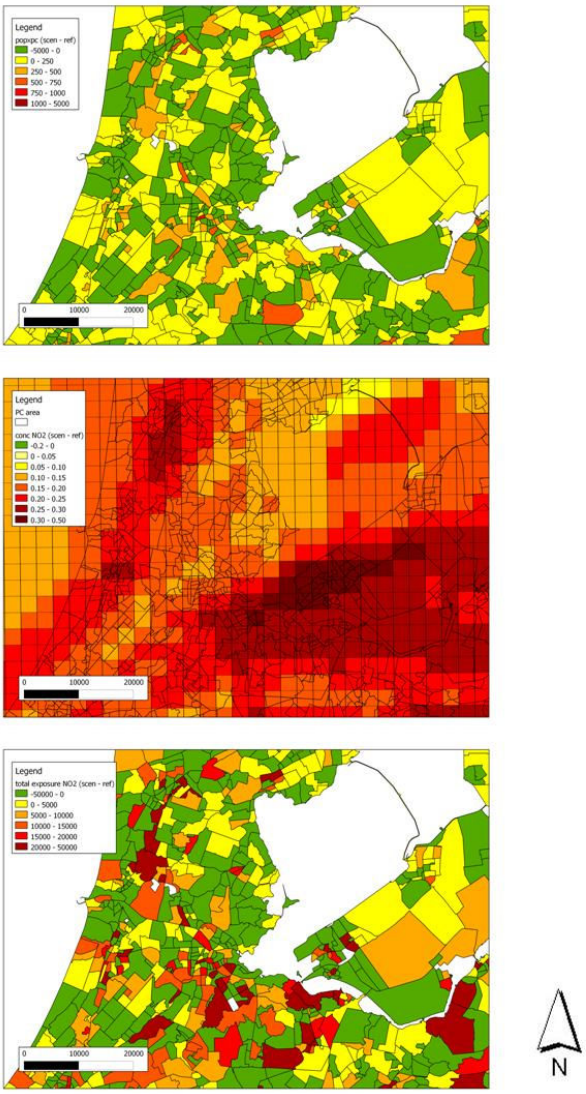


Figure 1: *The dynamic population is computed for each postal code area. For every hour of the day concentration levels for NO_2 are calculated on a 3 x 3 km² grid. Multiplying those two sets of data gives the total exposure. The maps show the difference between the scenario and the reference situation. (Monday April 4th 2005, 5 p.m., Amsterdam region, The Netherlands)*

The scenario considered here involves a widening of shop opening hours for daily and non-daily shopping. In the scenario shops are open from 6 a.m. until 10 p.m., which allows shopping activities earlier in the morning as well as shopping later in the evening. After changing these institutional constraints in the activity-based model, every stage in the modelling framework is passed through once again. By analyzing the output, the effect of changing shop opening hours on the population level exposure to NO_2 in different time-periods, on different locations, for different subpopulations (e.g. according to gender or socio-economic status) and during different activities (residential, work, transport, shopping, leisure, …) can be evaluated.

The scenario leads to a considerable increase of up to 6% in time spent on non-daily shopping activities. We can see an increase in transport time as well, although the effect is smaller. This augmentation is offset by less time spent on leisure and in-home activities. Another effect that propagated from our data is the shift of shopping activities towards the early and late hours of the day.

An evaluation (both geographically and statistically) of the differences between the scenario and the reference situation gives insight into the impact of the scenario on the population distribution across postal code areas,

the impact on the concentrations and finally on the total exposure of people in each municipality. In figure 1 this is illustrated for a random hour in April. As an example, on the first Monday in April at 5 p.m., concentrations will rise in certain neighbourhoods with a substantial 0,30 µg/m³. In addition to this, from our data we can conclude that the overall exposure of the population in the Netherlands will rise as a consequence of the scenario. Depending on the hour of the day, an average increase of population exposure to NO_2 of 0,5% is seen.

Examples of measures or scenarios that can be evaluated by such an approach are changing shop opening hours, ageing of the population, teleworking, introduction of congestion charging, etc…. This paper presented the first analysis of a scenario generated with the activity-based model ALBATROSS. We demonstrated that by using this approach the effect of a policy measure on population exposure can be assessed.

References

Arentze TA, Timmermans HJP (2000) ALBATROSS: A Learning Based Transportation Oriented Simulation System Eindhoven, The Netherlands, European Institute of Retailing and Service Studies

Arentze TA, Timmermans HJP (2004) A learning-based transportation oriented simulation system. Transportation Research Part B: Methodological 38:613-633

Beckx C, Int Panis L, Arentze T, Janssens D, Torfs R, Broekx S, Wets G (2009a) A dynamic activity-based population modelling approach to evaluate exposure to air pollution: Methods and application to a Dutch urban area. Environmental Impact Assessment Review 29:179-185

Beckx C, Int Panis L, Uljee I, Arentze T, Janssens D, Wets G (2009b) Disaggregation of nation-wide dynamic population exposure estimates in The Netherlands: Applications of activity-based transport models. Atmospheric Environment 43:5454-5462

Beckx C, Int Panis L, Van De Vel K, Arentze T, Lefebvre W, Janssens D, Wets G (2009c) The contribution of activity-based transport models to air quality modelling: A validation of the ALBATROSS-AURORA model chain. Science of The Total Environment 407:3814-3822

Davidson W, Donnelly R, Vovsha P, Freedman J, Ruegg S, Hicks J, Castiglione J, Picado R (2007) Synthesis of first practices and operational research approaches in activity-based travel demand modeling. Transportation Research Part A 41:464-488

Int Panis L (2010) New Directions: Air pollution epidemiology can benefit from activity-based models. Atmospheric Environment 44:1003-1004

McNally MG (2000) The Activity-Based Approach, UC Irvine: Center for Activity Systems Analysis

Shiftan Y (2000): The Advantage of Activity-based Modelling for Air-quality Purposes: Theory vs Practice and Future Needs. Innovation 13:95-110

WEDNESDAY 15 SEPTEMBER 2010

ORAL PRESENTATIONS

GEOSTATISTICAL INTERPOLATION AND SIMULATION USING COPULAS

A. Bárdossy

Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, D-70569 Stuttgart, Germany

In geostatistical applications the spatial dependence structure is traditionally described with the help of variograms or covariance functions. These however are strongly influenced by the distribution of the related variables under investigation. Variograms calculated from highly skewed data are often dominated by a few outlying observations. Indicator geostatistics can be used to overcome this problem. Indicator variograms often reveal specific dependence structures, for example very different ranges for high or low values. Indicator kriging or indicators simulations can take these differences into account, but the theoretical variogram parameters have to be estimated for each considered cutoff. Further due to the incompleteness of the model the estimated distribution functions might require corrections to become monotonic.

The transformation of the observed values to a uniform distribution using their marginal distribution function offers another possibility. Multivariate distributions with uniform marginals are called copulas. Their formal definition of a copula is:

$$C : [0,1]^n \rightarrow [0,1] \tag{1}$$

$$C(\mathbf{u^{(i)}}) = u_i \quad \text{if} \quad \mathbf{u^{(i)}} = (1,\ldots,1,u_i,1,\ldots,1)$$

$$\sum_{j=0}^{2^n-1} (-1)^{\sum_{i=1}^{n} j_i} C(u_1 + j_1\Delta_1,\ldots,u_n + j_n\Delta_n) \geq 0$$

$$\text{if } 0 \leq u_i \leq u_i + \Delta_i \leq 1 \text{ and } i = \sum_{k=0}^{n-1} j_k 2^k$$

which means that probability corresponding to any *n* dimensional hypercube has to be non negative.

Copulas can describe the dependence of multivariate distributions with any kind of marginal distributions. Sklar proved that each continuous multivariate distribution $F(x_1; : : : ; x_n)$ can be represented with the help of their copula *C* and the corresponding marginal distributions:

$$F(t_1,\ldots,t_n) = C(F_{t_1}(t_1),\ldots,F_{t_n}(t_n)) \tag{2}$$

where $F_{t_i}(t)$ represents the *i*-th one dimensional marginal distribution of the multivariate distribution. If the marginal distributions are continuous then the copula *C* in (2) is unique. Hence, copulas can be regarded as the expression of the dependence without the influence of the marginal distributions.

Copulas are frequently used to describe multivariate dependence in problems encountered in finance and the economy. In geostatistics the use of the copulas offers a possibility to describe spatial dependence independently from the marginal distributions. In place of variograms and covariance functions, bivariate empirical copulas, corresponding to separation distances and directions are calculated and used to gain insight into the dependence structure. The data are transformed using the distribution function F of the observed values to the uniform distribution. These transformed data are then directly used to construct bivariate empirical copulas.

For the quantification of the dependence structure empirical rank correlation functions, and measures of asymmetrical dependence can be calculated from these empirical copulas to quantify the strength and the form of spatial dependence. The asymmetry of a copula can be defined as:

$$A = \int_0^1 \int_0^1 (u - \frac{1}{2})(v - \frac{1}{2})(u + v - 1)c(u, v)dudv \qquad (3)$$

where *c* is the density of the copula *C*.

The asymmetry can be calculated from the observed values corresponding to different vectors *h*:

$$A^*(h) = \frac{1}{N(h)} \sum_{x_i - x_j \approx h} (F(z(x_i)) - \frac{1}{2})(F(z(x_j)) - \frac{1}{2})(F(z(x_i)) + F(z(x_j)) - 1) \quad (4)$$

As the Gaussian is symmetrical $A = 0$ the asymmetry function can be used to detect deviations from the Gaussian dependence. Simulation can be used to find the confidence limits for the asymmetry of a Gaussian dependence - which then can be used to check if the Gaussian dependence can be rejected. The traditional normal score transformation assumes a Gaussian copula.

Examples from groundwater quality parameters and rainfall data show how the real inter-dependence can be understood using copula-based measures in contrast to how it is distorted through skewed marginals when using variograms and covariance functions. Easy to visualize bivariate copulas often show unusual dependence of the variables for different ranges of the investigated values. The dependence is often asymmetrical; for example high values might be strongly related, while low values are only weakly linked.

Many natural variables, such as topographical surfaces, rainfall, groundwater quality parameters and soil hydraulic conductivity, show asymmetrical spatial dependence. If an asymmetrical dependence is observed, a normal score transformation of the variable is not appropriate, so that other multivariate assumptions have to be applied.

In order to perform interpolation, unconditional or conditional simulations a theoretical copula model has to be selected and its parameters have to be estimated from the observations. The multivariate copulas which can be used to describe the spatial dependence have to fulfill a few assumptions. These include spatial stationarity of the dependence structure and invariance of the multivariate distributions with respect to their multivariate marginals. Furthermore the multivariate copula should allow the spectrum from full dependence through to independence of the corresponding variables.

Non-Gaussian copula models which fulfill the above conditions can be obtained through selected transformations. One possibility is to use non monotonic transformations of the marginals of a multivariate Gaussian distribution, and to use the copula of the resulting distribution. Another possibility is to transform the copula densities of known multivariate distributions using periodic functions. Both approaches allow the description of the dependence using only a few additional parameters. On the other hand the calculation of the corresponding distributions is numerically demanding. The copula parameters can be estimated using a partition of the observations into multipoint subsets followed by an appropriate maximization of the corresponding likelihood function.

Interpolation of arbitrary quantiles can be performed using copulas. The fully conditional copulas provide the distribution functions for estimation at the unobserved location. They can be used to calculate medians, means or to define confidence intervals which depend on both the observed data values and geometry. The quality of the estimators and the conditional distributions may be verified using cross validation or split sampling approach. Using the full estimated conditional distributions a sequential conditional simulation can be set up.

Maps obtained by copula based interpolation often do not differ considerably from other geostatistical interpolation results. The main difference is in the corresponding confidence intervals which provide a realistic estimator for the interpolation uncertainty.

Examples, ranging from small scale soil samples through regional aquifers, large scale weather fields and landscapes from Mars are used to demonstrate the occurrence of non-Gaussian fields and to show the steps of a copula based investigation of spatial data.

ACCOUNTING FOR UNCERTAINTIES AND CHANGE OF SUPPORT IN SPATIO-TEMPORAL MODELLING OF INDIVIDUAL EXPOSURE TO AIR POLLUTION

Lydia E. Gerharz*, Edzer Pebesma
Institute for Geoinformatics, University of Münster, Germany

Individual exposure modelling

Relations between negative health effects like cardio-vascular diseases, asthma and lung cancer and elevated levels of air pollution are detected in several large scale exposure studies (e.g. Brunekreef and Holgate, 2002). However, correlations found in these studies are based on summary health statistics and annual air quality measurements over a city. It is likely that the causality on the individual level looks differently when the inhaled dose of the airborne pollutant could be related to direct health outcomes. Instead of performing time- and resource-consuming personal measurements, modelling of individual exposure to air pollution is commonly used to estimate the true exposure. This requires the combination of two types of information: concentration fields and individual trajectories. Whereas the first one is usually obtained by applying deterministic models to estimate the spread of air contaminants, the latter one is increasingly gathered by tracking devices such as GPS receivers providing precise positions in space and time for the trajectories, e.g., of persons difficult to track by diaries (Elgethun et al., 2007). Having data of individuals on such a high spatio-temporal resolution requires air pollution data on the same resolution. This yields the problem of higher uncertainties for the air pollution modelling at the higher resolution due to limitations of modelling techniques and data availability. Here, the question turns up on which scale or aggregation level in space and time we can still make sensible conclusions and reveal significant differences in the exposure between individuals, given uncertainties in air pollution estimates. In this paper, we will focus on the uncertainties and the effect of combining data at different scales when modelling individual exposure to air pollution.

Representation and sources of uncertainty

The concentration $C(t,s)$ of an air pollutant is a spatio-temporal variant phenomenon and is commonly represented by modelled or interpolated grids for discrete time steps, with the grid cell value representing a spatially and temporally averaged estimate of the true concentration. For $C(t,s)$ attribute uncertainties are often dominant to positional and temporal uncertainties due to the relatively coarse scale of modelling. Excluding explicit street canyon models, the scale of air quality models is usually a few hundreds of meters and larger than 15 min. The main uncertainties are caused by the imperfect dispersion or interpolation model and assumptions that have to be drawn throughout the model process.

Individual trajectories $I(t,s)$ are space-time paths as defined by Hägerstrand (1970). Their representations used for exposure modelling are discrete points (locations) in space and time, e.g., GPS tracks. By definition only positional and temporal uncertainties can occur for $I(t,s)$, as the only attributes are position and time. For GPS data the spatial uncertainty usually lies within 3 to 5 m and the temporal < 1 sec.

*Corresponding author: Lydia.Gerharz@uni-muenster.de

Quantified uncertainties are characterised by their probability distribution functions which are joint probabilities if the variables are correlated which is the case for both data types (Heuvelink et al., 2007). By combining these two datasets, it is obvious that in this case the uncertainty of the GPS tracks is negligible compared to that of the concentration estimates. Furthermore, the uncertainty of the concentration is the uncertainty of the grid cell average which cannot be simply transferred to the individual exposure as shown in the next section.

Change of support

Drawing conclusions from the relation of two data sets with different scales leads to the problem of ecological inference. Robinson (1950) showed that a correlation of two aggregated variables does not have to be valid for the individual variables. Thus, in our case, the causality between health burdens of urban populations and elevated ambient air pollution does not allow conclusions about the causality between the individual exposure and health effects of a random inhabitant from the cities. Young and Gotway (2007) classified the ecological inference problem as a special case of the change of support problem. Spatial support is defined as the physical size, shape and orientation of the considered variable. In the case of exposure modelling the support of the air pollution model is the size of the grid cells. Individual position and exposure are very small and treated as point support. Thus, the support of $C(s,t) >> I(s,t)$. Aggregation of the individual positions is not feasible as usually only a small fraction of possible points within a grid cell are visited. Therefore, disaggregation of the cell average is necessary to estimate concentration at point locations which demands further information about the variability within the grid cell.

In fact, the uncertainty of the exposure experienced in a grid cell crossed by an individual depends on the within-variability of the concentration in the grid cell. The within-variability can be described by a geostatistical model of spatial autocorrelation $\rho(h)$ for different lags h between the points. A stronger autocorrelation between points up to lags h larger than the grid cell size yield less within-variance for the cell so the cell mean is more representative for the point values. In the extreme case of $\rho(h)=1$ for all lags, the points correlate perfectly. The cell mean would be the same as for the points and the uncertainty of the cell mean is the only uncertainty of the point values. The other extreme $\rho(h)=0$ would mean that all points are varying completely at random (white noise) within the cell, which would yield for individual points an uncertainty equally to the within-grid cell variance times the uncertainty of the mean. For the latter case aggregating over a larger number of points per grid cell would average out the uncertainty completely whereas for the former case the uncertainty between simulations remains large.

Case study/Illustrative example

As an illustrative example for the change of support effect we combined ambient air quality data from a deterministic model with individual trajectories, recorded by GPS. Fig. 1 shows a map for Münster, Germany with estimated 1 hour averaged PM10 concentrations for 250 x 250 m² grid cells and a 1 h GPS track with tracked positions every 5 sec. Uncertainties of the cell average and temporal uncertainties were not included here.

The spatial autocorrelation of the points in each cell is assumed to be only dependent on the distance *h* between the points and is represented by:

$$\rho(h) = exp\left(\frac{-h}{a}\right)$$

Three exponential variogram models were created using the variance σ^2 estimated from a set of mobile measurements and three different strengths for $\rho(h)$

$$\gamma(h) = C(0) - C(h) = \sigma^2 - \sigma^2 \rho(h)$$

For each GPS point 1000 samples were calculated using unconditional Gaussian simulation.

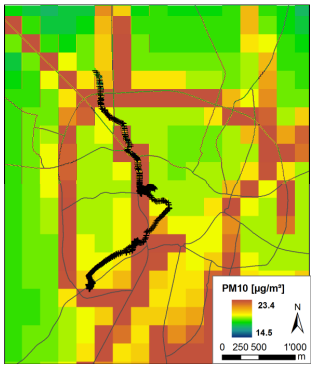


Figure 1. PM10 map of Münster with GPS track.

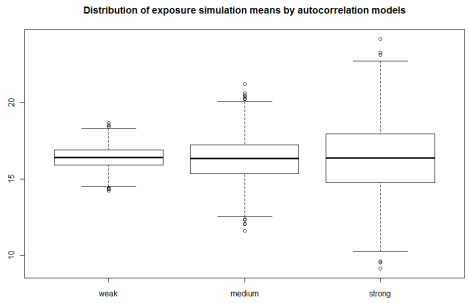


Figure 2. Distribution of exposure means over the whole GPS track for each simulation using variogram models with different spatial autocorrelation (weak: a=10;medium: a=50; strong: a=200).

The results for the GPS track in fig. 2 show clearly an increase in variability for the simulation averages with increasing strength of autocorrelation. Thus, averaging over a larger number of points reduces uncertainty faster if autocorrelation is weak. However, for a single point in a grid cell, the uncertainty depends mainly on the within-grid cell variability. Methods to assess the spatio-temporal autocorrelation and within-grid cell variances by using covariates for prediction are essential to reduce uncertainties in individual exposure modelling and are under current development.

References

Brunekreef B, Holgate ST (2002) Air pollution and health. Lancet 360:1233–1242

Hägerstrand T (1970) What about people in regional science? In: Papers of the Regional Science Association 24:7-21

Heuvelink GBM, Brown JD, van Loon EE(2007) A probabilistic framework for representing and simulating uncertain environmental variables. Int J Geogr Inf Sci 21:497-513

Elgethun K, Yost MG, Fitzpatrick CTE, Nyerges TL, Fenske RA (2007) Comparsison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns. J Expo Sci Env Epid 17:196-206

Robinson WS (1950) Ecological correlations and the behavior of individuals. Am Sociol Rev 15:351–357

Young LJ, Gotway CA (2007) Linking spatial data from different sources: the effects of change of support Stoch Environ Res Risk Assess 21:589-600

APPLYING AN EXHAUSTIVE SECONDARY DATASET FOR GEOSTATISTICAL ANALYSIS OF AIR POLLUTION

D.P. Robinson[1], C.D. Lloyd[1] and J.M. McKinley[1]

[1]School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast, N.Ireland, drobinson13@qub.ac.uk +44(0)28 9097 3929

Introduction/Background

Oxides of nitrogen (NO_x) can have a negative health impact on persons already susceptible to respiratory illness. NO_x is defined as a group of pollutants, one exemplar pollutant being nitrogen dioxide (NO_2). These pollutants have a strong link with the fuel combustion processes and secondary pollutant formations associated with transport networks (Carslaw and Beevers 2005). Dispersion modelling is currently used to generate mapped air pollution outputs in the UK. It is hypothesised that geostatistical analysis can aid the mapping process by accounting for the potential variation not explained through dispersion modelling. The study area was the UK and the primary dataset consisted of NO_2 data from automatic monitoring and NO_2 data derived using diffusion tubes. Annual mean levels (measured in μgm^{-3}) for each location were derived from automatic monitoring and diffusion tubes. The secondary dataset consisted of NO_x data (measured in μgm^{-3}) derived from dispersion modelling of emissions. It is available from UK Air Quality Archive (www.airquality.co.uk) and appears as a 1km × 1km gridded output for the study area. Dispersion modelled data were available at 1km × 1km scale for the UK, making this an exhaustive secondary dataset. The locally strong relationship between NO_x and NO_2 means that regression can be used in conjunction with geostatistics with the potential to generate more accurate pollution maps than can be provided by dispersion modelling alone. Previous work using such techniques includes Goovaerts (2000) and Lloyd and Atkinson (2004).

Methods

Variograms for geostatistical analysis were estimated and modelled using the Gstat software (Pebesma and Wesseling 1998). Ordinary Kriging (OK) (Isaaks and Srivastava 1989) was performed on the primary dataset as a 'baseline' approach. Simple Kriging with locally varying mean (SKlm) (Goovaerts 1997) was examined using different methods for deriving the varying local means. Global regression (GR) and geographically weighted regression (GWR) using various bandwidths (Fotheringham et al. 2002) were performed, with the corresponding predictions serving as the varying local means. Moving window regression (MWR) and global ordinary least squares (OLS) were also investigated for their use as locally varying means (Lloyd 2006). Alternative geostatistical methods to SKlm explored were collocated co-kriging (Collocated CK) and kriging with an external drift (KED) (Goovaerts 1997). Routines from Geostatistical Software Library (GSLIB) (Deutsch and Journel 1998) were used for kriging. Cross validation prediction errors were computed for comparison of techniques.

Hybrid approaches can be beneficial, particularly in examples such as this case study where the primary dataset is sparse and the secondary dataset is exhaustive. Hengl et al. (2007) highlighted the potential value of hybrid approaches, such as kriging and linear regression and expressed the benefits through numerous case studies. This study demonstrated the versatility of hybrid approaches and their potential cross-disciplinary application. Such approaches are further verified by the work of Lloyd and Atkinson (2004) and Goovaerts (2000) who used multivariate hybrid approaches in their analyses, yielding positive increases in accuracy in comparison to other techniques. Following on from the work of the aforementioned authors this study incorporated hybrid approaches.

Results

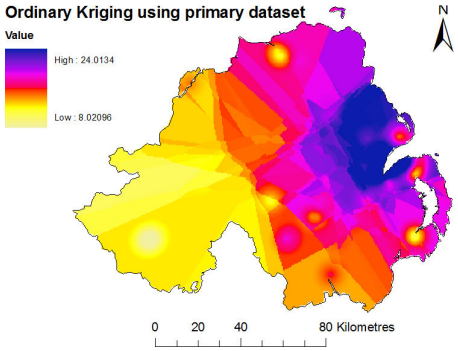


Figure 1: *OK map of NO_2 from primary dataset in N.Ireland*

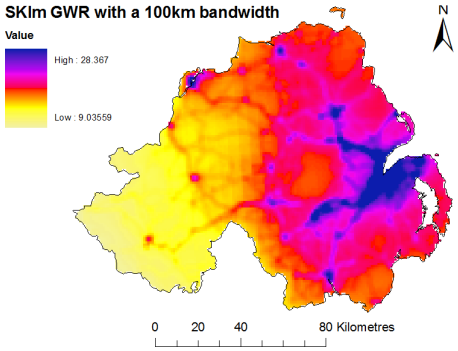


Figure 2: *SKlm map of NO_2 with local means derived from GWR with a 100km bandwidth (using both primary and secondary datasets) in N.Ireland*

Discussion

The results section displays estimates based on data for N.Ireland. Figure 1 shows the NO_2 pollution map derived from the primary dataset, using OK. The highest concentrations are located in the Belfast Urban Area (in the mid east area) and in the south and east of county Antrim. It can be observed from the map that there are too few data points in the primary dataset to produce an accurate representation of air pollution in the region. The root mean square error (RMSE) for OK was 5.528 μgm^{-3}, which is larger than the RMSE value for SKlm with local means derived using GWR with a 100km bandwidth, which was 3.878 μgm^{-3}. Similarly other cross validation error statistics such as the maximum, minimum, mean and standard deviation are all greater than the equivalent values for SKlm GWR with a 100km bandwidth (or further from zero in the case of the mean). Figure 2 shows the NO_2 pollution map derived from using the exhaustive secondary dataset in conjunction with the primary dataset, using SKlm with local means derived using the local regression method GWR. This figure is intuitively more realistic than Figure 1, as the road network is clearly visible, which would be expected of a pollutant whose formation and production have strong links to the transport network. The benefits of having the exhaustive secondary dataset are clear both

intuitively through the visual comparison of both concentration maps, but also through the cross validation prediction error statistics.

References

Carslaw DC, Beevers SD (2005) Estimations of road vehicle primary NO_2 exhaust emission fractions using monitoring data in London. Atmospheric Environment 39:167-177.

Deutsch CV, Journel AG (1998) GSLIB: Geostatistical Software and User's Guide. 2nd edn. Oxford University Press, New York.

Fotheringham AS, Brunsdon C, Charlton ME (2002) Geographically Weighted Regression, the analysis of spatially varying relationships. Wiley, United Kingdom.

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York.

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology 228:113–129.

Hengl T, Heuvelink GBM, Rossiter DG (2007) About regression-kriging: From equations to case studies. Computers and Geosciences 33:1301-1315.

Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics. Oxford University Press, New York.

Lloyd CD (2006) Local models for spatial analysis. CRC Press, Boca Raton.

Lloyd CD, Atkinson PM (2004) Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. International journal of applied earth observation and geoinformation 5:293-305.

Pebesma EJ, Wesseling CG (1998) Gstat, a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24:17-31.

SPATIOTEMPORAL GEOSTATISTICS IN CROSS-METHODOLOGICAL MONITORING OF AIR POLLUTION

A. Kolovos [1], A. Skupin [2], W. Pang [2], G. Christakos [2], M. Jerrett [3]

[1]SAS Institute, Inc, 100 SAS Campus Dr., Cary, NC 27513, USA, Email: alexander.kolovos@sas.com, Telephone: +1-919-531-2165, Fax: +1-919-677-4444
[2] Department of Geography, San Diego State University, San Diego, CA 92182-4493, USA,
[3] Department of Env. Health Sciences, School of Public Health, University of California, Berkeley, CA 94720-7360, USA

Abstract

This study investigates creative, cross-disciplinary interaction of spatiotemporal geostatistics with a non-geostatistical methodology. Spatiotemporal prediction is employed to provide the foundation for the so-called spatialization transformation in the field of geographic analysis. This sequence of steps begins with the geostatistical analysis of monthly-averaged nitrogen dioxide (NO2) and mean annual sulfate (SO4) observations in California over a 15-year study period between 1988 and 2002, and results into a lower-dimensional geometric representation of these pollutants. This composite approach combines the Bayesian Maximum Entropy (BME) methodology to perform geostatistical prediction, and spatialization to make high-dimensional spaces accessible to human cognition through processes of dimensionality reduction and symbolization. The proposed blending reveals new patterns and relationships suggested by the original data, provides new perspectives and broadens our understanding of environmental attributes, as shown in the present study of the NO2 and SO4 air pollutants.

Data and BME spatiotemporal prediction

The California Air Resources Board (CARB) provided two data sets of monthly-averaged NO2 measurements from 137 monitoring sensor locations, and annual averages of SO4 measurements from 166 monitoring sensor locations for the 15 years between 1988 and 2002. A separate geostatistical analysis was performed for each pollutant data set to predict the pollutant values on a spatiotemporal regular grid. The spatial grid encloses the state of California and consists of 21x24 spatial nodes, and the temporal grid consists of 180 monthly temporal instances for NO2 and 15 annual temporal instances for SO4. The predictions are used consequently as the input for the spatialization analysis.

The geostatistical analysis step enabled us to overcome the data sparseness and heterogeneity in space and time, to explore the underlying correlations in the pollutant observations, and to provide predicted values on the specified grid. To this end, we used the BME methodology, which has been employed in atmospheric studies (e.g., Bogaert et al. 2009; Christakos et al., 2004; Christakos and Kolovos, 1999) for accurate prediction of atmospheric attributes across space and time by successfully assimilating input from sensor measurements, a variety of certain and uncertain observations, and physical laws. In the present study, the prediction results yield informative maps of the two pollutants and their distributions at all locations in

space and time on the specified grid. For illustration, Figure 1 displays maps of the mean of the predicted distributions (*BMEmean*) for both pollutants at selected instances.

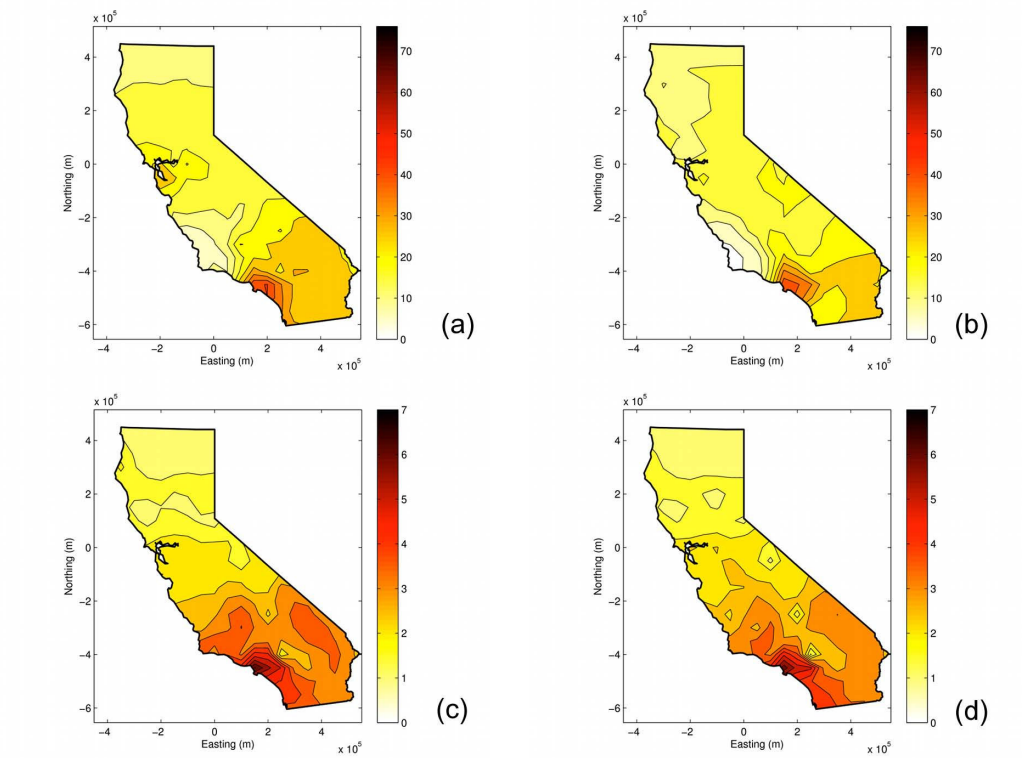


Figure 1: *BMEmean for monthly-averaged NO2 and annual-averaged SO4 concentrations across California at selected time instances. The pollutant concentrations are in ppb units. The plot shows the NO2 predicted distributions means at months (a) January 1988 and (b) July 1988, and the SO4 predicted distributions means at years (c) 1988 and (d) 1989.*

Spatialization analysis

The gridded geostatistical output makes it possible for the spatialization methodology to apply its analytical techniques for the reduction of its input into a lower-dimensional space (Skupin and Fabrikant 2008). In particular, the original multiplicity in dimensions comes from dealing with sensors at *L* geographic locations, and taking measurements of *A* attributes at *T* different times. This leaves us with a *single* set of *L*A*T* observations (an *LAT* data set). One can now systematically explore this *tri-space* of *LAT* values by constructing a series of matrices in which rows and columns are constructed from combinations of tri-space elements.

Depending on what combinations are considered, the same set of *L*A*T* observations could be transformed into six different matrices of the *LAT* source data by rearranging the original input values. Each matrix allows different questions to be explored about the data and the similarities of location-attribute composites; for example, one could ask whether NO 2 concentrations at one location have behaved over time similar to SO4 locations at another location. To answer such questions, the dimensionality of row vectors is almost always too large to be directly plotted. The method of self-organizing maps (SOM) is used to perform a

type of dimensionality reduction when dealing with very voluminous and high-dimensional data sets (Kohonen 2001). An example of dimensionality reduction with spatialization is shown in Figure 2, which presents monthly NO2 values for 1988-2002 as 12-dimensional vectors. The component planes in Figure 2 illustrate the dominant pattern of November through January as corresponding to the highest NO2 concentrations, no matter what the absolute values are.

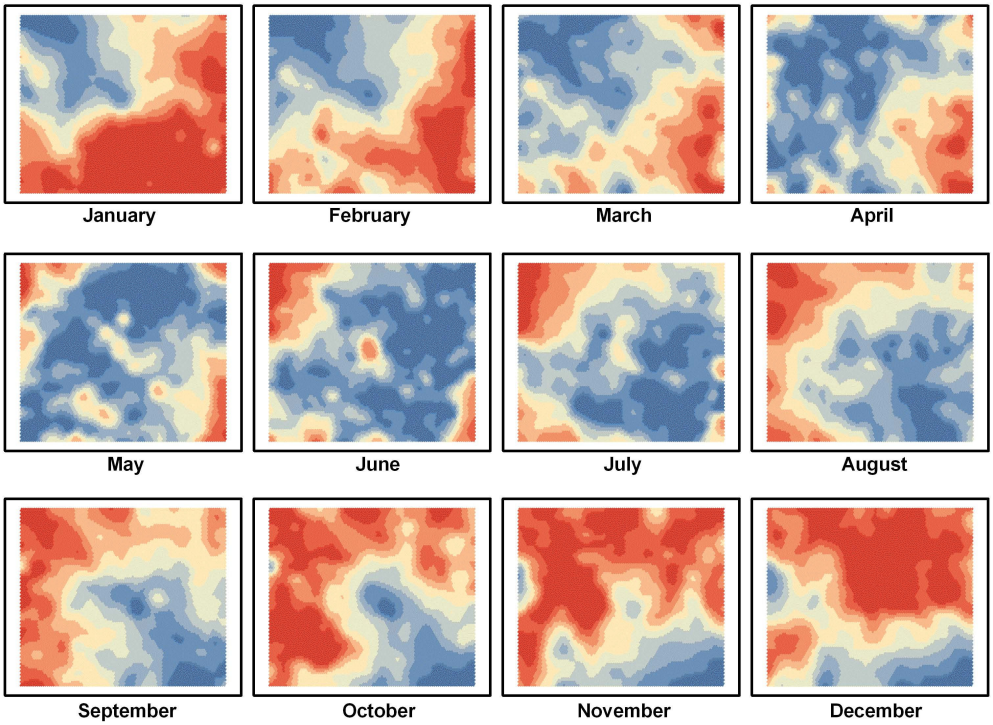


Figure 2: *Monthly NO2 values for 1988-2002 modeled as 12-dimensional vectors, with each cell contributing 15 separate vectors. The 12 component planes are shown, with low-high values indicated as a blue-red color range.*

Acknowledgement

The research was supported by the California Air Resources Board grant No. 55245A.

References

Bogaert P, Christakos G, Jerrett M, Yu H-L (2009) Spatiotemporal modelling of ozone distribution in the State of California. Atm. Envir. 43(2009):2471–2480.

Christakos G, Kolovos A, Serre M L, Vukovich F (2004) Total Ozone Mapping by Integrating Data Bases From Remote Sensing Instruments and Empirical Models. IEEE Trans. on Geosc. and Rem. Sensing 42[5]:991-1008.

Christakos G, Kolovos A (1999) A study of the spatiotemporal health impacts of the ozone exposure. J. Exposure Anal. and Env. Epidemiol. 9:322-335.

Skupin A, Fabrikant S I (2008) Spatialization. The Handbook of Geographic Information Science. J. P. Wilson and A. S. Fotheringham. Blackwell Publishing:61-79.

Kohonen, T. (2001). Self-Organizing Maps. Berlin, Springer-Verlag.

ACCOUNTING FOR SPATIO-TEMPORAL UNCERTAINTY ASSOCIATED TO EMISSION ALLOCATION IN AN ENERGY-AIR QUALITY ASSESSMENT MODEL

U. Leopold[1], L. Drouet[1], D.S. Zachary[1]

[1] Resources Centre for Environmental Technologies, Public Research Centre Henri Tudor, P.O. BOX 144, 4002 Esch-sur-Alzette, Grand Duchy of Luxembourg, Email: Ulrich.Leopold@Tudor.lu, Tel.: +352-425991-618, Fax: +352-425991-555.

This paper presents a methodology to model spatio-temporal uncertainties associated to the dis-aggregation of sectoral emissions of primary air pollutants in Luxembourg. We apply it to an integrated energy-air quality assessment tool for the urban and regional scale (Zachary et al., 2010). National aggregated sectoral emissions of primary air pollutants are computed by an energy model which minimises the cost of the reference energy system. The model describes five sectors, i.e. agriculture, transport, industry, residential and commercial at five-year periods. The sectoral emissions are allocated in space and time to obtain hourly to daily emission maps. The air quality model simulates the dispersion of the emitted pollutants and their chemical reactions to produce ozone for typical episodes for each five-year period. Both models are coupled by means of an oracle-based optimisation method to find the optimal energy system with a constraint on ozone concentrations.

When dis-aggregating emissions from the national scale to the urban scale, small scale variation and its associated uncertainty within and across land-use boundaries have to be modelled in order to obtain realistic spatio-temporally distributed emission maps. In a previous work we modelled the spatial uncertainty without considering a temporal correlation (Leopold et al., 2010). Now, we propose to extend it to the temporal approach.

We decompose the emission value into its mean, standard deviation and a spatio-temporally correlated error term. The mean and standard deviation depend on land use and transportation whereas the error has zero mean and unit variance and has a variogram that only depends on distance in time and space between points. The spatio-temporal error can be modelled by stochastic simulation assuming a space-time semi-variogram with local and regional scale as well as short-term and long-term scale variability to account for within-boundary and across-boundary variation as well as hourly and daily variation (Bierkens et al., 2000). A Monte Carlo approach is used to simulate the spatio-temporally correlated error and its distribution for each grid cell at each time step.

Finally, the average dis-aggregated emission value and its corresponding ``local'' uncertainty can be computed for each grid cell accounting for space-time correlation at different scales and across land-use boundaries. The underlying assumptions made, influence the results of the dis-aggregation approach. Different assumptions are compared and discussed. Once

uncertainties are assessed, the dis-aggregated emissions and its associated space-time uncertainties can be incorporated in an uncertainty propagation analysis of the air-quality model or a stochastic version of the energy optimisation algorithm.

References

Bierkens, M.F.P., Finke, P.A. and de Willigen, P. (2000). Upscaling and downscaling methods for environmental research: Developments in Plant and Soil Sciences, vol. 88, Kluwer Academic Publishers, Dordrecht.

Leopold, U., Zachary, D.S., Drouet, L. (2010). Modelling Spatial Uncertainty associated to Emission Allocation in an Energy-Air Quality Assessment Model, Accuracy 2010 Conference. Leicester, England.

Zachary, D.S., Leopold, U., Drouet L., Aleluia Reis, L. (2010). An integrated geospatial energy - air quality impact assessment tool, Draft for Environmental Research Letters (2010).

MERGING AREAL AND POINT DATA IN GEOSTATISTICAL INTERPOLATION: APPLICATIONS TO SOIL SCIENCE AND MEDICAL GEOGRAPHY

P. Goovaerts

BioMedware, Inc. 3526 W Liberty, Suite 100, Ann Arbor, MI 48103, USA
goovaerts@biomedware.com, (734) 913-1098, (734) 913-2201

Introduction

A common issue in spatial interpolation is the combination of data measured over different spatial supports. For example, information available for mapping disease risk typically includes point data (e.g. patients residence) and aggregated data (e.g. socio-demographic and economic data at the census track level). Similarly, soil measurements at discrete locations in the field are often supplemented with choropleth maps (e.g. soil or geological maps) that model the spatial distribution of soil attributes as the juxtaposition of polygons (areas) with constant values. This paper presents a coherent geostatistical framework to accommodate both areal and point data in the spatial interpolation of continuous attributes. The procedure is illustrated using two datasets: 1) geological map and heavy metal concentrations recorded in the topsoil of the Swiss Jura, and 2) incidence rates of late-stage breast cancer diagnosis per census tract and location of patient residences in Michigan (Figure 1).

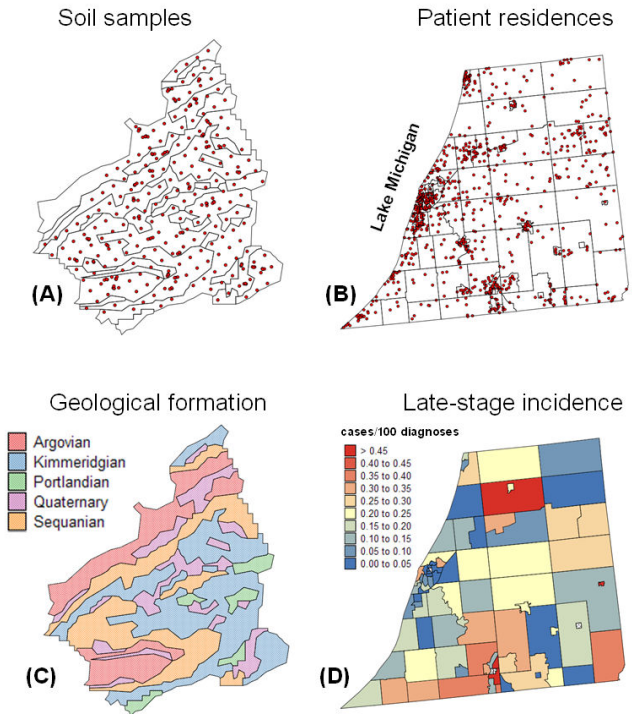


Figure 1: *Information available for mapping topsoil heavy metal concentration and late-stage breast cancer incidence. (A) Soil field measurements. (C) Choropleth map of the main geological formations. (B) Location of 937 patient residences. (D) Choropleth map of late-stage breast cancer incidence rate in three Michigan counties, by census tract, 1985-2002.*

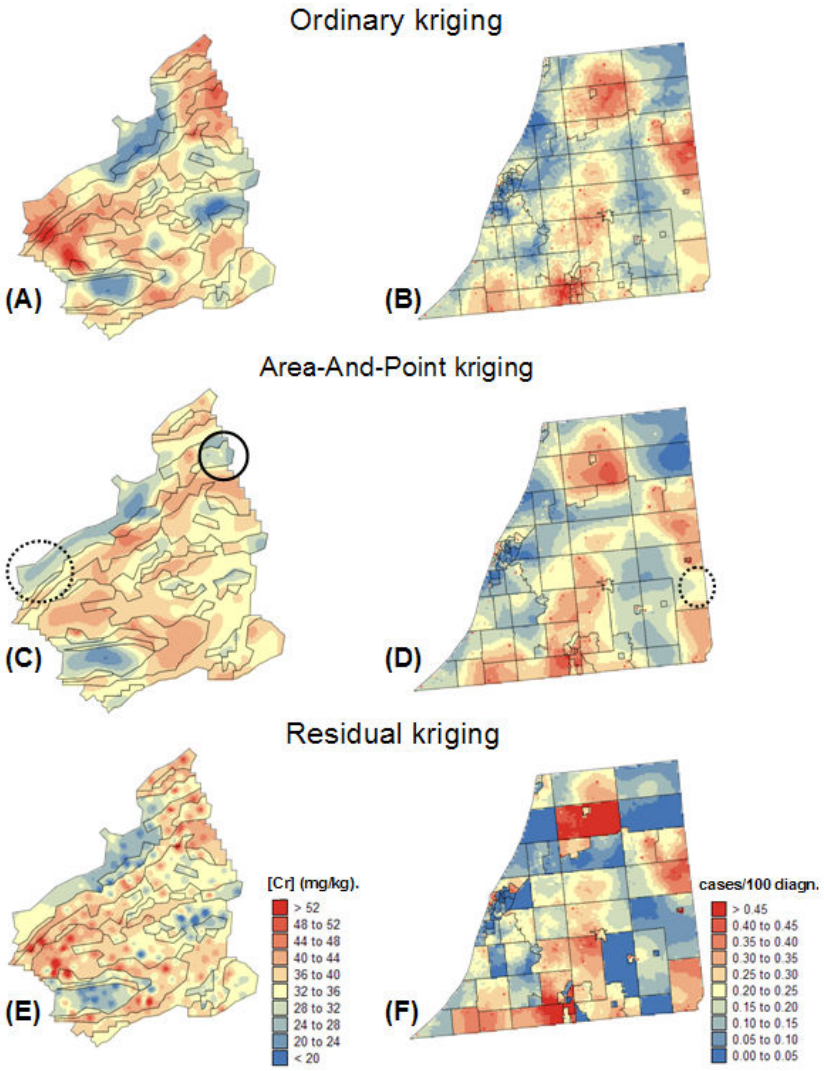


Figure 2: *Maps of chromium concentration and late-stage breast cancer incidence rate created by alternative interpolation techniques. (A,B) Ordinary kriging. (C,D) Kriging that combines both areal and point data "AAP kriging". (E,F) Residual kriging with a choropleth trend model. The same color scale is used for each series of three maps.*

Results and discussion

The interpolation approach capitalizes on: 1) a general formulation of kriging that allows the combination of both point and areal data through the use of area-to-area, area-to-point, and point-to-point covariances in the kriging system, 2) the availability of GIS to discretize polygons of irregular shape and size, and 3) knowledge of the point-support semivariogram model that can be inferred directly from point measurements, thereby eliminating the need for deconvolution. For health data, the kriging system includes an error variance term derived according to the binomial distribution to account for varying degree of reliability of incidence rates depending on the total number of cases recorded in those tracts.

Fig. 2 (left column) shows the maps of chromium concentration estimated using alternative interpolation techniques. The reference approach is ordinary kriging (OK) that

uses only field data (Fig. 2A). The other two maps incorporate areal data that take the form of average chromium concentration per geological mapping unit. These concentrations were used either as local means in residual kriging or directly incorporated into the Area-and-point estimator. In the later case, the average of kriged estimates equals the mapping units' mean (coherence constraint). The residual semivariogram model has a short range, leading to "bull's-eye" effect around sample points in the map created by residual kriging (Fig. 2E). In contrast, the AAP map (Fig. 2C) is much smoother and clearly displays the lower concentrations expected on the Argovian formation. Differences between the three maps are the largest in sparsely sampled areas where the choice of a trend model becomes preponderant. In particular, incorporating the geological information leads to smaller estimates on the section of Argovian formation where no sample was collected (dashed circle in Fig. 2C) and in a small Argovian mapping unit that must satisfy the coherence constraint despite the presence of larger concentrations recorded in the field (solid circle in Fig. 2C).

A similar analysis was conducted for the health outcome data in Figs. 1B-D. All incidence maps were created using the 32 closest point indicator data and, for AAP kriging, the rates recorded in census tracts that share a boundary or vertex with the tract including the interpolation node (1st order adjacency). Incorporating census-tract information through residual kriging adds more details to the map but generates discontinuities at the tract boundaries. On the other hand, accounting for adjacent areal data in AAP kriging leads to a map that has more compact spatial features than the indicator kriging map.

The performance of the proposed approaches, relatively to ordinary kriging or a traditional residual kriging with choropleth trend model, is assessed using jackknife. Performance criteria include the magnitude of prediction errors, the accuracy of the model of uncertainty, the smoothness of interpolated maps, and the ability to discriminate between early and late-stage cancer cases. Results demonstrated the overall better prediction performance of AAP kriging over ordinary kriging and residual kriging with the choropleth map trend model. In particular when sampling is sparse, incorporation of areal data improves the prediction accuracy while the exactitude property of areal data decreases the smoothness of interpolated surfaces.

Conclusions

The ability to combine data measured at various scales and over different spatial supports in kriging is becoming a pressing need, in particular as the field of geostatistical applications now encompasses social and health sciences. Whereas the first analytical developments of kriging clearly demonstrated its flexibility to accommodate different measurement and prediction supports, geostatistical analysis of a mixture of point data and irregular blocks has rarely been implemented in practice, mainly because of its lack of application in mining. Joint advances in GIS software and computational resources now allow the application of kriging to the complex geographies found in social and health sciences. In addition, the recent development of binomial and Poisson kriging allows one to take into account both the spatial extent of the geographical unit and the size of the population under study within that unit (i.e. number of breast cancer cases) in the interpolation.

DISTRIBUTION OF OVERWEIGHT AND IDENTIFICATION OF SPATIAL RISK

L.B. Nucci [1]; A.C.C.N. Mafra [1]; C. Stephan [1]; L.T.O Zangirolani; R. Cordeiro [1]

[1] State University of Campinas, Campinas, São Paulo, Brazil 13025-061 Phone: +55 (19) 9227.9105 e-mail: lbnucci@terra.com.br

The recent increase in the prevalence of obesity is widely recognized as constituting a major threat to health in most countries, as obesity reached epidemic proportions globally, with more than 1 billion adults overweight according to the World Health Organization (WHO) (World Health Organization (WHO) 2010). The prevalence of overweight and obesity is commonly assessed by using body mass index (BMI), defined as the weight in kilograms divided by the square of the height in meters (kg/m^2). A BMI over $25kg/m^2$ is defined as overweight, and a BMI of over $30\ kg/m^2$ as obese, resulting in a response variable with more than two categories (World Health Organization (WHO) 1995; World Health Organization (WHO) 2000; Ogden et al. 2007).

The number of studies on obesity increased significantly, mainly by the high risk that this condition leads to patients with other chronic diseases such as diabetes and hypertension (Mokdad et al. 2001; Resta et al. 2001; Peeters et al. 2003). Therefore, studies about the risk of obesity, to consider factors other than those already widely studied, have been proposed. As an example we can cite the inclusion of location variables to identify obesogenic environment (Drewnowski et al. 2007; Zangirolani 2009) and studies on food supply (Hill et al. 1998; Mokdad et al. 2001; Young et al. 2002).

Although risk measurement is widely used in epidemiology, the estimates of relative risk for variables with more than two categories is not yet part of the statistical analysis for many studies in this area, mainly when we use spatial analysis. The multinomial approach allows to research, in a single analysis, the association for each class observed. Furthermore, to estimate the spatial risk of a disease in a given region, we also include variables indicating location, which in this study are the geographic coordinates. As an application of this method, this study aims to identify the spatial distribution and risk of overweight, classified as eutrophic+underwieght, overweight and obese, according to World Health Organization criteria.

Methods

A cross-sectional study was conducted with 651 local residents in a municipality in Southeast Brazil. We applied a standardized questionnaire containing socio-economic information, and anthropometric measurements were obtained in the interview. A generalized additive model (GAM) that analyzes the risk related to spatial dependence, considering more than two levels of the response variable was applied (Kelsall and Diggle 1998). Specifically, the cumulative logit model was selected (Ananth et al. 1997).

Results

The prevalence of overweight was 54.2%, with 21.5% of whom obese. Most of the sample were women (55.8%), mean(sd) age was 41.6(12.5) and education measured by years of study was 8.6(4.2).

Figure 1 shows the estimated spatial risks via GAM using cumulative model with three categories. We compared the risk for obesity in relation to eutrophic and overweight (Figure 1-a), and the risk for obesity and overweight comparing to eutrophic (Figure 1-b). Taking into account areas where there is a resident population (represented by streets), note that there is an increased risk of obesity in the Northeast and low risk, or protection, to the Western region (Figure 1-a). Furthermore, when we look at risk of overweight, which includes obesity, and make a comparison with normal weight, we identify risk areas to the Northeast and Northwest (Figure 1-b).

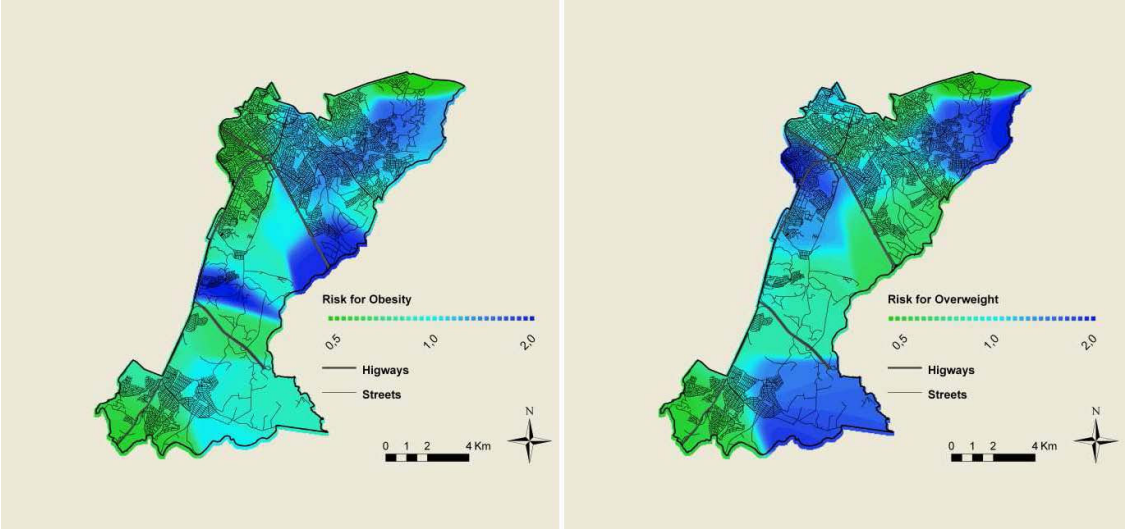


Figure 1: *Spatial Risk for Overweight. Region of a city in Southeast Brazil, 2007.*
a) Risk for Obesity (BMI ≥ 30kg/m^2) b) Risk for Overweight (BMI ≥ 25kg/m^2)

Discussion

The results identify geographical areas of higher and lower risk for overweight and obesity, setting the existence of a spatial component in the obesity epidemic in this region, allowing better clinical and Public Health interventions. New studies on the calculation of significance for the risk estimates are being made. In addition, forms of inclusion of covariates in the spatial analysis that also explain the distribution of the disease are in development.

References

Ananth, C. V. and D. G. Kleinbaum (1997). "Regression models for ordinal responses: a review of methods and applications." Int J Epidemiol **26**(6): 1323-1333.

Drewnowski, A., C. D. Rehm, et al. (2007). "Disparities in obesity rates: analysis by ZIP code area." Soc Sci Med **65**(12): 2458-2463.

Hill, J. O. and J. C. Peters (1998). "Environmental Contributions to the Obesity Epidemic."

Science **280**(5368): 1371-1374.
Kelsall, J. E. and P. J. Diggle (1998). "Spatial variation in risk of disease: a nonparametric binary regression approach." Journal of the Royal Statistical Society (Series C): Applied Statistics **47**(4): 559-573.
Mokdad, A. H., B. A. Bowman, et al. (2001). "The continuing epidemics of obesity and diabetes in the United States." JAMA **286**(10): 1195-1200.
Ogden, C. L., S. Z. Yanovski, et al. (2007). "The epidemiology of obesity." Gastroenterology **132**(6): 2087-2102.
Peeters, A., J. J. Barendregt, et al. (2003). "Obesity in adulthood and its consequences for life expectancy: a life-table analysis." Ann Intern Med **138**(1): 24-32.
Resta, O., M. P. Foschino-Barbaro, et al. (2001). "Sleep-related breathing disorders, loud snoring and excessive daytime sleepiness in obese subjects." Int J Obes Relat Metab Disord **25**(5): 669-675.
World Health Organization (WHO) (1995). Physical Status: the use and interpretation of anthropometry. WHO Technical Report Series. R. o. W. E. Committee.
World Health Organization (WHO) (2000). "Obesity: preventing and managing the global epidemic. Report of a WHO consultation." World Health Organ Tech Rep Ser **894**: i-xii, 1-253.
World Health Organization (WHO). (2010). "Obesity and Overweight." Retrieved 18/02/2010, 2010, from http://www.who.int/dietphysicalactivity/publications/facts/obesity/en/.
Young, L. R. and M. Nestle (2002). "The contribution of expanding portion sizes to the US obesity epidemic." Am J Public Health **92**(2): 246-249.
Zangirolani, L. T. O. (2009). Topologia do excesso de peso no Distrito Sul de Campinas, São Paulo. Faculty of Medical Science. Campinas, State University of Campinas.

RELATIVE RISK ESTIMATION FOR DENGUE DISEASE MAPPING BASED ON DISCRETE TIME-SPACE STOCHASTIC SIR-SI MODELS FOR DISEASE TRANSMISSION WITH TRACT-COUNT DATA

N. A. Samat[1] and D. F. Percy[1]

[1]Salford Business School, University of Salford, Greater Manchester, M5 4WT, United Kingdom, N.A.SamotSamat@pgr.salford.ac.uk

Relative risk estimation is one of the most important issues when studying geographical distributions of disease occurrence for disease mapping. Numerous studies of disease mapping use a regression-type model where observable (fixed effects) and unobservable (random effects) variables are included to give a clean map and hence depict the true excess risk. In spite of this, published studies that use structural disease transmission models for disease mapping are scarce.

Specifically for the case of dengue fever, few researchers use stochastic processes to estimate the relative risk for disease mapping. Rather, most dengue studies are based on exploratory data analysis accompanied by pictorial maps, which includes the study of covariates and their effects on dengue disease distribution; see, for example, Gubler (1998) and Rosa-Freitas *et al*. (2003). Furthermore, some authors use a geographic information system (GIS) to integrate the patient-related information; see, for example, Tran *et al*. (2004).

Consequently, our research introduces an alternative method to estimate the relative risk of dengue disease transmission based initially on discrete-time, discrete-space stochastic SIR-SI models (susceptible-infective-recovered for human populations; susceptible-infective for vector populations). This method is designed to overcome the drawbacks of relative risk estimation in disease mapping using the classical approach based on standardized morbidity ratios. It involves extending the fundamental Poisson-gamma model and developing a Bayesian analytic approach that has previously been applied to the basic SIR model.

In particular, we develop the discrete time-space stochastic SIR-SI models shown below based on a compartmental model that has been used in the study of dengue disease transmission; see, for example, Esteva and Vargas (1998) and Nishiura (2006). This is an extension of the discrete-time stochastic SIR model proposed by Lawson (2006). In general, for $i = 1,2,\ldots,M$ tracts and $j = 1,2,\ldots,T$ time periods, the discrete time-space stochastic SIR-SI models for dengue disease transmission in human populations are as follows:

$$S_{i,j}^{(h)} = \mu^{(h)} N^{(h)} - \mu^{(h)} S_{i,j-1}^{(h)} - \left(1+\mu^{(h)}\right) I_{i,j-1}^{(h)} - \left(1+\mu^{(h)}\right) R_{i,j-1}^{(h)}$$

$$I_{i,j}^{(h)} \left| \lambda_{i,j}^{(h)} \right. \sim \mathbf{Po}\left(\lambda_{i,j}^{(h)}\right)$$

$$\lambda_{i,j}^{(h)} = \exp\left(\beta_0^{(h)} + c_i^{(h)}\right) \frac{\beta^{(h)} b}{\left(1+\mu^{(h)}+\gamma\right)\left(N^{(h)}+m\right)} I_{i,j-1}^{(v)} S_{i,j}^{(h)}$$

$$R_{i,j}^{(h)} = \gamma I_{i,j}^{(h)} - \mu^{(h)} R_{i,j-1}^{(h)}$$

Similarly, the discrete time-space stochastic SIR-SI models for dengue disease transmission in vector (mosquito) populations are as follows:

$$S_{i,j}^{(v)} = A - \mu^{(v)} S_{i,j-1}^{(v)} - \left(1 + \mu^{(v)}\right) I_{i,j-1}^{(v)}$$

$$I_{i,j}^{(v)} \left| \lambda_{i,j}^{(v)} \right. \sim \mathbf{Po}\left(\lambda_{i,j}^{(v)}\right)$$

$$\lambda_{i,j}^{(v)} = \exp\left(\beta_0^{(v)} + c_i^{(v)}\right) \frac{\beta^{(v)} b}{\left(1 + \mu^{(v)}\right)\left(N^{(h)} + m\right)} I_{i,j-1}^{(h)} S_{i,j}^{(v)}$$

The terms that appear in these equations are:

$S^{(h)}$	number of susceptibles in the human population
$I^{(h)}$	number of infectives in the human population
$R^{(h)}$	number of recovered in the human population
$S^{(v)}$	number of susceptibles in the mosquito population
$I^{(v)}$	number of infectives in the mosquito population
$\mu^{(h)}$	per capita mortality rate of humans
$\mu^{(v)}$	per capita mortality rate of mosquitoes
γ	recovery rate of humans
b	biting rate (average number of bites per mosquito per day)
m	number of alternative hosts available as the blood source
A	constant recruitment rate for the vector
$\beta^{(h)}$	probability of transmission from the vector to the human
$\beta^{(v)}$	probability of transmission from the human to the vector
$N^{(h)}$	human population size $S^{(h)} + I^{(h)} + R^{(h)}$
$\beta_0^{(h)}$	constant term for the human population
$\beta_0^{(v)}$	constant term for the vector population
$c^{(h)}$	spatial random effect for the human population
$c^{(v)}$	spatial random effect for the vector population

We perform the calculations for our alternative method of relative risk estimation using WinBUGS software, which is ideal for implementing a Bayesian analysis of structural models such as our discrete time-space stochastic SIR-SI models for dengue disease transmission. In general, for $i = 1,2,\ldots,M$ tracts and $j = 1,2,\ldots,T$ time periods, a sample of observations from the posterior distribution of the mean number of infective humans $\lambda_{ij}^{(h)}$ can be written as $\lambda_{ij1}^{(h)}, \lambda_{ij2}^{(h)}, \ldots, \lambda_{ijn}^{(h)}$. Therefore the posterior mean of the mean number of infective humans can be approximated using the unbiased sample mean

$$\tilde{\lambda}_{ij}^{(h)} = \frac{1}{n} \sum_{k=1}^{n} \lambda_{ijk}^{(h)} .$$

The relative risk parameter $\theta_{ijk}^{(h)}$ can be written as

$$\theta_{ijk}^{(h)} = \frac{\lambda_{ijk}^{(h)}}{e_{ij}^{(h)}},$$

so it follows that the posterior mean of the relative risk may also be approximated using the unbiased sample mean

$$\widetilde{\theta}_{ij}^{(h)} = \frac{1}{n}\sum_{k=1}^{n}\theta_{ijk}^{(h)} = \frac{1}{n}\sum_{k=1}^{n}\frac{\lambda_{ijk}^{(h)}}{e_{ij}^{(h)}} = \frac{\widetilde{\lambda}_{ij}^{(h)}}{e_{ij}^{(h)}}.$$

In other words, the posterior mean for the relative risk is equal to the posterior mean of the mean number of infective humans divided by the expected number of infective cases for human populations. To conclude this presentation, we demonstrate the estimation of relative risk based on these discrete time-space stochastic SIR-SI models for dengue disease transmission using dengue data in the form of counts of cases within the states of Malaysia.

REFERENCES

Esteva, L. and Vargas, C. (1998) Analysis of a dengue disease transmission model. Mathematical Biosciences 150: 131-151.

Gubler, D.J. (1998) Dengue and dengue haemorrhagic fever. Clinical Microbiology Previews 11(3): 480-496.

Lawson, A.B. (2006) Statistical methods in spatial epidemiology. 2nd edition, England: John Wiley & Sons.

Nishiura, H. (2006) Mathematical and statistical analysis of the spread of dengue. Dengue Bulletin 30: 51-67.

Rosa-Freitas, M.G., Tsouris, P., Sibajev, A., Weimann, E.T, Marques, A.U., Ferreire, R.L. and Gards-Moura, F.C.L. (2003) Exploratory temporal and spatial distribution analysis of dengue notifications in Boa Vista, Roraima, Brazilian Amazon, 1999-2001. Dengue Bulletin 27: 63-80.

Tran, A., Deparis, X., Dussart, P., Morran, J., Rabarison, P., Remy, F., Polidori, L. and Gardon, J. (2004) Dengue spatial and temporal patterns, French Guiana, 2001. Emerging Infectious Diseases 10(4): 615-621.

SPATIAL RISK DISTRIBUTION OF WORK RELATED ACCIDENTS IN SOUTHEAST BRAZILIAN CITY UNDER A MULTINOMIAL APPROACH.

C.Stephan, A.C.C.N. Mafra, L.B. Nucci, R. Cordeiro

State University of Campinas, 13083-970, Campinas, São Paulo, Brazil,
celso@stephan.mus.br, +55 19 9675-8904.

Work related accidents are the most important healthy problem to Brazilian workers[1]. Beyond that, underreporting of almost 90% of cases[2] seriously difficult diagnosis and knowledge of its causes. complicating public policies to their prevention. The use of spatial data in epidemiology investigation is not recent, but improvement of computers capacity and development of new statistical methods could open new horizon on epidemiological findings. This spatial case-control study[2], includes data of 2,323 accidents and controls in a 300,000 inhabitants city in Southeast of Brazil. Data were analyzed by degree of severity reported by workers. This were classified at a 1:10 scale and then stratified into 3 levels: light (1-4), moderate (5-7) and serious (8-10) accidents, plus level 4 for the controls. This classification was compared with physician´s one obtaining 0.17 of Pearson correlation coefficient.

Using the definition of spatial risk from Bithell[3], and an extended procedure proposed by Kelsall and Diggle[4] to multinomial analysis, we estimated the work related accident spatial risk by generalized additive model[5] (GAM) for each of those 3 levels compared to controls, whith and without some non-spatial covariates that could be related to the outcome (age; sex; schooling; if the employee has formal contract; if the labor is regular; if is domestic and if is in streets). The multinomial model used was the polytomous logistic model[6]. All programming was made in software R2.7 and the maps were built on ArcMap 9.2.

Figure 1 shows results for the two models studied, one adjusted only with space as predictor and the other adjusted with space plus covariates. Table 1 shows the parametric analysis of the estimated GAM with covariates.

Table 1: *Estimated measures for work related accident. Response: serious, moderate, light and control. N=2323. Piracicaba - SP - Brazil (2007).*

	Severe vs Controls				Moderate vs Controls				Light vs Controls			
	OR	LL	UL	P-value	OR	LL	UL	P-value	OR	LL	UL	P-value
Age (Years)	0,980	0,970	0,990	0,000	0,962	0,946	0,978	0,000	0,977	0,968	0,987	0,000
Schooling (Years)	0,912	0,882	0,942	0,000	0,933	0,889	0,980	0,005	0,957	0,928	0,986	0,004
Sex (Male)	2,137	1,534	2,976	0,000	2,008	1,272	3,169	0,003	2,448	1,809	3,314	0,000
Street Work (Yes)	0,820	0,647	1,039	0,100	0,700	0,498	0,985	0,041	0,750	0,609	0,924	0,007
Regular Work (Yes)	1,712	1,096	2,673	0,018	1,517	0,891	2,582	0,125	1,535	1,082	2,176	0,016
Labor Rights (Yes)	0,899	0,577	1,401	0,638	1,851	1,062	3,226	0,030	1,498	1,047	2,144	0,027
Domestic Work (Yes)	0,678	0,427	1,077	0,100	0,666	0,353	1,255	0,209	0,651	0,423	1,003	0,052

OR: Odds Ratio; LL: OR Lower Limit; UL: OR Upper Limit.

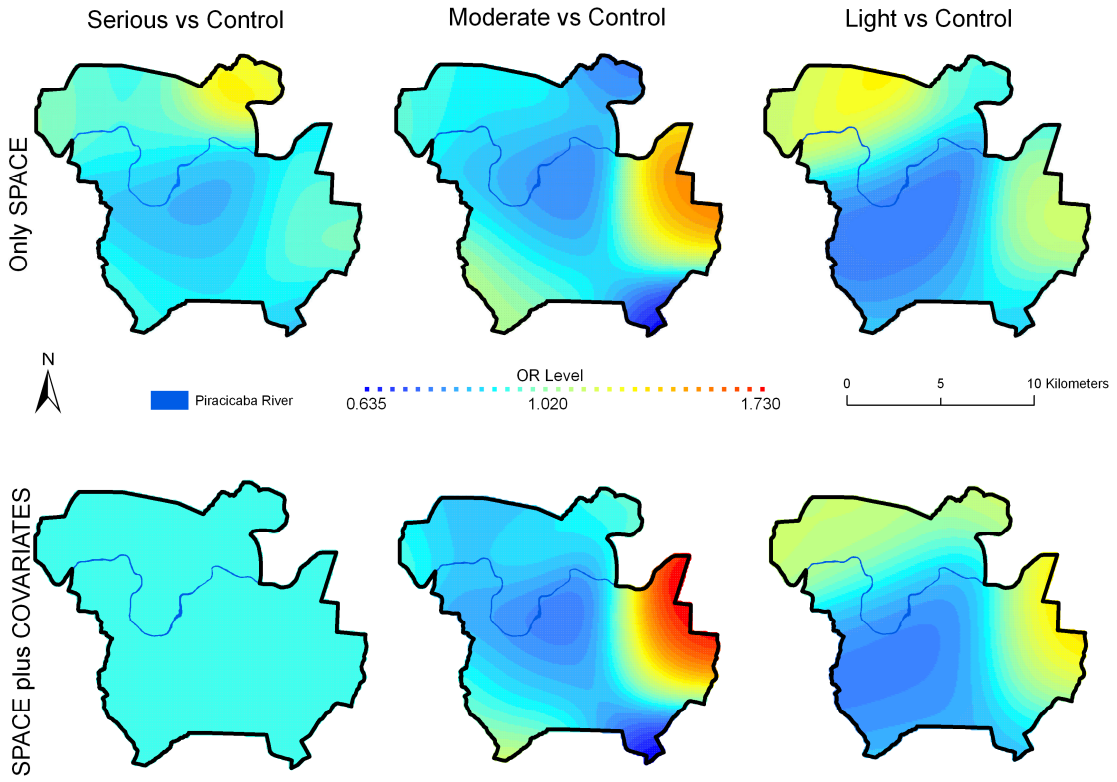


Figure 1: *Spatial Risk of Work Related Accident . Piracicaba - SP - Brazil (2007).*

One can see that different risk levels occur in each area according to the severity of the accident, in this case defined by self report worker. Addition of covariates in the models shows areas where other factors of interest must be investigated, as there is a remained risk in some regions. The method is an important way to detect areas of any disease risk including covariates, under the efficiency of case-control design.

REFERENCES

Hennington EA, Cordeiro R, Moreira Filho DC. Trabalho, violência e morte em Campinas. Cadernos de Saúde Pública. 2004;20:610-7.

Stephan C. Distribuição do risco de acidente do trabalho entre trabalhadores precarizados de Piracicaba. - Campinas, SP : [s.n.], 2008.

Bithell J. An application of density estimation to geographical epidemiology. *Statistics in Medicine* 1990;9:691-701.

Kelsall JE,.Diggle PJ. Spatial variation in risk of disease: a nonparametric binary regression approach. *Applied Statistics* 1998;47:559-73.

Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman and Hall., 1990.

Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology* 1997;26:1323-1333.

GENERALISED LINEAR MODELLING OF THE ASTHMA HOSPITALISATION RISK AND AIR POLLUTANT CONCENTRATION IN PERTH

Y. Yano[1], U. Mueller[1], A. Hinwood[1]

[1] Edith Cowan University, 100 Joondalup Drive, Joondalup, WA,. yyano@ecu.edu.au, umueller@ecu.edu.au, a.hinwood@ecu.edu.au

Asthma is a respiratory disease which carries a risk of hospitalisation in severe cases. Numerous studies have shown that air pollution is one of the factors that exacerbates asthma symptoms (Rumchev, et al, 2004) and while population increases in asthma hospitalisation have been shown for Perth, the nature of exposure and risk across areas within the metropolitan area have not been determined.

The preliminary statistical analysis of the air pollutant concentration data and monthly asthma hospitalisation records showed that the change in the hospitalisation counts was synchronised with the change in NO_2 and CO concentrations. This observation led to an assumption that the asthma hospitalisation risk could be estimated by a generalized linear model with the concentration of air pollutants as the explanatory variable. The measured air pollutant concentration did not provide sufficient data coverage across the Perth metropolitan area as only five monitoring stations measured the concentration of the pollutants of interest. Given the sparseness of hard data, it was not possible to use geostatistical interpolation as illustrated by Malherbe et al (2008). An alternative air pollutant concentration measure was required, which was provided by the combined application of a Gaussian Plume Model (Awasthi, Khare, & Gargava, 2006 and Builtjes, 2003) and Lognormal Kriging (LK) of emission inventory data (NO, CO and PM10) which were available for 2006. The NO emission data were used as the reference variable to NO_2. Higher concentrations of CO and NO are observed in winter, a trend which was also observed in the monitored air pollutant data and past environmental studies (Codde et al., 2003; Hinwood et al., 2006; Rodriguez et al., 2007), while the PM10 concentration becomes higher in summer.

The asthma hospitalisation risk was estimated from the monthly recordings of the asthma admission per postcode via a Bayesian hierarchical model. A variable risk model was adopted for the hospitalization risk. The variable risk modeling was performed in R statistical computing software using the Poisson-gamma exchangeable prior model . The results show that the asthma hospitalisation risk is higher in the outskirts of the central metropolitan area.

The log transformed asthma risk was regressed against the air pollutant concentration, which returned the effect of air pollutant concentration change on asthma hospitalisation risk as a regression coefficient in exponential scale. The results of GLM analysis show that there is a statistically significant relationship between asthma hospitalisation risk and air pollutant concentration. The GLM coefficients suggest that an increase in NO and CO concentration

results in an increase in estimated mean risk, while the PM10 concentration show opposite relationship. Although it is intrinsically unlikely that the increase in particulate concentration decreases the risk of hospitalisation, this outcome is deemed appropriate considering the seasonality of the modeled PM10 concentration and hospitalisation occurrence. The influence of air pollutants on the asthma hospitalisation risk is shown to vary over both time and space. Table 1 gives the estimation equation for the mean hospitalisation risk in 2006 for two postcodes.

Table 1: *Predictive equations for 2006 asthma risk: postcodes-6000 and 6010*

Year - 2006	***Postcode - 6000***	***Postcode - 6010***
CO	$\theta = \exp(-6.2692 + 0.33247 X_{CO})$	$\theta = \exp(-6.9491 + 0.03903 X_{CO})$
NO	$\theta = \exp(-6.038 + 0.2017 X_{NO})$	$\theta = \exp(-8.234 - 0.1833 X_{NO})$
PM10	$\theta = \exp(-5.977 - 0.1393 X_{PM10})$	$\theta = \exp(-6.449 - 0.1541 X_{PM10})$

where θ X_{CO}, X_{NO} and X_{PM10} denote the estimated mean risk and the concentration of CO (g/month), NO (g/month) and PM10 (μg/month) respectively.

A comparison of the yearly GLM output suggests that the asthma hospitalisation risk has decreased over the study period, although few acute rises in the risk were observed in some years. Overall a decrease in the risk over the period of 2000 to 2007 can be seen from the GLM outputs, the NO model output follows the similar distribution and temporal variation to the CO model output, while PM10 was deemed as having no effect on the asthma hospitalization risk.

References

Albert, J. (Ed.). (2007). *Bayesian Computation with R*: Springer.

Awasthi, S., Khare, M., & Gargava, P. (2006). General plume dispersion model (GPDM) for point source emission. *Environmental Modeling and Assessment, 11*(267-276).

Builtjes, P. (Ed.). (2003). *AIR QUALITY AND MODELLING - Theories, Methodologies, Computational Techiniques and Available Databases and Software. Vol 1 - Fundamentals*: The EnviroComp Institute and the Air & Waste Management Association.

Codde, J., De Klerk, N., Hinwood, A., Landau, L., Lyons, T., Murray, F., et al. (2003). *Research on Health and Air Pollution in Perth, Morbidity and Mortality: A Case-Crossover Analysis, 1992-1997*. Perth: Department of Environment.

Hinwood, A., De Klerk, N., Rodriguez, C., Jacoby, P., Runnion, T., Rye, P., et al. (2006). The relationship between changes in daily air pollution and hospitalizations in Perth, Australia 1992-1998: A case-crossover study. *International Journal of Environmental Health Research, 16*(1), 27-46.

Malherbe, L., Cardenas, G., Colin, P., & Yahyaoui, A. (2008). Using different spatial scale measurements in a geostatistically based approach for mapping atmospheric nitrogen dioxide concentrations. Application to the French Centre Region. *Environmetrics, 19*, 751-764.

Rodriguez, C., Tonkin, R., Heyworth, J., Kusel, M., De Klerk, N., Sly, P. D., et al. (2007). The relationship between outdoor air quality and respiratory symptoms in young children. *International Journal of Environmental Health Research, 17*(5), 351-360.

Rumchev, K., Spickett, J., Bulsara, M., Phillips, M., & Stick, S. (2004). Association of domestic exposure to volatile organic compounds with asthma in young children. Thorax, 59, 746-751.

MONDAY 13 SEPTEMBER 2010

POSTER PRESENTATIONS

8th International Conference on Geostatistics for Environmental Applications

POSTER PRESENTATIONS

Monday 13 September 2010

Chairs*: *Ph. Renard & J. McKinley

D. Ortega-Lenis, D. Arango-Londoño, M. Andrade-Bejarano	Analysis of homicides occurred at Cali, Colombia, during the period 2004-2009, through geostatistics methods and multivariate data analysis
S. Bek, J. Jezek	Optimization of interpolation parameters when deriving DEM from contour lines
F.Z. Benmostafa Largueche	Modelling soil contamination using geostatistical techniques, watford waste site a case study
P. Bossew	Is the radon: soil – indoor transfer factor a reasonable concept?
M. García Howlett and J.M. Cámara Vicario	Geostatistics and spatial analysis applied in monitoring and predicting urban pests in Madrid city
A. Gheris	Development of a combined scheme: mixed hybrid finite element-finite volume (MHFE-FV) for modeling of contamined transport in unsaturated porous media
A. Govaerts, A. Vervoort, P. Darius	Interactive modules for the visualization and teaching of geostatistical concepts
F. Kardel, K. Wuyts, M. Babanezhad, U.W.A. Vitharana, A.R. Khavaninzadeh, T. Wuytack, R. Samson	Spatial distribution of plant anatomical and morphological characteristics for biomonitoring of urban habitat quality
R. Kasparinski, O. Nikodemus	Influence of environmental factors on forest soil spatial distribution and diversity in Latvia
A.C.C.N. Mafra, R. Cordeiro, L.B. Nucci, C. Stephan	Spatial risk of disease in case-control studies with multinomial responses: an occupational accident example.
E. Meerschman, M. Van Meirvenne	Regional characterization of soil heavy metals in a former World War I battle area
J.M. Mirás-Avalos, P. Sande-Fouz, R.M. Mestas-Valero, A. Paz-González	Mapping monthly mean temperature in Galicia (NW Spain) using geostatistical procedures

L.A. Morales, E.V. Vázquez and J. Paz-Ferreiro	Spatial variability of pH, redox potential and extractable Fe, Mn and Zn on paddy soil as a function of crop stage and lime amendment
F.Pla, H. Herrero	A first approximation to mantle convection: a Rayleigh-Bénard problem with variable viscosity and non uniform heating
V. Sidorova, P. Krasilnikov	Improving the map of soil horizons by indicator kriging
K. Szopka, A. Karczewska, C. Kabała, P. Jezierski, A. Bogacz	Spatial differentiation of zinc and copper concentrations in the surface layers of forest soils in the area of Karkonosze national park, Poland
G.H. Zahedi Amiri and S. Ahmadi	Spatial dependence between soil carbon and nitrogen in pure beech forest stand in north of Iran

ANALYSIS OF HOMICIDES OCCURRED AT CALI, COLOMBIA, DURING THE PERIOD 2004-2008, THROUGH GEOSTATISTICS METHODS AND MULTIVARIATE DATA ANALYSIS

D. Ortega-Lenis[1], D. Arango-Lodoño[1], M. Andrade-Bejarano[1]

Universidad del Valle, School of Industrial Engineering and Statistics, Section of Statistics, Cali, Colombia, South America, E-mail: mandrade@pino.univalle.edu.co

The study of the homicides in a city is very important, because with the knowledge of the phenomenon and its behaviour, is possible to identify violence focus and to implement security plans and vigilance, with the objective of the habitants have better life conditions and coexistence.

The high rates of homicides registered daily at Cali, Colombia are a cause of concern for authorities and institutes in charge of the citizen security and social coexistence (CISALVA, 2005, 2006, 2007). Authors as Kaplan *et al*, 1996 have identified the relation among poverty, social inequality and violent crime, when they compared the homicides rates among countries or among geographical areas of the same country. They suggested to analyse the spatial variability of the homicides rates in a territory to explain its possible variations.

Data of this study belong to the period January 2004 to December 2008 and have been collected by a Vigilance Committee in charge to analyse the deaths for external causes in Cali, Colombia. Database contain information about homicide in 292 neighborhoods, and 49 variables belonging to 1) Personal information of the victim 2) Place of the homicide 3) Information of the attacker 4) Time variables 4) Context of the homicide and type of weapon.

Rates were fitted by Bayesian method (Marshall 1991). This method does a weighted having into account the number of habitants in a neighborhood, giving higher weight to the average rate calculated for the city in the case when a neighborhood has small number of habitants.

Two methodologies are used to analyse homicide data: 1) Geostatistics (Matheron, 1962), with the objective to analyse the possible spatial correlation of the homicides through the time, the possible direction of the phenomenon and the clustering of the homicides in the city through risk maps, 2) Multivariate Data Analysis with the Principal Components Analysis (PCA) (Hotelling, 1933) and Cluster Analysis , to characterise the homicides by towns, by using variables as type of weapon, Place of the homicide, Context of the homicide, etcetera.

The spatial statistics is a set of techniques suitable for the data analysis of random variables, measured in several sites (points of the space or spatial aggregations) of a region. In a formal way, spatial statistics works with realisations of the random process (random field) , where S, is the location in the Euclian space d – dimensional and is a random variable in the location S (Chilès and Delfiner, 1999).

PCA as a reduction technique and data representation is applied to data tables where the columns are quantitative variables, and the rows are the subjects. This is a mathematical methodology where is no necessary to assume a probabilistic distribution. PCA transforms the original set of variables in a smaller set of variables, which are linear combinations of the original variables and they contain the highest variability.

Figure 1(a) shows the empirical variogram of homicides rates for the year 2004. This variogram shows a good fit to the exponential model (blue line). Figure 1(b) shows an anisotropic behaviour in the direction 90° (green line). This direction belongs to the neighborhoods from the east of Cali. These neighborhoods belong "Distrito de Aguablanca" are identified as vulnerable zones, with high levels of poverty.

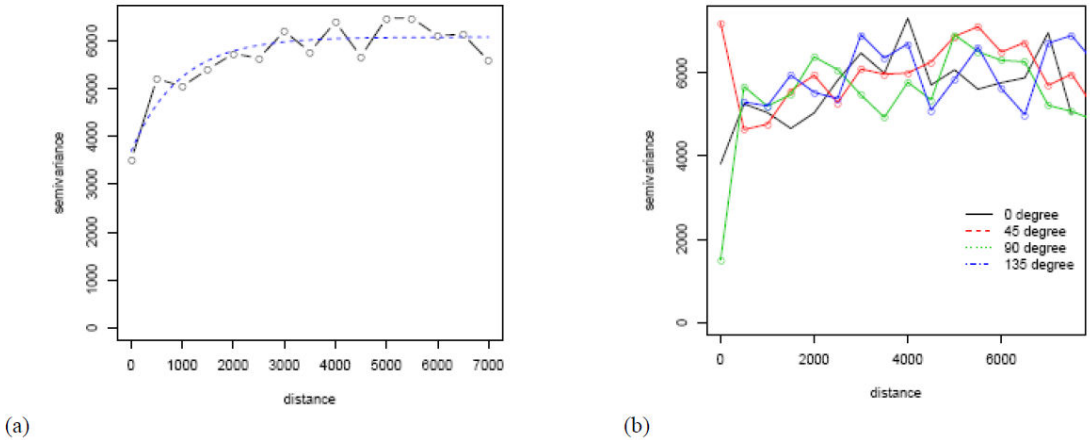
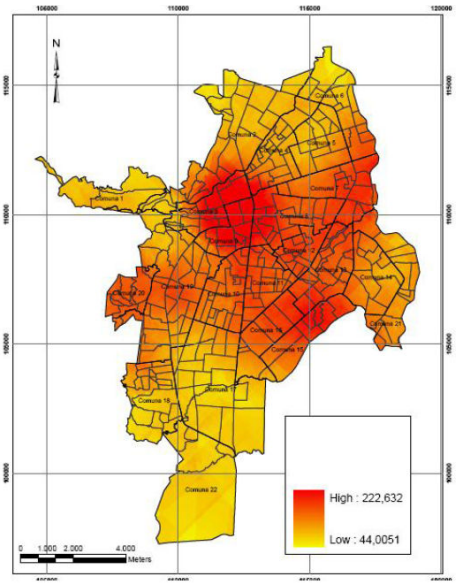


Figure 1: (*a) Empirical and Theoretical Variograms (b) Directional Variograms*

Maps were obtained using by the kriging results of the exponential model fitted. Figure 2 shows that the neighborhoods with the highest risk of homicides belong to central and eastern part of the Cali city. This behaviour is seen in 2004 and in the rest of years analysed

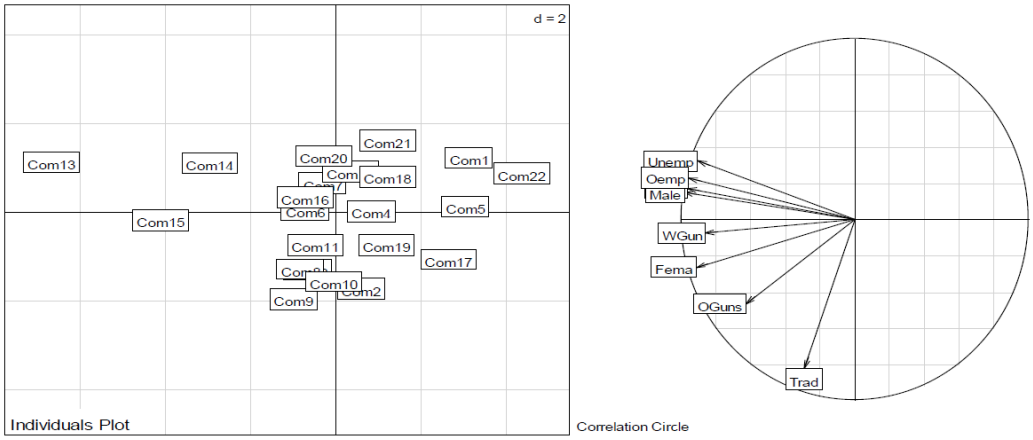


Data of Santiago de Cali

Latitude (North)	3°27'26"
Longitude (West)	76°31'42"
Altitude (m)	1,070 m
Total Area	560,3 km²
Territory Division	292 Neighborhoods
Population (2006)	2'075,380

Figure 2: *Risk Map of Homicides, Estimated by Kriging, for year 2004*

Figure 3(b) shows that the variables are located at left size of the first factorial axis; the white weapon and the masculine gender have the highest contribution in the conformation of the first factorial axis. In addition, homicides of the masculine gender are associated with the use of the white weapon and the unemployment. Figure 3(a) indicates that the towns 13, 14 and 15 are away of the first factorial axis, in the sense that the number of homicides increase. Furthermore, these towns seem clustered as of high risk of homicide; On the contrary, towns 1, 5 and 22 seem clustered as of low risk of homicide.



(a) (b)

Figure 3: *(a) Subject graphic (b) Variable graphic*

BIBLIOGRAPHY

CISALVA (2005,2006 y 2007). Boletín anual de muertes por causa externa.

Chilès, J. and Delfiner, P. (1999), Goestatistics. Modelling Spatital Uncertainty, John Wiley and Sons, Inc. United States.

Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psychology. 24. 417-441.

Kaplan, G., Pamuk, E., Lynch, J., Cohen, R., & Balfour, J. (1996). Inequality in income and mortality in the United States: analysis of mortality and potential pathways. BMJ.

Matheron, G. (1962). Traite de Geostatistique Apliquee, Tome I. Emoires bureau de Recherches Geologiques et Minieres, N 24. Editions Bureau de Recherche et Minieres, Paris.

Marshall, R. J. (1991). Mapping Disease and Mortality Rates using empirical Bayes Estimator. Applied Statistics 40.

OPTIMIZATION OF INTERPOLATION PARAMETERS WHEN DERIVING DEM FROM CONTOUR LINES

S. Bek[1], J. Ježek[1]

[1] Charles University in Prague, Faculty of Science, Institute of Applied Mathematics and Information Technologies, Albertov 6, 128 43 Praha 2, Czech Republic, e-mail: bek@natur.cuni.cz, 00420221951584

In digital soil mapping and erosion numerical modeling, the primary elevation data have to be processed and topographic analysis carried out before we can use it for the particular application. We are interested in optimizing an interpolation method used for contour lines data which models a terrain.

The regularized spline with tension (RST), constructed by Mitáš & Mitášová (1993), proved to be one of the most appropriate for digital modeling and particularly for sediment-flow applications (e.g. Hofierka et al. 2007 or Cebecauer et al. 2002). The RST function S(x) is based on the condition of minimizing the deviation from the measured points and its smoothness seminorm. S(x) is computed as: $S(x) = a + \sum_{j=1}^{N} \lambda_j R(x, x^{[j]})$, where $R(x, x^{[j]})$ is radial basis function with an explicit form depending on *interpolation parameters* of the method. The coefficients a and λ_j are obtained by solving the system of linear equations corresponding to the input points. The interpolation parameters are: (i) *tension* which tunes the character of the resulting surface from thin plate to membrane, (ii) smoothing parameter, *smooth,* which controls the deviation between the given points and the resulting surface. *Smooth* plays an important role when the data contains measurement error. Otherwise, a small value is recommended to avoid overshoots. (iii) *sampling step:* Since the method requires points as its input, the contours have to be sampled first.

In our study area, which is a steep slope with an evidence of intensive soil erosion with area of about 25 hectars in Czechia, we use the above described RST method to create a reference DEM based on scatter data points measured by laser theodolite. Due to a high accuracy of the measurements, we have chosen a very low smoothing parameter (0.001), just to avoid overshoots. A wide range of tension parameters have been tested, while the one that generated minimal standard error of the cross-validation have been chosen (30). Terrain characteristics, important for erosion modeling (such as slope, aspect and curvatures) are computed simultaneously with the interpolation. Hence, it is not necessary to compute them numerically from the DEM during which the spatial resolution plays the important role. We consider the resolution of the grid of 1 meter as a good compromise that assures the necessary precision by comparing the surface models and the acceptable computational cost.

The contour lines were provided by the national Geodetic and Cadastral Office. It concerns the vectorized contours from basic topographic maps at a scale of 1: 10 000 with the contour interval of 5 m, ranging from 375 to 465 m. Kučera (1961) empirically derived the standard

error of the contour altitude information: $m_v = \sqrt{0{,}88^2 + 4{,}2tg\varepsilon}$, where ε denotes the slope steepness.

The deviations of the reference DEM and contours were biased. Therefore, we had to increment the reference DEM by 2,76 m to obtain the calibrated reference DEM (REF). In order to optimize the interpolation parameters for contours, we look for the three interpolation parameters (sampling step, tension and smooth) which generate a surface (M1) whose standard deviation of the differences from REF ($s_{M1\text{-}REF}$) is the lowest. For a dense net of the three parameters, we have repeatedly created DEMs and computed the $s_{M1\text{-}REF}$. We have found the minimum $s_{M1\text{-}REF}$ = 1,18 m for sampling step = 20, tension = 18, smooth = 0,4 whereas it was possible to optimize the interpolation parameters of tension and smooth for large interval of the sampling steps (in our study it was approximately 10 – 60 meters). The value of $s_{M1\text{-}REF}$ is comparable to the standard deviation of the vertical difference between contour points and the reference model $s_{CON\text{-}REF}$ = 1,3 m. We have found that the minimizing function was only very slightly sensitive to the sampling step which could allow us to use a higher sampling step (e.g. 50 m) to spend the computational cost. The resulting surface M1 seems to be plausible to REF also regarding the topographic potential for erosion and deposition which is a derived characteristic of DEM described by Mitáš & Mitášová (1998).

In common applications the reference DEM is not at our disposal and the cross-validation is used to solve the problem. Nevertheless, the use of the cross-validation to find the optimal triplet of the interpolation parameters (sampling step, tension and smooth) brings some problems. For short sampling steps the cross-validation ceases to indicate a goodness of the interpolation. The standard cross-validation deviation is very low also for unrealistic results. It is due to the fact that the distances between points on the same contour are much smaller than the distance to the neighboring contours. Regarding the experience with the optimization using reference DEM, we could use reasonably large sampling step of 20 m. We obtained the optimal parameters (tension = 21, smooth = 0.01). Nevertheless, we observed unrealistic step-like features at the bottom of valley. We suppose that the problem is caused by the fact that the cross-validation procedure does not account for vertical error of input data. We obtained much better results after we enlarged the sampling step or increased the smoothing parameter which we demonstrate on the topographic potential for erosion and deposition, see Figure 1.

We came to the conclusion that the sampling step does not play very important role when using a reference DEM. It was possible to obtain very similar "optimal" surfaces with $s_{M1\text{-}REF}$ close to minimum for sampling steps ranging approx. from 10 to 60 meters. The choice of the sampling step becomes important, if a reference model is not at our disposal. This is the case of common applications when often the cross-validation is performed to obtain the optimal interpolation. For a good cross-validation performance, we have to use larger sampling steps (over 20 m). The use of cross-validation is suitable for data without measurement error, however for our contour data we had to use larger sampling steps or smoothing parameter, than proposed by CV, to obtain more accurate results. Further, we are studying the

relationships between smooth and sampling step parameters in order to adjust their combination properly to the known measurement error of input points.

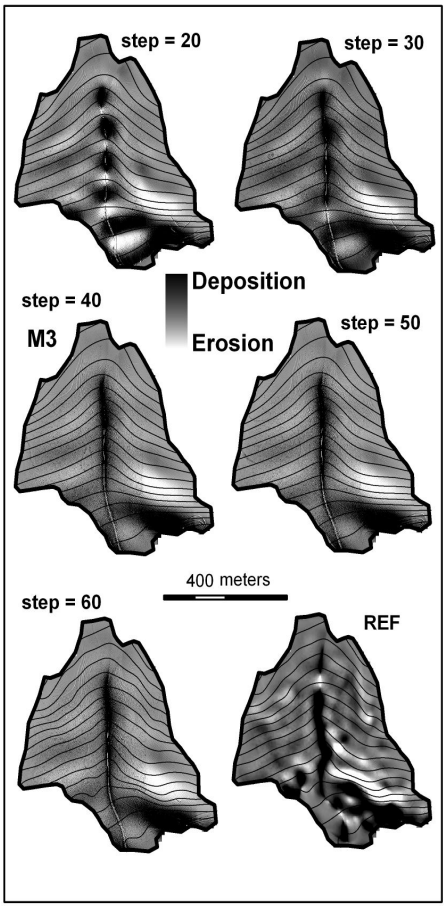


Figure 1: *Topographic potential for erosion and deposition on the optimized surfaces for the different sampling steps.*

References

Cebecauer, T, Hofierka, J, Šúri, M (2002) Processing digital terrain models by regularized spline with tension: tuning interpolation parameters for different input datasets. Proceedings of the *Open source GIS - GRASS users conference 2002* - Trento, Italy.

Hofierka, J, Cebecauer, T, Šúri, M (2007) Optimisation of Interpolation Parameters
Using a Cross-validation. In: Peckham, R J, Jordan, G (Eds.): Digital Terrain Modelling - Development and Applications in a Policy Support Environment. Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, pp. 67-82.

Kučera, K (1961) Kritéria přesnosti topografického mapování v měřítku 1: 5000 a 1: 10 000. In: Sborník výzkumných prací IV. Praha, VÚGTK. [in czech]

Mitáš, L, Mitášová, H (1998) Distributed erosion modeling for effective erosion prevention Water Resources Research Vol. 34, No. 3, pp. 505-516.

Mitášová, H, Mitáš, L (1993) Interpolation by regularized spline with tension : I. Theory and implementation. Mathematical Geology 25: 641-655.

MODELLING SOIL CONTAMINATION USING GEOSTATISTICAL TECHNIQUES, WATFORD WASTE SITE A CASE STUDY

F.Z. Benmostefa Largueche

benmostefafatima@yahoo.fr
University of Badji-Mokhtar; Annaba; Algreria

Abstract:
The main objective of this study was to investigate whether Kriging is a useful tool to estimate the spatial distribution of ground pollutants in contaminated land. The second objective of this work was a more practical one. It consists on the identification of areas that should be subjected to remedial actions and also on deciding which contaminant needs to be considered when remediation processes are taken. To achieve the described objectives, a contaminated site has been studied and the following steps have been followed: The contamination concentration limits beyond which action needs to be taken to remediate the ground contamination, in which case it is important to determine the areas that should be subjected to the appropriate remediation measures. A presentation of a case study will follow. A brief site description is given. Next, a spatial analysis of the site has been carried out. It consists essentially of: Firstly a primary process of the data which means that histograms and an unprocessed representation of the pollutant's distribution has been plotted for each contaminant. Secondly a graphic presentation of the pollution using Kriging interpolation technique is shown. Finally co clusions concerning Kriging applications are given. An assessment concerning Kriging is presented and a balance between the advantages and disadvantages of its use is discussed.

Key words:
Ground contaminants, remediation, assessment, regionalised variables, semivariogramms, geostatistics interpolations techniques.

References

1. Aboufirassi, M and Marino, M. A. 1983. "Kriging of Water Levels in the Souss Aquifer, Morocco". Mathematical Geology. Vol. 15, No 4, pp 537-551.
2. Aboufirassi, M and Marino, M. A. 1984. "Co-Kriging of Aquifer Transmissivities from Measurements of Transitivities and Specific Capacity". Mathematical Geology. Vol. 16, No 1, pp 19-35.
3. Abtew, W; Obbeysekera, J and Shih, G. 1993. "Spatial Analysis for Monthly Rainfall in South Florida". Journal of Water Resources. Vol. 29, No 2, pp 179-188.
4. Agterberg, F. P. 1970. "Autocorrelation Functions in Geology". In Geostastics, Edit. D. F. Merriam, pp 113-141. Plenium, New York.
5. Agterberg, F. P. 1974. "Geomathematics: Mathematical Background and Geo-Science Applications". Elsevier Scientific Publishing Company. pp596.

6. Armstrong, M. 1984 a. "Common Problems seen in Variograms". Mathematical Geology. Vol. 16, No 3, pp 305-313.
7. Armstrong, M. 1984 b. "Problems with Universal Kriging". Mathematical Geology. Vol. 16, No 1, pp 101-108.
8. Armstrong, M and Diamond, P. 1984. "Testing Semi-Variogram for Positive Definiteness". Mathematical Geology. Vol. 16, No 4, pp 407-421.
9. Armstrong, M and Jabin, R. 1981. "Variogram Models must be Positive-Definite". Mathematical Geology. Vol. 13, No 5, pp 455-459.
10. Atkinson, P. M and Curran, P. J. 1995. "Defining an Optimal Size of Support for Remote Sensing Investigations". Vol. 33, No 3, pp 768-776.
11. Atteia, O; Dubois, J. P and Webster, R. 1994. "Geostatistical Analysis of Soil Contamination in the Swiss Jura". Journal of Environmental Pollution. Vol. 86, No 3, pp 315-327.
12. Bardossy, A and Bardossy, G. 1984. "Comparison of Geostatistical Calculations with the Results of Open Pit Mining at the Iharkut Bauxite District, Hungary: A Case Study". Mathematical Geology. Vol. 16, pp 173-192.
13. Barnes, M.G; Giacomini, J. J; Reiman, R. T and Elliot, B. 1980. NTS Radiological Assessment Project Results for Frenchman Lake Region of Area 5. DOE/ DP/ 01253-17. Desert Research Institute, University of Nevada, Reno, Nev. pp 108.
14. Bourgault, G. 1994. "Robutness of Noise Filtering by Kriging Analysis". Mathematical Geology. Vol. 26, No 6, pp 733.
15. Burgess, T. M., and Webster, R. 1980 b. "Optimal Interpolation and Isorithmic Mapping of Soil Properties 2. Block Kriging". Journal of Soil Science. Vol. 32, pp 643-659.
16. Buxton, B. E. 1982. "Coal Reserve Assessment: A Geostatistical Case Study". M.Sc Thesis presented to Stanford, C A.
17. Byers, W; Meyers, M. B and Mooney, D.E. 1994. "Analysis of Soil from a Disused Gasworks". Journal of Water, Air Soil Pollution. Vol. 73, No 1-4, pp 1-9.
18. Carr, J. R. Roberts and Kevin P. "Application of Universal Kriging for Estimation of Earth".
19. Christakos, G and Thesing, G. A. 1993. "Intrinsic Random Field Model in the Study of Sulfate Deposition Process". Atmospheric Environment, Part A: General Topics. Vol. 27 A, No 10, pp 1521-1540.
20. Clark, I. 1979. "Practical Geostatistics". Applied Science Publishers, Lodon. pp 129.
21. Clark, I; Guarascio, M; David, M and Huijberghts, C. 1976. "Some Practical Computational Aspects of Mine Planning". Advanced Geostatistics in the Mining Industry, Publishing, Dordrecht, Holland, pp 391-399.
22. Collins, M. 1996. Personal Communication.
23. Cressie, N and Hawkins, D. M. 1980. "Robust Estimation of the Variogram: 1". Mathematical Geology. Vol. 12, No 2, pp 115-125.
24. Cressie, Noel. 1990. "The Origins of Kriging". Mathematical Geology. Vol. 22, No 3, pp 239.

25. Davis, M. W. D., and David, M. "Automatic Kriging and Contouring in the Presence of Trends (Universal Kriging made Simple)". Journal of Canadian Petroleum Technology. Vol. 17, No 1, pp 1-10.
26. Davis, M. W. D; David, M., and Belisle, J. M. 1978. "A Fast Method for the Solution of a System of Simultaneous Linear Equations- A Method Adapted to a Particular Problem". Mathematical Geology. Vol. 10, No 4, pp 369-374.
27. De Kwaadsteniet, J. M. 1986. "Is the Kriging Technique a useful Tool for the Optimization of a Monitoring Network?". R I V M. Vol. 19, No 2, pp 22-23.
28. Delfiner, P. 1976. "Linear Estimation of Non-Stationary Spatial Phenomena". Advanced Geostatistics in the Mining Industry, M. Guarascio, M. David, and C. Huijbreghts, Eds, D. Reidel, Publishing, Dordrecht, Holland. pp 49-68.
29. Ferguson, C. C. 1992. "The Statistical Basis For Spatial Sampling of Contaminated Land". Ground Engineering. pp 34-38.
30. Finke, P. A; Bourna, J and Stein, A. 1992. "Measuring Field Variability of Disturbed Soils for Stimulation Purposes". Soil Sciences Society of America. Vol 56, No 1, pp 187-192.
31. Fleming, G. 1991. "Recycling Derelict Land". Institution of Civil Engineering, Thomas Telford Publishing Company.
32. Fleming, P. R. 1994. "Geostatistics for Contaminated Land". pp 41-47.
33. Gallichand, J; Marcotte, D and Prasher, S. O. 1992. "Including Uncertainty of Hydraulic Conductivity into Drainage Design". Journal of Irrigation and Drainage Engineering. Vol. 118, No 5, pp 744-756.
34. Goovaerts, P; Sonnet, P and Navarre, A. 1993. "Factoriel Kriging Analysis of Sringwater Contents in the Dyle River Basin, Belgium". Water Resource Research. Vol. 29, No 7, pp 2115-2125.
35. Hawkins, D. M. and Cressie, N. 1984. "Robust Kriging- a Proposal". Mathematical Geology. Vol. 16, No 1, pp 3-18.
36. Henley, S. 1981. "Non-Parametric Geostatistics". Applied Science Publishers. pp 145
37. Hosseini, E; Gallichand, J and Marcotte, D. 1994. "Theoritical and Experimental Performance of Spatial Interpolation Methods for Soil Salinity Analysis". Vol.37, No 6, pp 1799-1802.
38. Houlding, S. W. 1994. "3D Geoscience Modelling. Computer Techniques for Geological Characterisation". Springer Publishers.
39. ICRCL (1987). Guidance on the assessment and redevelopment of contaminated land. ICRCL 59/83 2nd edition. Department of Environment, London.
40. Interdepartmental, Committee of the Redevelopment of Contaminated Land. "Notes on the Redevelopment of Gasworks Sites".
41. Journel, A. G, and Huijbreghts, Ch. J. 1978. "Mining Geostatistics". Academic Press. pp 600.

IS THE RADON: SOIL – INDOOR TRANSFER FACTOR A REASONABLE CONCEPT?

P. Bossew

Bundesamt für Strahlenschutz / German Federal Radioprotection Authority; Köpenicker Allee 120-130, 10318 Berlin. e-mail: pbossew@bfs.de , ph. +49-30-18333-4231.

Abstract

It has been suggested to use a "transfer factor" (TF) (more correctly, but less commonly called concentration ratio) for estimating indoor radon (Rn) concentrations from Rn concentrations in soil air. We investigate the spatial distribution of estimated transfer factors, for a particular region of Germany known for high Rn potential in some areas. The results are, that (1) there is a significant spatial pattern of the TF, with the TP negatively correlated to soil air concentration; and (2) that the local distributions of the TP are quite dispersed, with geometric standard deviations (GSD) up to 2, and hence problematic to predict locally.

Data and estimation

We use soil and indoor Rn concentrations from the N part of the Oberpfalz region of Germany, as described in Bossew (2010; this conference). Also the estimation procedure by OK on nscore transformed data is the same. Back-transform for E-type estimates and estimation of local distributions is done by simulation. For each grid node 2000 samples are taken from the local distributions in nscore- (labelled Y) space, known from theory to be (asymptotically) normally distributed, $y \sim F_Y = N(y^*(x), s_Y(x))$, with $y^*(x)$ the OK estimate and $s_Y(x)$ the kriging SD at location x, in Y-space. Each sample is back-transformed into the original (labelled Z) space by simple linear interpolation in the $Z \leftrightarrow Y$ transform table, and statistics are built. In Z-space, the ratio indoor / soil is calculated and its statistics computed. At this stage Y(soil) and Y(indoor) were estimated independently, since so far we do not have the technical means for co-estimation; for the E[Y] and the TF, however, as a compromise, the Y(soil, indoor) were random-sampled from bivariate normal with correlation $r^2=0.638$ as found in Bossew (ibd., fig. 1).

Results

The E-type estimates of Rn concentrations in soil and indoor are mapped in Figure 1. The spatial pattern known from the measurements is reproduced well, although extremes are underestimated. For examples, the highest measured concentration in soil air is above 1000 kBq/m³, whereas the highest estimated grid node equals somewhat less than 400. This is due to the relatively high nugget of the variogram, which could however not be set lower, in view of the current data situation. Global distributions of the measurements and the estimates of soil Rn are shown in

Figure ***2***, as well as local distributions of the estimates at two locations: One low- and one high-Rn point was chosen, (214000,192000) and (204000,232000), respectively. The spatial distributions of the means (E), medians (M) and GSDs of the transfer factors are shown in

Figure 3. Their global, and two local distributions (same points as for the concentrations) are given in Figure 4.

Conclusions

We observe the following:

(1) The TF are spatially distinctly structured, and negatively correlated to the concentration in soil air. Reasons may be found in non-linear transfer from soil into indoor air, regionally different transport properties of subsoil, or regionally different building styles and living habits, which all influence indoor Rn, apart from geogenic Rn occurrence, as quantified by the concentration in soil air.

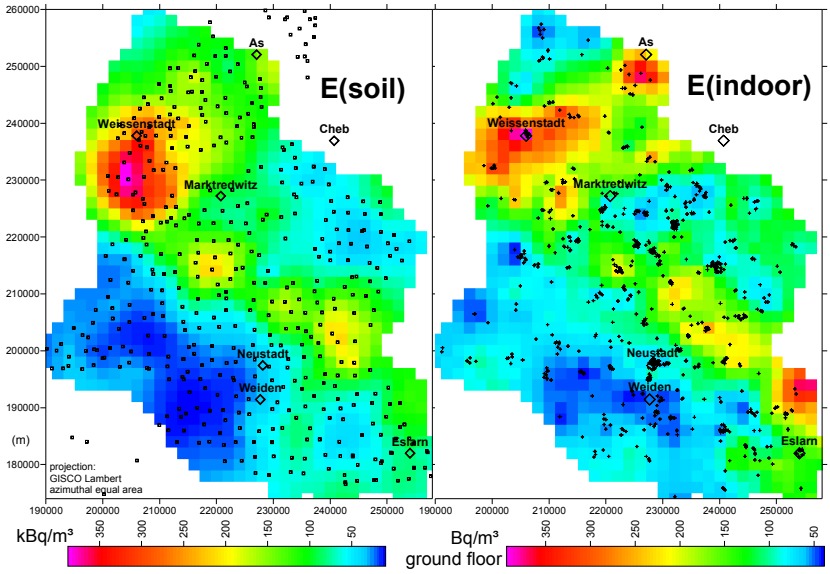


Figure 1: *estimated soil and indoor Rn concentrations. Symbols: sampling locations.*

(2) The estimated local distributions of the TF are quite dispersed, with local GSDs up to 2 (Median over all nodes 1.79; median of the coefficient of variation (CV): 66%). In a future, improved attempt, co-estimation of soil- and indoor concentrations in Y-space may reduce the variability.

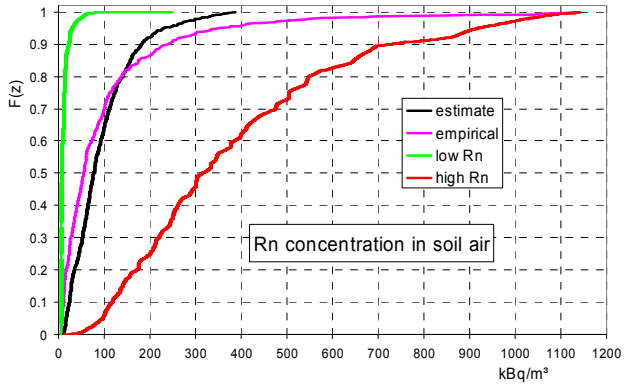


Figure 2: *cumulative frequency distributions F of concentrations in soil air; global: estimates E(Z) over all grid nodes, empirical data (measurements); and local distributions at one point with low, and one point with high mean concentration.*

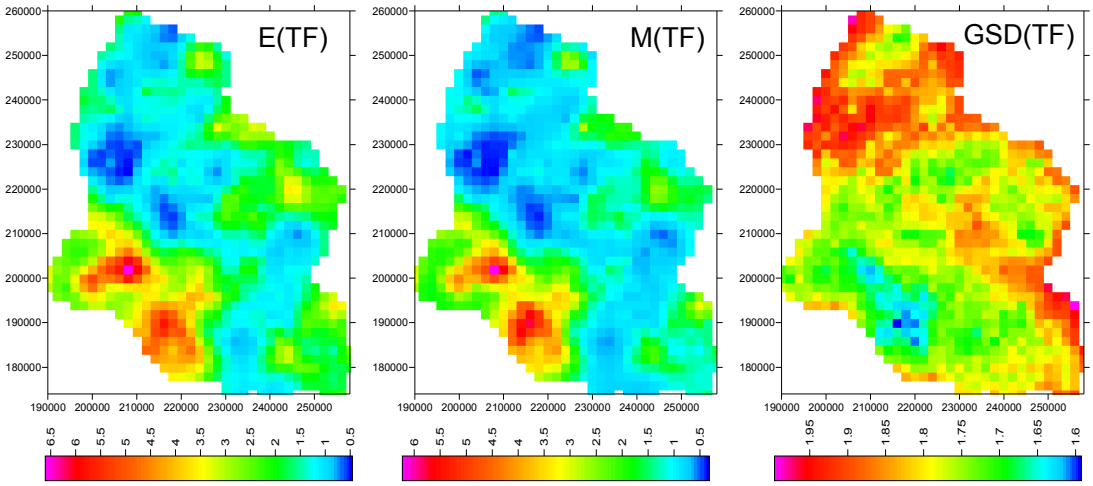


Figure 3*: geographical distribution of estimated mean (E), median (M), and geometric standard deviation (GSD) of the soil – indoor transfer factor. Units of E and M: (Bq/m³)/(kBq/m³).*

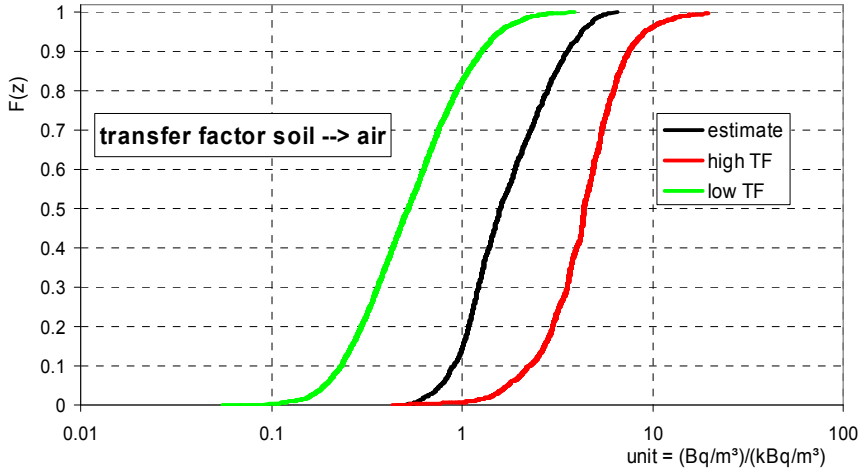


Figure 4: *cumulative frequency distributions F of soil – indoor transfer factors TF; global: estimates E(TF) over all grid nodes; and local distributions at one point with low, and one point with high mean TF.*

The higher GSDs at the edge of the investigated region are due to the higher estimation uncertainty of soil and indoor Rn, as kriging SD in Y-space, which basically reflects the geometry of sampling design (sparser towards the edges).

In particular factor (1) renders the usefulness of the concept of a TF somewhat questionable. A way out – not trivial, however – may be "factorizing out" influences on Rn transport other than the geogenic potential.

GEOSTATISTICS AND SPATIAL ANALYSIS APPLIED IN MONITORING AND PREDICTING URBAN PESTS IN MADRID CITY

M. García Howlett[1] **and J. M. Cámara Vicario**[1]

[1] Vector Control Unit, Environmental Health Department, Public Health Institute-Health Madrid (Madrid City Council), Madrid, Spain, mspvectores@munimadrid.es, 0034 915 113 089 (telephone), 0034 915 113 085 (fax).

Health Madrid is the municipal health administration of the City Council of Madrid and environmental health is a very important subject in the Health Madrid organization. Nowadays, for Health Madrid there are many reasons to consider vector-pest risk management as an essential item on public health, some examples are:

- The emergence and reemergence of vector-borne parasitic and infectious diseases.
- The global warming and its potential effect on vector populations (specially in south European countries due to climate and biogeographical proximity of vector hotspot areas).
- The globalization of trade and the migratory movements as a driving force to transport vectors and pathogens.
- The complex relationship between urban and periurban zones (the role of wild animals on urban pest prevention/control operations).
- The alarming reduction on biocide (insecticide and rodenticide) availability.
- The increasing news about vector resistance to biocides.
- The peoples rejection to the use of biocides because of the associated health risks and environmental pollution.

In this context, the professional that work on pest control need to identified, evaluate and manage the entire environmental factors that interact and determine a pest problem. As Integrated Pest Management (IPM) strategies (see Brenner BL et al., 2003) seek and in order to accomplish this it is fundamental to identify and acquire the biological, social, economical and environmental data that determine the different pest situations. Understanding the environmental determinants of vector-borne disease transmission is critical to the development of an IPM plan (Glass et al., 2002). Furthermore, it is necessary to fore fill the scientific approach that will allow us to apply effective and environmental friendly solutions.

Geographical Information Systems (GIS) are tools that permit a technical and scientific based collection, processing and management (understanding) of data (Bosque Sendra J, 1997). For example, global positioning systems (GPS) and GIS are now essential vector-control tools used to control mosquitoes (WHO Public Health Significance of Urban Pests, 2008). In Health Madrid, GIS are not only a complex software solution for health problems, but a corporative strategy for the global processing of information.

The pest control sector, as almost all sectors, has always used maps. The inclusion of GIS technology incorporates a new perspective, allowing spatial data to perform a much more

important role due to the quickness in for filling complex geographical analysis or processes. This is a great opportunity for Health Madrid's Vector Control Unit to achieve a GIS-based programs (methodologies) for the prediction, prevention and control of cockroaches, rats, pigeons and other pests in Madrid city (García Howlett M and Cámara Vicario JM, 2009). Vector Control Unit already uses this technology in the following fields:

- Monitoring citizens complains-incidents.
- Monitoring pest populations.
- Pest control programs evaluations.
- Health surveillance.
- Spatial analysis and factors correlation. And continues exploring new research and development GIS opportunities, such as:
- Environmental risk assessment maps.
- Spatial epidemiology.
- Multivariant factor analysis.

The most complex and alluring task is the prediction of spatial pest population behaviour, geoestadistics are fundamental in succeeding in this objective. Analytical tools, such as spatial filtering (Kitron U et al., 1996; Thomson MC et al., 1999) discriminant analysis and temporal Fourier-series representation of satellite data (Hay SI et al., 1998; Rogers D et al., 1996) have been successfully applied to risk mapping of trypanosomiasis and malaria. Variograms, correlograms and fractals allow for consideration of heterogeneity over a range of spatial scales (Cullinan VI and Thomas JM, 1992). These and other spatial statistics and analytical tools have many applications for other vector-borne disease, where the interactions between environmental determinants, transmission patterns and disease risk may be more complex (Kitron U, 2000)

Before using any classical statistics or geostatistics support data is necessary. Vector Control Unit has recollected data for more than ten years, mostly citizens complains but also pest monitoring studies and census. All monitoring and census have been based on scientific methodology, therefore can be examined and confidence limits are known.

Urban heterogeneous environment and its geographical scale makes the prediction processes particularly difficult, as well as most urban pest ecological dynamics does. But the Vector Control Unit has began applying spatial analysis and geostatistics to pest population prediction, such as cockroaches (figure 1, after literature cited), rats and pigeons. Principally, prediction has been based on Kriging (see Webster R and Oliver MA, 2007). Thus they are simple geostatistical and spatial analysis approaches they already allow improvements in the Vector Control Units programs. This means better knowledge, efficiency and performance executing our role in local public health.

Brenner BL et al. (2003). Integrated pest management in an urban community: a successful partnership for prevention. Environmental Health Perspectives, 111:1649-1653.

Bosque Sendra J (1997). Sistemas de Información Geográfica. Madrid, Rialp (Ed.).

Cullinan VI and Thomas JM (1992) A comparison of cuantitative methods for examining landscape pattern and scale. Landscape Ecology, 7:211-227.

García Howlett M and Cámara Vicario JM (2009). Introducción de las tecnologías de la información geográfica en el control de las poblaciones de animales plaga/vector en la ciudad de Madrid. Revista de Salud Ambiental, 9(1):44-45.
Glass GE et al. (2002). Satellite imagery characterizes local animal reservoir populations of Sin Nombre virus in the southwestern United States. Proceedings of the National Academy of Sciences of the United States of America, 99:16817-16822.
Hay SI et al. (1998). From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. Parasitology Today, 14:306-313.
Kitron U et al. (1996). Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, using Landsat TM satellite imagery and GIS. J. Animal Ecology, 65, 371-380.
Kitron U (2000). Risk Maps: Transmission and burden of vector-borne diseases. Parasitology Today, 16:324-325.
Rogers D et al. (1996). Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Ann. Tropical Medicine and Parasitology, 90.225-241.
Thomson MC et al. (1999). Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. American J. of Tropical Medicine and Hygiene, 61:2-8.
Webster R and Oliver MA (2007). Geostatistics for environmental scientists. West Sussex, John Wiley & Sons Ltd..
WHO Public Health Significance of Urban Pests (2008). Copenhagen, World Health Organization Regional Office for Europe.

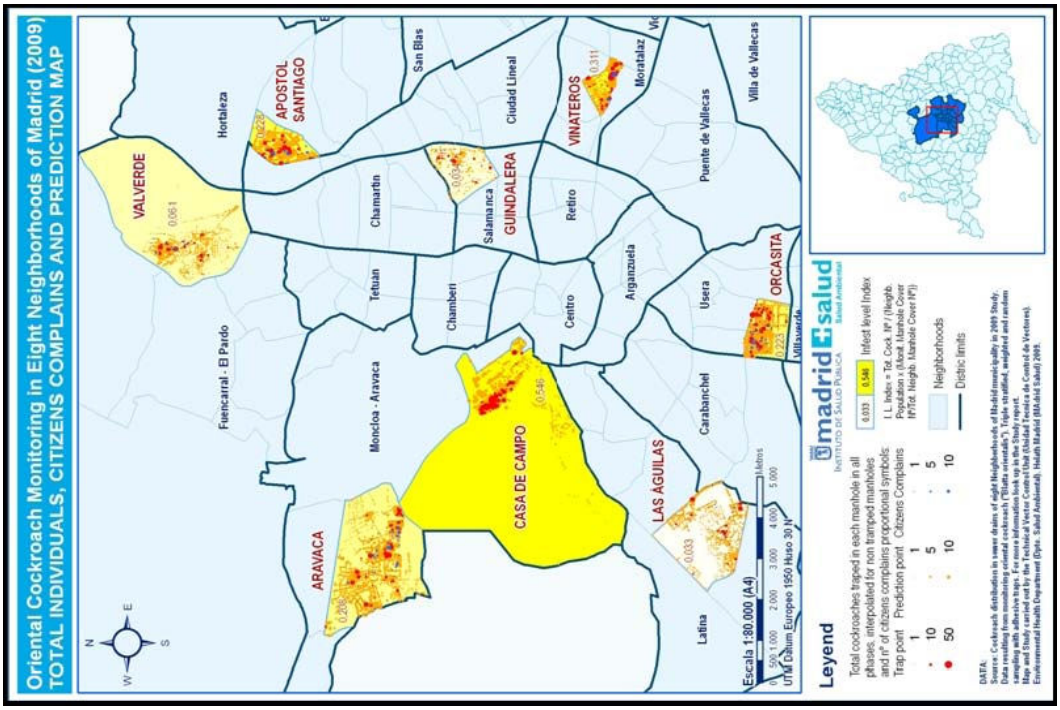


Figure 1: *Map example of pest population prediction in eight neighbourhoods of Madrid city.*

DEVELOPMENT OF A COMBINED SCHEME: MIXED HYBRID FINITE ELEMENT-FINITE VOLUME (MHFE-FV) FOR MODELING OF CONTAMINED TRANSPORT IN UNSATURATED POROUS MEDIA

A. Gheris[1]

[1]Department of Civil Engineering, University of Souk Ahras, BP1553, Souk Ahras 41000, Alegria, a.gheris@hotmail.com

Abstract

This work is devoted to developing numerical method for simulating flow and solute transport in porous media with multiphase taking tale of interaction solid/solute. In more precise, the problem studied is modeled by a coupled system consisting of an elliptical equation (for drainage), and an equation like convection-diffusion-reaction (for transfers). Numerical simulations were realistic for the two-dimensional problems confirm the stability and efficiency of the combined scheme, in the characterization of pollutant transport through the unsaturated zone of an industrial site.

Introduction

In the classical theory of drainage, the flow is assumed to be vertical in the unsaturated zone (USZ) and horizontal in the groundwater. This is the concept of the mixed layer. However, many authors (Vauclin et al, 1979, Clement et al 1996, Romano et al 1999, Kao et al, 2001; Silliman et al, 2002) emphasize the existence of a significant part of the infiltration horizontal elapsing in this zone. The importance of the unsaturated zone as an integral part of the hydrological cycle has long been recognized. The zone plays an inextricable role in many aspects of hydrology, including infiltration, soil moisture storage, evaporation, plant water uptake, groundwater recharge, runoff and erosion. Interest in the unsaturated zone has dramatically increased in recent years because of growing concern that the quality of the subsurface environment is being adversely affected by agricultural, industrial and municipal activities. In these areas the calculated and numerical simulation are essential because the experiments are very difficult if not impossible, by cons the predictions are vital.

The application of the classical conforming finite element method to fluid flow in porous media, in general, does not locally conserve mass [14] due to its global continuity requirement. Many specialized methods have been developed to resolve these difficulties. Eulerian methods use fixed spatial grids and incorporate some form of upstream weighting or other dissipation techniques in their formulations, [05], [13] and the references therein). Hence, they can eliminate the nonphysical oscillations present in the standard finite difference and finite element methods. Some of the Eulerian methods, such as the Godunov scheme, the total variation diminishing (TVD) schemes, and the ENO/WENO schemes, can resolve shock discontinuities from the nonlinear hyperbolic conservation laws. Characteristic methods take advantage of the hyperbolic nature of the convection-diffusion equations by utilizing a characteristic tracking to treat the advective component of the governing equation [07].These methods symmetrize the governing equations and significantly reduce the

temporal truncation errors. Thus, they allow large time steps to be used in numerical simulations without loss of accuracy, and lead to a greatly improved efficiency. However, most characteristic methods have difficulties in treating flux boundary conditions when a characteristic is tracked to the boundary of the domain, and fail to conserve mass. More recently, FV methods were developed and analyzed for convection-diffusion problems (see, for instance, [01], [02], [04], [06], [08], [09], [10], [11], [12]). There are various approaches in deriving FV approximations of convection-diffusion equations. The most general classification is obtained depending on the choice of: (1) the finite volumes and (2) the discrete space to which the approximate solution belongs. The domain is meshed and depending on whether the finite volumes are the elements from the original splitting or volumes around the vertices of the original splitting, we have correspondingly cell-centered and vertex-centered finite volume methods. For the vertex-centered finite volumes, depending on whether the discrete space is piecewise constant over the finite volumes or piecewise linear over the original mesh, we have correspondingly vertex-centered finite volume difference methods or vertex- centered finite volume element methods. The cell-centered finite volumes can lead to cell-centered finite volume difference methods or mixed methods.

In the first part of this work, we present a mathematical model that describes the flow and the solute transport through variably saturated porous media and multiphase, taking into account the interaction between the solid phase and liquid phase (effect of delay, degradation, and adsorption). The second part is devoted to analysis and numerical solution of the problem addressed in the first part, by using a combined diagram of mixed finite element and finite volume. The mixed finite element method is used for discretize the flow equation and the scheme of the finite volume is used to approximate the equation of convection-diffusion-reaction. For this we have developed using the computer code FreeFem++[03] a calculation program allowing therefore to solve a coupled system consisting of an elliptical equation discretized by the finite element method mixed hybrid and a linear parabolic equation type convection-diffusion-reaction, discretized by finite volume method type "vertex centered" semi-implicit in time. Finally, in the last section a numerical simulations are performed to verify the one hand the validity of our calculation scheme, and secondly to test the computational efficiency and the prediction of our model on a real example (case of the factory paint the town of Souk AHRAS, ALGERIA).

References

[01] A. Michel (2003), A finite volume scheme for two-phase immiscible flow in porous media, SIAM J. Numer. Anal. 41, No.4, pp1301-1317.

[02] D. Burkle and M. Ohlberger (2002), Adaptive finite volume methods for displacement problems in porous media, Comput. Vis. Sci. 5, No.2, pp95-106.

[03] F. Hecht, O. Pironneau, A. Le Hyaric, K. Ohtsuka, 2002. FreeFem++ Version 2.11. http://www.freefem.org/ff++

[04] J. Droniou (2003), Error estimates for the convergence of a finite volume discretization of convection-diffusion equations, J. Numer. Math. 11, No.1, pp1-32.

[05] M. A. Celia, T. F. Russell, I. Herrera, and R. E. Ewing (2002), An Eulerian-Lagrangian localized adjoint method for the advection-diffusion equation, Advances in Water Resources 13, pp187.206.
[06] M. Feistauer, J. Felcman and I. Straskraba (2003), Mathematical and Computational Methods for Compressible Flow (Oxford Sciences Publications).
[07] O. Pironneau (1982), On the transport-diffusion algorithm and its application to the Navier-Stokes equations, Numer. Math. 38 pp309.332.
[08] R. Eymard, T. Gallou¨et and R. Herbin (2000), The Finite Volume Method, in: Handbook of Numerical Analysis, eds P.G. Ciarlet and J.L. Lions. (North-Holland, Amsterdam) pp. 715.1022.
[09] R. Herbin (1995), An error estimate for a finite volume scheme for a diffusion-convection problem on a triangular mesh, Numer. Methods Partial Differ. Equations 11, No.2, pp165-173.
[10] R. Lazarov, I. D. Mishev, and P. S. Vassilevski (1996), Finite volume methods for convection-diffusion problems, SIAM J. Numer. Anal. 33, No.1, pp31-55.
[11] R. Lazarov and S. Tomov (2002), A posteriori error estimates for finite volume element approximations of convection-diffusion-reaction equations, Comput. Geosci. 6, No.3-4, pp 483-503.
[12] T. Gallouet, R. Herbin, and M. H. Vignal (2000), Error estimates on the approximate finite volume solution of convection diffusion equations with general boundary conditions, SIAM J. Numer. Anal. 37, No.6, pp1935-1972.
[13] T. F. Russell and M. A. Celia (2002), An overview of research on Eulerian.Langrangian localized adjoint methods (ELLAM), Advances in Water Resources 25. pp1215-1231.
[14] Z. Chen, R. E. Ewing and Z.-C. Shi, Editors (2000), Numerical Treatment of Multiphase Flows in Porous Media (Springer.Verlag, Berlin).

INTERACTIVE MODULES FOR THE VISUALIZATION AND TEACHING OF GEOSTATISTICAL CONCEPTS

A. Govaerts[1], A. Vervoort[2], P. Darius[3]

[1] Research unit mining, K.U.Leuven, Kasteelpark Arenberg 40 bus 2448, B-3001 Leuven, Belgium, annelies.govaerts@bwk.kuleuven.be, tel: +32(0)16321774, fax:+32 (0)16 321988
[2] Research unit mining, K.U.Leuven, Kasteelpark Arenberg 40 bus 2448, B-3001 Leuven, Belgium, andre.vervoort@bwk.kuleuven.be
[3] MeBios and Leuven Statistics Center, K.U.Leuven, Kasteelpark Arenberg 30 bus 2456, 3001 Leuven, Belgium, paul.darius@biw.kuleuven.be

The concepts of geostatistics are commonly difficult to understand for students, even for those with a strong mathematical background. Furthermore first time users of geostatistics often do not want to struggle through different books with the whole geostatistical background before using it. Too often this leads to using geostatistics as a black box. The data is imported in a software program, which calculates an experimental semivariogram, determines the 'best' semivariogram model and computes kriging estimates. These results are then considered as the only and best results, because they are based on geostatistics. This paper describes a number of generally accessible R-based (www.r-project.org) interactive modules, developed by the authors, which can be used in geostatistical courses but also by an individual user to easily learn the key concepts of geostatistics. These modules should allow the student or practitioner to see, in real time, the impact of deleting, moving or adding data points, changing parameters (e.g. lag distance, lag tolerance, directional tolerance for calculating an experimental semivariogram), using different semivariogram modeling techniques, using different kriging techniques, and so on. All the modules are made to work completely autonomous. This means that it is not required to have a dataset of your own. Those modules are therefore not meant to do a geostatistical study on a certain dataset (for this purpose there exists already a broad spectrum of good software packages). After learning the key concepts of geostatistics it should be easier for the user to do a good quality geostatistical study on their own dataset and to do a good interpretation of these results. The developing of these modules is based on experience with web-based tools for teaching statistical concepts, namely 'Vestac' and 'env2exp' (Darius et al. 2003; Darius et al. 2007) (see http://lstat.kuleuven.be/java/ and http://lstat.kuleuven.be/env2exp/). Those tools got several good critics (Cobb 2007; Wild 2007) and are already used in statistical courses all over the world.

In this paper different modules are discussed. The first module allows interactive experimentation with the experimental semivariogram of easy 1D-functions (e.g. Figure 1). Here the user can learn to understand the experimental semivariogram, but also about the effect of the number of sampling points, the lag distance, and other parameters.

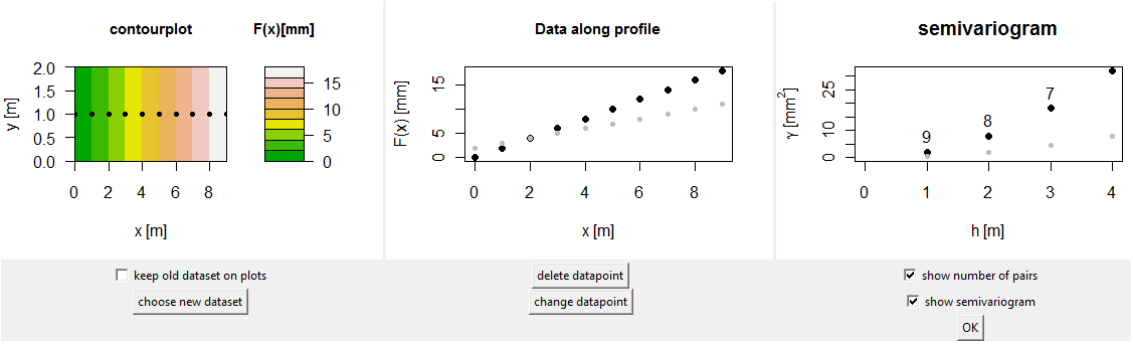


Figure 1 *Part of the first module, which allows interactive experimentation with the experimental semivariogram of easy 1D-functions. In this figure the semivariogram of two linear datasets (F(x)=ax+b) are compared. The user can also delete or change data points and see the effect on the semivariogram in real time.*

One of the other modules focuses on the effect of sampling on the semivariogram. The following questions are addressed in this module: Is it better to sample on a regular grid or randomly if nothing is known about the sampled parameter. And does this change if one has an idea about the spatial distribution of the sampled parameter? How does this sampling affect the experimental semivariogram that is deduced based on these data points?
In the coming months other modules are to be developed. The covered themes are e.g. the effect of the semivariogram model and the kriging parameters (for example search radius) on the kriging estimates, and thus the decisions based on these estimates; and the effect of taking into account the difference between volume and point measurements and point and volume estimates.
All modules will be shared on the internet. According to the freeware philosophy of R, all users will be able to adjust and develop the modules further.

Cobb G (2007) One possible frame for thinking about experiential learning. International Statistical review 75(3): 336-347
Darius P, Schrevens E, van der Knaap H, Portier K, Massonnet G, Lievens L, Duchateau L, Thas O. (2003) Using web-based tools for teaching statistical concepts and experimental concepts and experimentation skills. IASE/ISI Satellite 2003.
Darius P, Portier K, Schrevens E (2007) Virtual experiments and their uses in teaching experimental design. International Statistical review 75(3): 281-294
Wild C (2007) Virtual environments and the acceleration of experiential learning. International Statistical review 75(3), 332-335

SPATIAL DISTRIBUTION OF PLANT ANATOMICAL AND MORPHOLOGICAL CHARACTERISTICS FOR BIOMONITORING OF URBAN HABITAT QUALITY

F. Kardel[1], K. Wuyts[1], M. Babanezhad[2], U.W.A. Vitharana[3], A.R. Khavaninzadeh[1], T. Wuytack[1], R. Samson[1]

[1] Department of Bioscience Engineering, University of Antwerp, 2020 Antwerpen, Belgium, E-mail: fatemeh.kardel@ua.ac.be, Tel: +32 3 265 3682, Fax: +32 3 265 3225
[2] Department of Statistics, Golestan University, Gorgan, Golestan, Iran
[3] Department of Soil Science, University of Peradeniya, Sri Lanka

Introduction

Air quality is an important issue in many urbanised and industrialised cities all over the world. Monitoring of air quality is generally done by physico-chemical measurements of single atmospheric pollutants, thereby ignoring the combined effects pollutants might have on biological tissues. In addition, high costs associated with such measurements limit their spatial distribution, and thus the spatial resolution of the measured air quality. Biological monitoring procedures, however, enable low cost, high spatial resolution assessments of the combined effect of air quality, or habitat quality in general. The aims of this study were to evaluate the potential of anatomical and morphological leaf characteristics of *Plantago lanceolata* L. as bio-indicators of urban habitat quality and to assess their spatial distribution in an urban environment as a case study.

Material and methods

This study examined 169 sampling locations in the city of Ghent (Belgium), divided over four land use classes: green land (GL), urban green (UG), urban (U) and harbour and industry (HI). The spatial distribution of leaf characteristics of *P. lanceolata* L., including specific leaf area (SLA, $cm^2\ g^{-1}$), abaxial (the lower leaf side) stomatal density (SD, number of stomata per mm^2), stomatal pore surface (SPS, μm^2) and stomatal resistance (Rs, $s\ m^{-1}$), are investigated for evaluating urban habitat quality in the city of Gent. Details on the sampling are give in Kardel et al. (2010).

Firstly, normality of the data was tested for all considered leaf characteristics using a Kolmogorov-Smirnov test. Sampling date, expressed as the day of the year, was considered as a possible factor which could affect the leaf characteristics. Therefore, the effect of sampling date was checked using a linear regression model between the leaf characteristics and the day of year. Then, differences in characteristics between the four land use classes were tested for significance by using a one-way analysis of variance (ANOVA) procedure and a Tukey-HSD test using R 2.6.2 software.

The geostatistical analysis procedure 'simple kriging with varying local means' (SKLM) (Balasooriya et al., 2009) was used to map the spatial distribution of the leaf characteristics. For each characteristic, the mean values of the different land use classes were calculated. Subsequently, the residuals were calculated by taking the difference between each observed

value and the corresponding land use class mean. These residuals were used to compute the residual variograms required for the SKLM. The geostatistical analysis was carried out using the Geostatistics Software Library (GSLIB) software.

Results and discussion

SD, SPS and Rs were significantly different between all land use classes; the most extreme variations were observed between the two land use classes GL and HI. SD and Rs were increased by 34 and 8 % in the HI in comparison with the GL, respectively. SPS was decreased (by 36 %) in the HI in comparison with the GL.

Table 1 summarises the variogram parameters range, sill, nugget and relative nugget effect of the modeled variograms. The maximum value of semivariance (γ) is known as the sill. The lag distance (h) where the variogram reaches the sill is known as the range: the observations located within the range are spatially correlated. The semivariance where $h = 0$ is known as the nugget, and is an estimate of the residual error or spatially uncorrelated noise. The relative nugget effect is the ratio between the nugget and the sill, and quantifies the strength of spatial structure. All variograms of the considered characteristics show a bounded structure with ranges of 646-3740 m. The small relative nugget effects (8–51 %) depicted by all characteristics indicated a strong spatial correlation inside their ranges.

Table 1: *Omnidirectional variogram model parameters of SLA, SD, SPS and Rs with indication of the used model, the range, sill, nugget and the relative nugget effect*

Variable	Model	Range(m)	Sill[a]	Nugget[a]	Rel. nugget effect (%)
SLA ($cm^2\ g^{-1}$)	Exponential	1243	1584	572	36
SD (No. mm^{-2})	Exponential	1872	244	125	51
SPS (μm^2)	Exponential	3740	90	27	30
Rs ($s\ m^{-1}$)	Spherical	646	27	2	8

[a] The units are equal to the square of the variable units.

Geostatistical analysis revealed that the spatial distribution of SD, SPS, and Rs varied considerably across the study area, indicating clear differences and contrasts between different city areas. More specifically, these maps allowed us to distinguish between less and more polluted places in the considered area. For example, the nature reserve in the vicinity of the city centre, the city centre itself, areas with a high traffic density, both railway stations and the harbour area could easily be distinguished (e.g., Fig.1).

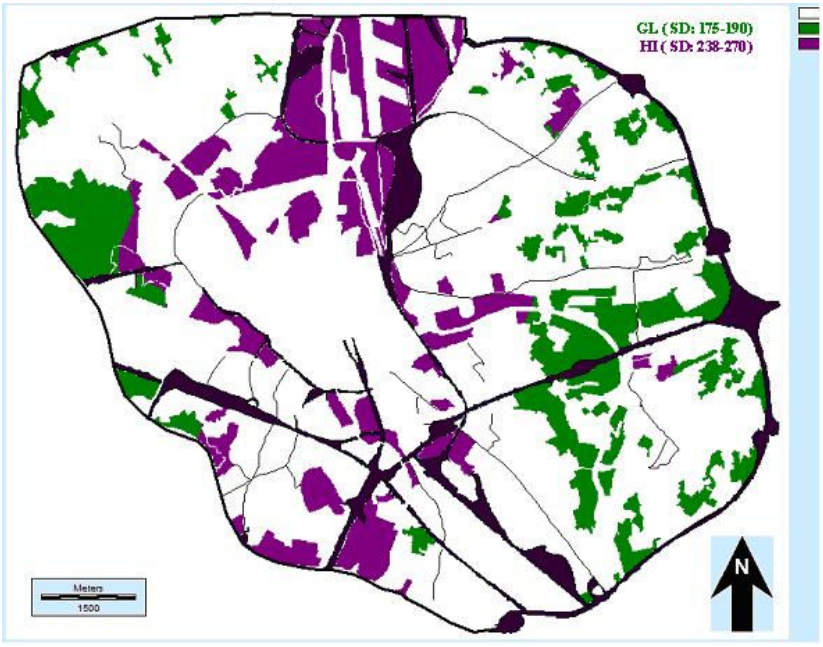


Figure 1: *Spatial distribution of SD (number of stomata per mm^2) in the two contrasting land use classes 'green land (GL)' and 'harbour and industry (HI)' in the study area, obtained by kriging (SKLM) (adapted from Kardel et al., 2010)*

Conclusion

Spatial analysis of anatomical and morphological leaf characteristics of the herb *P. lanceolata* showed that stomatal characteristics particularly are potentially good indicators for urban habitat quality. Reason is that stomata are the connection between the interior of the leaf (and the plant) and the atmosphere (Wuytack et al. in press). Our results indicate that bio-monitoring with common herbs is a promising, cost effective tool for urban habitat and air quality assessment, and, therefore, is a valuable tool for urban policy makers.

References:

Balasooriya BLWK, Samson R, Mbikwa F, Vitharana UWA, Boeckx P, Van Meirvenne M (2009) Biomonitoring of urban habitat quality by anatomical and chemical leaf characteristics. Environmental and Experimental Botany 65: 386-394.

Kardel F, Wuyts K, Babanezhad M, Vitharana UWA, Wuytack T, Potters G, Samson R (2010) Assessing urban habitat quality based on specific leaf area and stomatal characteristics of *Plantago lanceolata* L. Environmental Pollution 158: 788-794

Wuytack T, Verheyen K, Wuyts K, Adriaenssens S, Kardel F, Samson R (in press) The potential of bio-monitoring of air quality by using leaf characteristics of white willow (*Salix alba* L.). Environmental Monitoring and Assessment (DOI 10.1007/s10661-009-1271-4).

INFLUENCE OF ENVIRONMENTAL FACTORS ONFOREST SOIL SPATIAL DISTRIBUTION AND DIVERSITY IN LATVIA

R. Kasparinskis[1], O. Nikodemus[1]

[1]University of Latvia, Faculty of Geography and Earth Sciences, Riga, Latvia, e-mail: Raimonds.Kasparinskis@lu.lv; telephone: +371 67332704, fax: +371 67332704

SPATIAL RISK OF DISEASE IN CASE-CONTROL STUDIES WITH MULTINOMIAL RESPONSES: AN OCCUPATIONAL ACCIDENT EXAMPLE.

A.C.C.N. Mafra[1], R. Cordeiro[1], L.B. Nucci[1], C. Stephan[1]

[1]State University of Campinas, 13083-970, Campinas, Brazil, anacarol.nunes@gmail.com, +55 (0) 19 3544-1000

The spatial distribution of disease risk has always been a concern in Epidemiology. Particularly, in the last two decades, several techniques of spatial analysis for epidemiological data have been developed to estimate the variation of risk in space.

However, several studies based on scales or multinomal response still use logistic regression models for data analysis without incorporating spatial effects. This is partially due to the reduced number of techniques and computational tools which allow treating adequately multinomial responses occurring in epidemiological data within the spatial setting.

This work aims to contribute to overcoming this limitation. To do this, when the response of a case-control study is multinominal, the authors define the spatial risk and present how to obtain, analyse and map it, including acquaintance of the estimates statistical significance.

The definition of spatial risk in multinomial analysis comes from Bithell (1990) that define the spatial risk functions for epidemiological studies with binomial responses.

Kelsall and Diggle (1998) proposed a procedure to estimate the spatial risk by means of a generalized additive model (GAM) in case-control studies, which allows for the inclusion of other non-spatial covariates in the study that could be related to the outcome of interest, where the space is analised as a bidimensional non-parametric function.

Here the authors, after defined the new spatial risk, presented an algorithm to estimate GAM when the response is multinomial and the non-parametric function, to study the space, is estimated as a bidimensional kernel. This development is based on article and book written by Kelsall and Diggle (1998) and Hastie and Tibshirani (1990), respectively. The multinomial model used was the polytomous logistic model (Ananth and Kleinbaum 1997). To obtain a significance analysis of the estimated spatial risk a simulation method was introduced.

All programming was made in software R2.7 and the maps were built on ArcMap 9.2.

Data from a population-based case-control study of occupational accidents in a Brazilian city (Stephan 2008) were adjusted with the response variable classified in three categories: serious cases, mild cases and controls. Along the spatial analysis, the informations about: age; sex; schooling; if the employee has formal contract; if, despite any formal contract, the worker has some rights; if the labor is domestic and if is in streets were included.

Table 1 shows the parametric analysis of the estimated GAM. Table 2 gives a deviance analysis comparing the complete model with the models without space and without parametric covariates. Figure 1 shows the spatial risk estimated by GAM with covariates and Figure 2, without covariates. Both figures include significance analysis.

Table 1: *Estimated measures for occupational accident. Response: serious, mild and control. N=2323.*

	Serious vs Controls				Mild vs Controls			
	OR	LL(OR)	UL(OR)	P-value	OR	LL(OR)	UL(OR)	P-value
Age (Years)	0.989	0.977	1.000	0.053	0.969	0.962	0.976	<0.001
Sex (Male)	2.481	1.682	3.660	<0.001	2.146	1.722	2.674	<0.001
Schooling (Years)	0.953	0.917	0.990	0.013	0.929	0.909	0.949	<0.001
Formal Contract (Yes)	2.674	1.489	4.800	0.001	1.314	1.038	1.663	0.023
Labor Rights (Yes)	0.546	0.304	0.979	0.042	1.779	1.392	2.275	<0.001
Domestic Work (Yes)	0.458	0.237	0.887	0.021	0.787	0.587	1.057	0.111
Street Work (Yes)	0.868	0.667	1.128	0.289	0.765	0.657	0.892	0.001

OR: Odds Ratio; LL: Lower Limit; UL: Upper Limit.
Source: Stephan (2008).

Table 2: *Deviance Analysis of the adjusted models. Response: serious, mild and control. N=2323.*

	Res. DF	Res. Deviance	DF	Deviance	P(>|Chi|)
Semiparametric Model (complete model)	4628	3621.361	-	-	-
Parametric Model (without space)	4630	3685.968	2	64.607	<0.0001
Non-Parametric Model (without covariates)	4642	3828.129	14	206.768	<0.0001

DF: Degrees of Freedom; Res.:Residual.
Source: Stephan (2008).

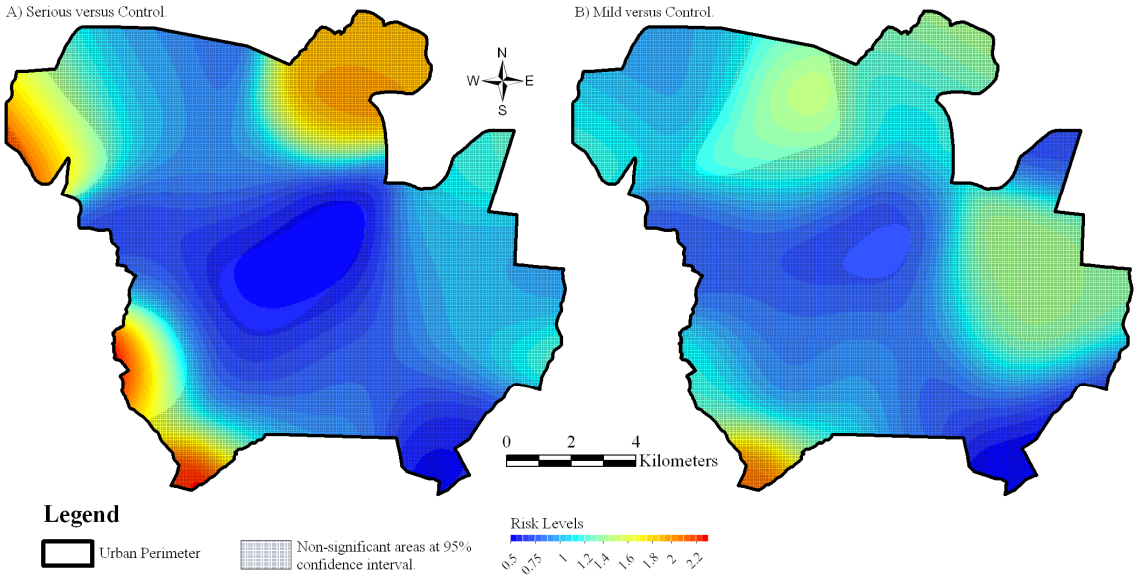


Figure 1: *Spatial Risk of Occupational Accident adjusted with covariates. Piracicaba - SP - Brazil (2007).*

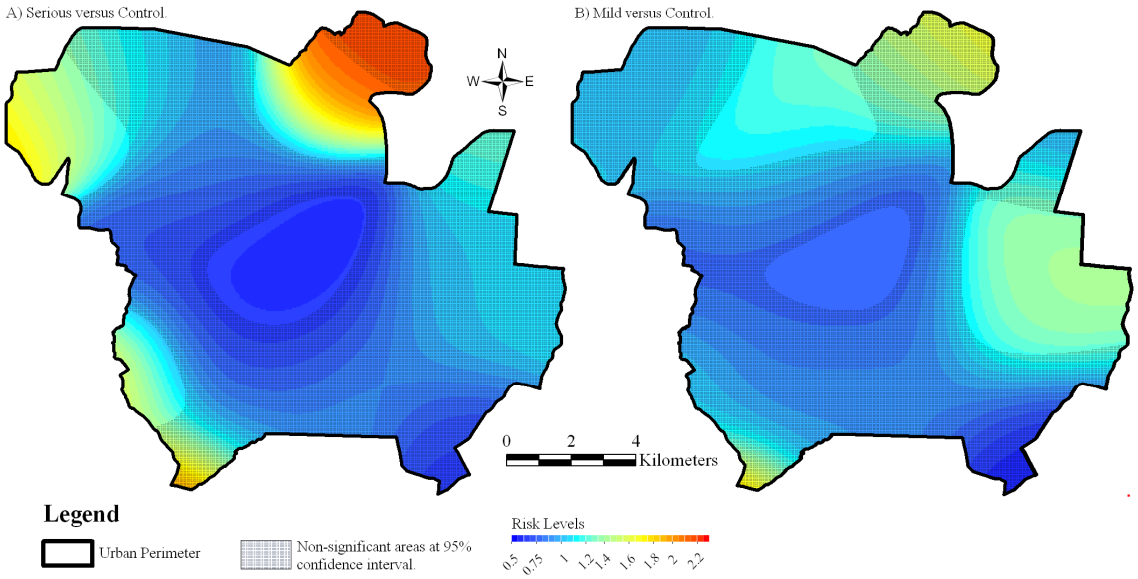


Figure 2: *Spatial Risk of Occupational Accident adjusted without covariates. Piracicaba - SP - Brazil (2007).*

In the example, the analysis has found areas of significant increased relative risk and protection for occupational accidents that varied depending on the level of comparison. Some areas had twice the risk compared to the average of the region studied when considering serious accident, meanwhile the mild cases had lesser. In parametric variables studied, different risk and protection factors were found for the two levels of the cases. The deviance analysis indicates the global model as the best.

This work brings the complete way to analyze data from case-control studies with multinomial response where the spatial risk has to be analyzed. As a bonus, gives the opportunity to include other covariates in the spatial analysis, controlling all study informations in one global model.

REFERENCES

1. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology* 1997; 26:1323-1333.
2. Bithell J. An application of density estimation to geographical epidemiology. *Statistics in Medicine* 1990; 9:691-701.
3. Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman and Hall, 1990.
4. Kelsall JE,.Diggle PJ. Spatial variation in risk of disease: a nonparametric binary regression approach. *Applied Statistics* 1998; 47:559-73.
5. Stephan C. Distribuição do risco de acidente do trabalho entre trabalhadores precarizados de Piracicaba. - Campinas, SP : [w.n.], 2008.

REGIONAL CHARACTERIZATION OF SOIL HEAVY METALS IN A FORMER WORLD WAR I BATTLE AREA

E. Meerschman, M. Van Meirvenne

Research Group Soil Spatial Inventory Techniques, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, B-9000 Gent, Belgium, eef.meerschman@ugent.be, Tel: 0032 09 2646042, Fax: 0032 09 2646247

Introduction

Only recently soil contamination has been acknowledged to be a form of collateral damage of the First World War (WWI). Van Meirvenne et al. (2008) showed an enrichment of copper content by 6 mg kg^{-1} in the soil around Ypres (Belgium) to be a legacy of shelling during WWI. However, their conclusion was based on a regional data base of the entire province of West-Flanders (3144 km^2) containing a limited number of soil samples in the war zone (640 km^2). Furthermore, the chemical composition of the WWI artillery consisted not only of Cu. Also Zn was used in the top fuse of a shell, shrapnel balls and bullets were made out of Pb and As was used in toxic gas.

Even though no risk for human health was found, the Flemish government became conscious of the possible effects of WWI on the geochemical soil composition. As such, they provided the means to take 300 additional soil samples in order to (1) increase the detail of the inventory and to (2) expand the data base to include eight heavy metals As, Cd, Cr, Cu, Hg, Ni, Pb and Zn. To constrain the logistic costs we opted for a two-phase sampling strategy with each phase consisting of 100 samples. The other third was foreseen to be used as an independent validation dataset. This paper discusses the two-phase sampling strategy and shows the main estimation maps created with the enlarged dataset.

Study site

The study area of 640 km² corresponds to the former WWI war zone along the Western front, which extended from the Belgian coast to the Swiss border (Figure 1). The front line curved along the east of Ypres, which was the centre of intense and sustained warfare between the German and the Allied forces. The soil in the war zone consists of Pleistocene (Weichsilian) wind-blown sediments deposited over Tertiary marine clayey sediments (Ypresian). The topsoil texture is mainly sandy silt in the north and centre of the area and changes gradually to silt in the south. In the utmost south and north, clayey soils occur. At present, most of the area is used for agriculture.

Data collection

Initially, samples were available at 253 locations. Most of them were provided by the Public Flemish Waste Agency (OVAM), which systematically records data from soil investigation studies. The procedure to determine the heavy metal concentrations involved a microwave destruction of the air-dry fine-earth fraction using HCl, HNO_3 and HF. In the digest, the metals were analyzed using ICP-AES, in conformity with the Flemish soil legislation.

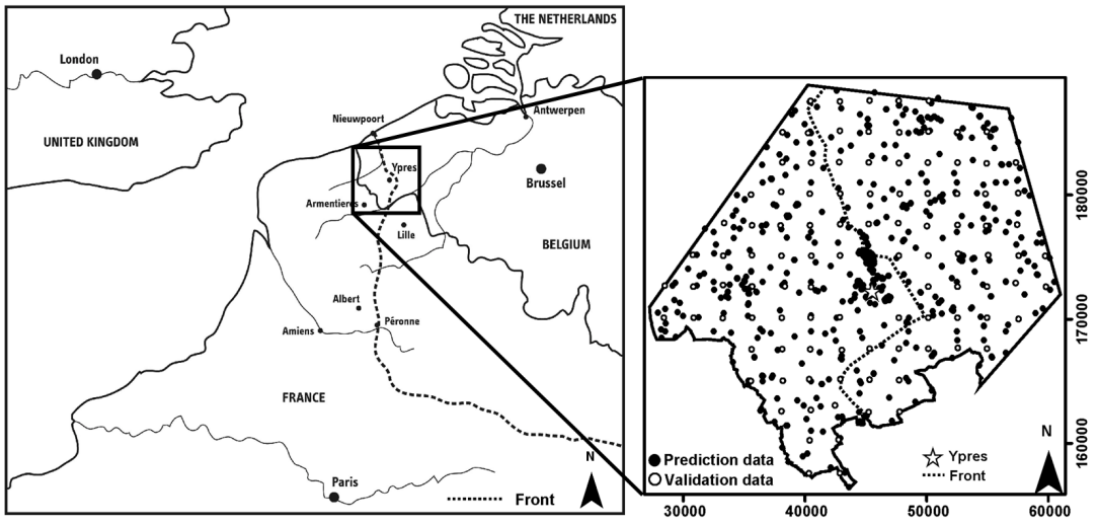


Figure 1: *Map of the Western front (Flanders Fields Documentation Centre Ypres) with indication of the study area (left) and the available dataset with 629 prediction data and 102 validation data (right).*

We performed a two-phase sampling strategy aiming to increase the level of detail inside the study area rather than to delineate a contaminated area in respect to a critical threshold. Because the initial dataset was clustered, sampling in less densely sampled regions was the first objective. Therefore, selection criteria based on distance measures and the ordinary kriging variance were used. After filling the spatial gaps, we focused on the local uncertainty in the second sampling stage. Therefore, we used both the conditional variance and the conditional coefficient of variation as complementary selection criteria. A clear improvement of four validation indices demonstrated the benefits of our data collection. Finally, after a screening of the OVAM data base for the most recent measurements, we ended up with 731 data points (Figure 1).

Summary statistics revealed that Cu, Pb and Zn are the dominant heavy metals in the soil around Ypres. Because their mean is higher than the legal background value, we can conclude that there is a regional enrichment for Cu, Pb and Zn, whereas there is no indication of a regional enrichment for As. On average the soil is not contaminated since for each heavy metal, the median is below or equal to the background value. Nevertheless, at a local scale concentrations above the sanitation threshold were detected.

Final maps

To yield the prediction maps, all available data were used, including validation data. Because of the large amount of censored data for As, Cd and Hg, indicator kriging was applied with the detection limit as the first threshold. The other data sets were interpolated using sequential Gaussian simulation. Before, the independent validation had revealed that the E-type was the best estimation for Cr and Ni and the M-type the best for Cu, Pb and Zn, which can be explained by the strong positive skewness of the latter data distributions.

Whereas the prediction maps for Cr and Ni show a link between increased concentrations and the occurrence of clay soils, patterns of increased Cu and Pb estimations were not related to soil texture (Figure 2). Apart from some isolated patches with elevated concentrations, a continuous band appeared around the WWI front line.

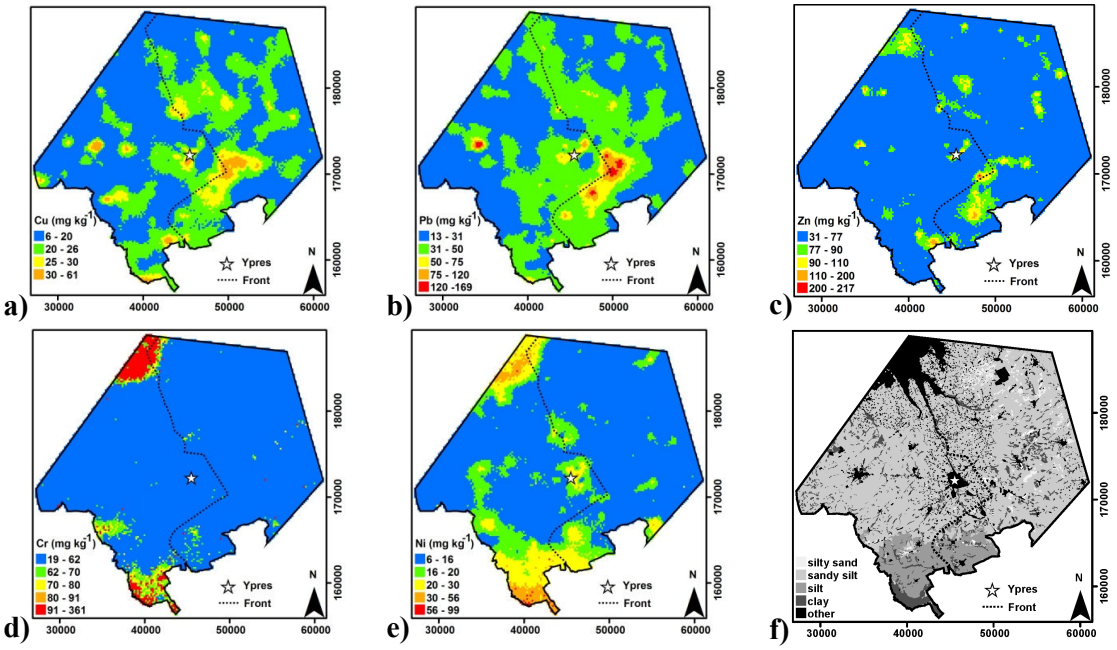


Figure 2: *Prediction maps for the topsoil a) Cu, b) Pb, c) Zn, d) Cr, e) Ni content with indication of the centre of Ypres and the frontline which remained more or less stable between 1915 and 1917 and f) texture map of the study area modified from the digital soil map of Flanders.*

The largest zone with increased Cu and Pb concentrations was situated in a prominent part of the Ypres Salient, where intensive battles took place. For Zn 90% of the estimations were below or equal to the background value. This might be due to the higher content of Cu compared to Zn in brass and the higher mobility of Zn in soil. We considered the differences in spatial structure of the heavy metals as an indication of the uniqueness of the source of Cu, Pb and, to a lesser degree, of Zn.

Conclusion

Even though the soil around Ypres is in general not contaminated according to critical legal thresholds of heavy metals, it was clearly shown that WWI affected the concentrations and spatial occurrence of Cu and especially Pb. Differences in parent material, metallurgical industry, amendments by slurry and sewage sludge as possible heavy metal sources were considered and excluded by Van Meirvenne et al. (2008). Consequently, we identified WWI activities as being responsible for the occurrence of elevated concentrations of Cu, Pb and Zn.

Reference

Van Meirvenne M, Meklit T, Verstraete S, De Boever M, Tack F (2008) Could shelling in the First World War have increased copper concentrations in the soil around Ypres? European Journal of Soil Science 59: 372-379.

MAPPING MONTHLY MEAN TEMPERATURE IN GALICIA (NW SPAIN) USING GEOSTATISTICAL PROCEDURES

J.M. Mirás-Avalos, P. Sande-Fouz, R.M. Mestas-Valero, A. Paz-González

Facultad de Ciencias, Universidade da Coruña, Campus A Zapateira s/n, A Coruña, Spain, jmirasa@udc.es, +34 981167000 ext. 2102, +34 981167065

Climatic variables, such as temperature, are important controls on surface energy balance and ecosystem processes, and they are also inputs to ecological models that yield estimates over large spatial domains. However, meteorological networks provide point data observations, which are irregular and scarce, so this information must be extended to the entire study area through a process of spatial interpolation. The results from different interpolators depend on the sampling density (Dirks et al. 1998); when this is low, Geostatistics is more convenient than deterministic procedures. Kriging methods have shown considerable advantages compared to deterministic interpolations in rainfall (Goovaerts 2000) and temperature (Jarvis and Stuart 2001) estimation.

The aim of this study was to create continuous surfaces from point monthly mean temperature observations in Galicia (NW Spain). Thus, a characterization of the spatial variability of temperature was carried out. Then monthly mean temperature was mapped by inverse distance weighting (IDW) and three different kriging techniques: ordinary kriging (OK), simple kriging (SK) and kriging with external drift (KED) using altitude as secondary information.

Galicia is the most northwestern region of Spain, with 29570 km^2, approximately. The region has a climate with local characteristics, due to maritime environment and rough relief. A decrease in temperature of 0.65 ºC each 100 m in height was estimated. The experimental dataset used in this exercise includes one year (2009) of mean monthly air temperature over 141 stations. These stations cover the region almost homogeneously. The mean density of meteorological stations is low, one for every 210 km^2 (14.5 km x 14.5 km). The range of distances between the stations extends from 5 to 280 km, approximately.

The main statistical moments were analyzed. Shapiro-Wilk's test was used to check normality on the datasets. Pearson's r coefficient was used to determine the correlation between temperature and altitude in order to justify the use of KED (Webster and Oliver 2001). IDW was used as a reference to map mean monthly temperature. Spatial variability was evaluated through semivariogram calculation, model fitting and comparison of each dataset. Model performance was checked using cross-validation (Webster and Oliver 2001). Only omnidirectional semivariograms were calculated due to the limited number of available data. Geostatistical interpolation was performed through OK, SK and KED. Theoretical background and mathematical formulation can be found in Webster and Oliver (2001). KED incorporates secondary information related to the primary variable; in this exercise, derived from a digital elevation model. Geostatistical analyses were performed using gstat (Pebesma 2004).

Mean monthly air temperature during the study period varied from 6.5 ºC in January to 18.7 ºC in August. Shapiro-Wilk's test showed that temperature datasets followed a normal distribution only in a few months however, no data transformation was performed. Correlation between altitude and mean monthly temperature ranged from –0.5 to –0.97, significant at the 1% level in all cases, justifying the use of KED.

Two different theoretical functions (exponential and spherical) with variable nugget effects, depending on the month, were fitted to the experimental semivariograms. Spatial dependence was observed in the 12 months analyzed. Spherical models were fitted to 6 months and exponential models to the other 6 months. Figure 1 presents the models fitted to April and November 2009 as examples. Hence, monthly mean temperature in Galicia is not temporarily stationary, namely its spatial structure may vary from one month to another. The nugget effects varied between 12.5 and 72.3% of the sill value. Ranges were less than 15 km once, between 15 and 30 km in 3 months and greater than 30 km 8 times. Cross-validation results showed that the theoretical structures adequately described the spatial dependence of monthly mean temperature.

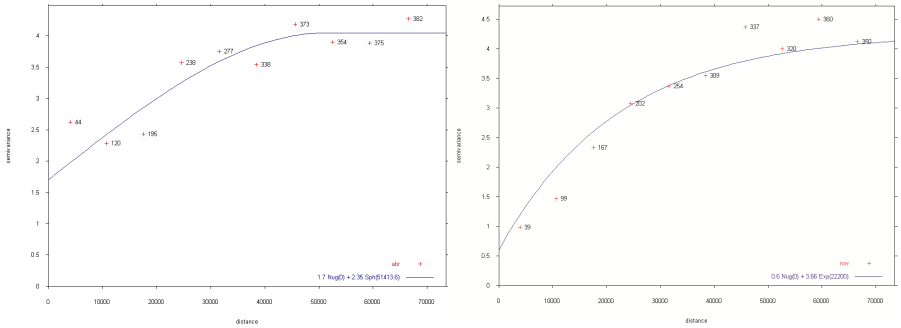


Figure 1. *Experimental semivariograms and their fitted models for monthly mean temperature of April (left) and November (right) 2009.*

In general, OK and SK maps were very similar due to the characteristics of both interpolators and gave mean monthly temperature distribution patterns similar to those obtained by IDW. However, the maximum and minimum values were slightly underestimated. Kriging errors were high in most of the study area except in areas near to the meteorological stations due to the low semivariogram ranges (Figure 2).

KED tended to overestimate maximum values. Usually, KED errors showed lower values than those from OK. KED estimates reflected the variability of the secondary attribute however this is not proof that these trends pertain to monthly mean temperature (Figure 2).

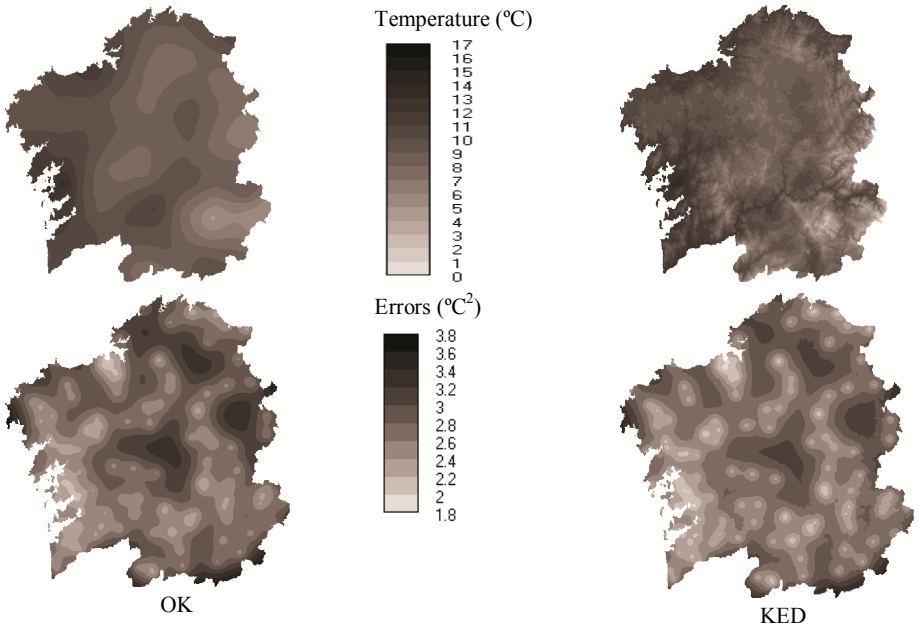


Figure 2. *Estimation and error maps of monthly mean temperature for April 2009 generated by OK and KED.*

The results from the different geostatistical procedures were very similar thus, no great improvements were obtained by incorporating altitude as secondary information through KED. Extending the study to longer datasets and shorter time scales may lead to different conclusions due to trends which are masked at the monthly level.

References

Dirks KN, Hay JE, Stow CD, Harris D (1998) High-resolution studies of rainfall on Norfolk Island part II: interpolation of rainfall data. J Hydrol 208: 187-193.

Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. J Hydrol 228: 113-129. doi: 10.1016/S0022-1694(00)00144-X

Jarvis EH, Stuart N (2001) A comparison among strategies for interpolating maximum and minimum daily air temperatures. J Appl Meteorol 40: 1075-1084

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30: 683-691. doi: 10.1016/j.cageo.2004.03.012

Webster R, Oliver MA (2001) Geostatistics for environmental scientists. John Wiley and Sons, Chichester, England, p. 149.

SPATIAL VARIABILITY OF PH, REDOX POTENTIAL AND EXTRACTABLE Fe, Mn AND Zn ON A PADDY SOIL AS A FUNCTION OF CROP STAGE AND LIME AMENDMENT

L.A. Morales[1], E. Vidal Vázquez[2], J. Paz-Ferreiro[2]

[1]Facultad de Ciencias Agrarias. Universidad Nacional de Corrientes. Sargento Cabral, sn, Corrientes, Argentina. Email: luiaber@yahoo.es
[2]Facultad de Ciencias. Universidade da Coruña. Campus A Zapateira. C.P. 15071. A Coruña (Spain). Email. jpaz@udc.es

Soil spatial variability is a natural occurring and/or management induced feature that is important for site-specific management practices such as variable rate fertilization. Since rice paddy fields are flooded and flat, apparently they should be homogeneous and, therefore, it had been hypothesized that spatial variability in yields and soils might be negligible. However, significant levels of variability in soil general properties, soil nutrients and rice yields have been observed, and this even in small paddy fields. Moreover, soil properties (pH, organic mater content, available P and K) have been found to be spatially structured on paddy soils both under anaerobic and aerobic land uses (Sun et al., 2003).

This study compares the effect of lime additions on the spatial variability of pH, redox potential (Eh) and available Fe, Mn and Zn from an acid rice soil, during three different growth stages. The objectives of this research were to assess the effect of lime amendment on the studied soil properties and to assess its patterns of spatial variability during the successive sampling periods.

The study site was a paddy field cropped with rice (*Oryza sativa* L.) of 5.1 ha in area, situated in Corrientes province, Argentina According to the U.S. Soil Taxonomy, the soil is a Typic Plintacualf. A control with no lime addition and two dolomite doses in amount of 625 and 1250 $kg.ha^{-1}$ were tested. Soil samples were taken at three different times during the vegetative period (seedling, tillering, i. e. bunch formation and flowering) of the rice crop. A systematic sampling scheme was used. In each of the three treatments and the three sampling dates, ninety-six soil samples were taken on a 20 x 11.9 m grid covering the 5.1-ha site. The used sampling scheme was intended to provide sufficient numbers of data pairs over a wide range of distances, thus allowing identification of short- and long- range variations. Determinations were made in all samples for pH and Eh. Samples also were analyzed for extractable Fe, Mn and Zn (Abreu et al., 2007).

Exploratory statistical analysis included examination of mean values, coefficients of variation, maximum and minimum values. Proximity to the normal distribution was judged on the basis of Kolmogorov-Smirnov test. Spatial variability was assessed by means of semivariogram analysis. The kriging approximation was used for interpolation and mapping purposes. As expected, soil pH increased, whereas soil Eh decreased as a function of time after flooding. Very high increases of extractable mean Fe contents and high increases of Mn contents were also recorded after flooding due to liberation of these elements as a

consequence of reduction processes induced by anaerobic conditions. For example, Fe content was about two times higher during the second sampling date and about four times higher during the third sampling date when compared with the initial content in aerobic conditions. Mean extractable Zn contents significantly decreased ($P<0.05$) after flooding indicating fixation of this element. Fe, Mn and Zn contents in general increased with increasing dolomite dose, mainly in anaerobic conditions. Therefore, there was an effect of lime amendment on extractable Fe, Mn and Zn concentrations.

Table 1. *Best fitted model parameters and determination coefficients during validation for extractable Fe, Mn and Zn at three sampling dates, and on three lime treatments*.

Treatment	Code	Model	C_0	C_1	a (m)	r^2	$C_0/(C_0+C_1)$
			First sampling				
control	Fe-10	Spherical	0.0	268.1	41.3	0.917	0.0
control	Mn-10	Spherical	6.2	52.3	47.3	0.878	10.5
control	Zn-10	Spherical	0.023	0.059	57.9	0.784	28.0
625 kg ha^{-1}	Fe-11	Exponential	0.0	271.5	64.4	0.918	0.0
625 kg ha^{-1}	Mn-11	Spherical	19.56	93.5	56.4	0.892	17.3
625 kg ha^{-1}	Zn-11	Spherical	0.017	0.046	57.9	0.885	3.5
1250 kg ha^{-}	Fe-12	Exponential	62.5	227.3	48.0	0.926	21.6
1250 kg ha^{-1}	Mn-12	Spherical	9.28	113.7	68.3	0.888	7.5
1250 kg ha^{-1}	Zn-12	Spherical	0.00072	0.050	59.1	0.903	12.5
			Second sampling				
control	Fe-20	Exponential	0.0	3868.3	47.6	0.892	0.0
control	Mn-20	Exponential	0.0	465.1	58.3	0.918	0.0
control	Zn-20	Spherical	0.011	0.051	79.9	0.849	17.2
625 kg ha^{-1}	Fe-21	Spherical	1511.4	9139.6	54.0	0.894	14.2
625 kg ha^{-1}	Mn-21	Spherical	24.7	277.7	65.4	0.961	8.2
625 kg ha^{-1}	Zn-21	Spherical	0.007	0.049	73.6	0.936	13.0
1250 kg ha^{-}	Fe-22	Spherical	94.5	10954.1	44.6	0.952	0.9
1250 kg ha^{-1}	Mn-22	Spherical	119.9	512.4	44.0	0.914	19.0
1250 kg ha^{-1}	Zn-22	Spherical	0.013	0.039	66.5	0.916	24.8
			Third sampling				
control	Fe-30	Exponential	1378.2	4150.3	63.1	0.968	24.9
control	Mn-30	Spherical	82.3	1245.1	45.6	0.932	6.2
control	Zn-30	Spherical	0.012	0.044	82.5	0.937	20.9
625 kg ha^{-1}	Fe-31	Exponential	0.0	13967.9	62.8	0.892	0.0
625 kg ha^{-1}	Mn-31	Spherical	100.6	220.5	73.6	0.934	31.3
625 kg ha^{-1}	Zn-31	Spherical	0.001	0.023	70.8	0.953	4.9
1250 kg ha^{-}	Fe-32	Spherical	20701.1	100817.3	55.7	0.930	17.0
1250 kg ha^{-1}	Mn-32	Spherical	0.0	861.7	73.4	0.974	0.0
1250 kg ha^{-1}	Zn-32	Spherical	0.005	0.020	52.4	0.908	19.5

(C_0 = nugget effect; C_1 = sill; a = range; r^2 = correlation coefficient).

Spatial variability of Fe, Mn and Zn on the study rice field was far from negligible both on aerobic and on anaerobic conditions. In general, semivariograms for pH, Eh, Fe, Mn and Zn could be adjusted quite well, over the spatial scale of interest. A stable sill was reached in all the 45 data sets (five variables x three treatments x three sampling dates).Experimental

semivariograms were best fitted by spherical or exponential models. The nugget variances were all below 31.3 % of the sill value (Table 1).

Kriging contour maps that were drawn using the fitted semivariogram models show discrete patches or small zones with distinct pH, Eh, Fe, Mn or Zn values. The position of patches with maxima and minima values for the study variables in the kriging maps changed between successive sampling dates illustrating the lack of temporal stability of the pattern of spatial distribution. Moreover, patterns of spatial variation clearly show disparities between the three dolomite treatments for a given sampling date. Kriging contour maps also show that there was no correspondence between patches with maximum or minimum values of Fe, Mn and Zn. So, in some cases it could be observed that areas with maximum Fe values show a trend to match those with minimum Mn values and vice versa.

Dynamics of extractable Fe, Mn and Zn is considered to be complex and it is affected by numerous internal factors or soil properties such organic matter content, elemental composition or element speciation, pH and potential redox (Eh), etc, which determine a number of chemical reactions and over a rice field are far from homogeneous. On the other hand, although all the experimental units were managed similarly uneven water application due to inaccuracies of the flooding system and/or microrelief features could be also a source of spatial variability. Therefore, texture, soil mineral and organic composition, uneven flooding and microtopographic irregularities are possible factors influencing the variability in the studied soil properties

The above results show that there is a great potential to apply site specific technologies and management strategies for rice production in the studied soil. Measurement and management of the small scale variability in rice fields raises both, challenges and opportunities for precision agriculture. Small scale variability requires more careful consideration regarding both determination of design parameters and data processing issues. Precision agriculture strategies should be further investigated, depending on the feasibility of controlling spatiotemporal variability.

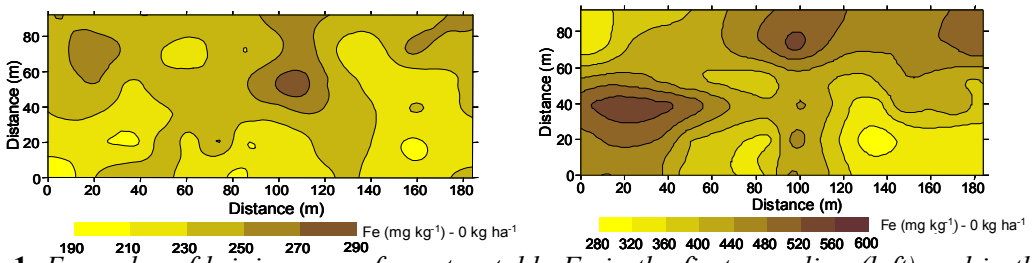


Figure 1. *Examples of kriging maps for extractable Fe in the first sampling (left) and in the second sampling (right).*

References

Abreu CA, Lopes AS, Santos GCG (2007) Micronutrientes. In *Fertilidade do solo*, *eds*. R.F. Novais, V.H. Alvarez, N.F. Barros, R.L.F. Fontes, R.B. Cantarutti, and J.C.L. Neves, 645-736. Viçosa: SBCS.

Sun B, Zhou S, Zhao Q (2003) Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. Geoderma, 115: 85-99.

A FIRST APPROXIMATION TO MANTLE CONVETION: A RAYLEIGH-BÉNARD PROBLEM WITH VARIABLE VISCOSITY AND NON UNIFORM HEATING

F. Pla, H. Herrero

Depto. de Matemáticas, Facultad de Ciencias Químicas, Universidad de Castilla-La Mancha
13071, Ciudad Real, Spain, Francisco.Pla@uclm.es, Henar.Herrero@uclm.es

Thermal convection in the upper mantle is the driving force of the plate tectonics which causes the continental drift [4, 7, 9]. So, in recent years interest in the study of thermal convection with characteristics of the mantle increased [1, 8, 12]. Some of these characteristics are depth-dependent viscosity or thermal conductivity or temperature-dependent thermal conductivity or viscosity [3, 5, 8, 13]. In Refs. [2, 8, 11] large viscosity variations are considered. On this line, this work deals with convection in a fluid which is subject to large viscosity variations, by imposing on the equations of motion a strong dependence of the viscosity with temperature.

Horizontal gradients are also present in mantle convection, as the heating in not homogeneous [6]. Therefore, we propose the study of a Rayleigh-Bénard problem with non uniform heating and temperature-dependent viscosity as a first approximation to mantle convection. The viscosity has an exponential dependence on temperature [8, 10] and the domain of the problem is a rectangular layer finite in the x-z plane and infinite in the y-direction. A numerical linear stability analysis of the basic state solutions is performed and the results are compared with the constant viscosity and uniform heating case. It is found that horizontal temperature gradients are responsible of the three dimensional structure of the bifurcation solutions. This suggests that localizad self generated horizontal gradients may influence the mantle convection evolution.

[1] Bercovici D (2003) The generation of plate tectonics from mantle convection. Earth and Planetary Science Letters 205: 107-121
[2] Booker JR (1976) Thermal convection with strongly temperature-dependent viscosity. J Fluid Mech 76: 741-754
[3] Bunge HP, Richards MA, Baumgardner JR (1996) Effects of depth-dependent viscosity on the planform of mantle convection. Nature 379: 436-438
[4] Elder JW (1968) Convection key to dynamical Geology. Science Progress 56: 1-10
[5] Foster TD (1969) Convection in a Variable Viscosity Fluid Heated from Within. J Geophys Res 74: 685-693
[6] Labrosse S, Jaupart C (2007) Thermal evolution of the Earth: Secular changes and fluctuations of plate characteristics. Earth and Planetary Science Letters 260: 465-481
[7] Mckenzie DP, Parker RL (1967) The North Pacific: An example of Tectonics on a Sphere. Nature 216: 1276-1280

[8] Moresi LN, Solomatov VS (1995) Numerical investigation of 2D convection with extremely large viscosity
variations. Phys Fluids 7: 2154-2162
[9] Oxburgh ER, Turcotte DL (1968) Mid-Ocean Ridges and Geotherm Distribution during Mantle Convection. J Geophys Res 73: 2643-2661
[10] Pla F, Mancho AM, Herrero H (2009) Bifurcation phenomena in a convection problem with temperature dependent viscosity at low aspect ratio. Physica D 238: 572-580
[11] Torrance KE, Turcotte DL (1971) Thermal convection with large viscosity variations. J Fluid Mech 47: 113-125
[12] Trompert R, Hansen U (1998) Mantle convection simulations with rheologies that generate plate-like behaviour. Nature 395: 686-689
[13] Yanagawa TKB, Nakada M, Yuen DA (2004) A simplified mantle convection model for thermal conductivity stratification. Physics of the Earth and Planetary Interiors 146: 163-177

IMPROVING THE MAP OF SOIL HORIZONS BY INDICATOR KRIGING

V. Sidorova, P. Krasilnikov

Institute of Biology, Karelian Research Centre, RAS, Petrozavodsk, Russia,
sidorova@krc.karelia.ru+78142769810

The soil profile is a sequence of soil horizons. For classification purposes, at any level of taxonomy, soil surveyors usually first establish genetic horizons in the soil, and then make a conclusion about the taxonomic position of the soil. Laboratory analysis either confirms or disproves the field diagnostics, but soil division into horizons is the first step in soil mapping anyway. The depth of a particular soil horizon is often of major interest from a practical viewpoint. In many countries the depth of surface horizons is used as a diagnostic criterion in soil taxonomy at the species level, where classes such as “shallow”, “medium“, and “deep” are used to indicate the depth of the A or E horizons; the limits of these classes depend on the soil type. In other systems, the depth of horizons is used at the series level. Correct soil diagnostics at lower levels of taxonomy is an important task in soil survey.

The depth of a particular soil horizon varies significantly in space, and often does not depend directly on easily observed external factors. This is evident from the data of the study of long transects, especially where continuous trenches are studied. Traditional statistical methods are not always effective for managing spatially distributed data, where the neighboring samples cannot be regarded as independent one from another. Thus, geostatistical methods are recommended for the study of the spatial structure of soils, and for spatial interpolation of data. The objective of this study was to study the spatial variation of the presence and depth of soil horizons in the forest.

The study was made at the territory of the Karelia Republic, Northwest Russia. The study plot Gomselga was situated in the subzone of middle taiga. The study plot was established in hilly glacial and glaciofluvial landscapes with intensive tectonic discontinuities in pre-Cambrian crystalline rocks. The territory of the plot used to be covered with spruce forests, which later were almost completely clearcut, and the territory grew occupied by secondary forests about 50 years old, where the dominant species were birch, aspen with admixture of pine and spruce. The soils of the plot are diverse: iron-humus podzols (Haplic Podzols), podburs (Entic Podzols), raw-humus burozems (Dystric Cambisols), high-moor peat soils (Dystric Histosols), mud gley soils (Histic Gleysols), and sod-gley-podzolic soils (Dystric Planosols).

We studied spatial variability of various soil horizons, using the data of soil surveys of 16 km^2 plot. At each point we recorded the depth of O, A, E, and B horizons. We recorded data from more than 100 profiles. The spatial variability of the depth of soil horizons was estimated using variography. At first, experimental semivariograms were calculated, and

corresponding models were approximated using the least squares method. They quantified the distance (range), at which samples became uncorrelated from each other, and gave an idea of direction of the best and the worst spatial correlation. Then, using kriging and variograms as the basis, we generated maps of soil horizons depth.

Kriging is the good local estimators in some sense. But it always generates a smoothed model. Thus, the model variability is much less than it is in reality. The study plot was characterized by high soil diversity, and there were many points where the thickness of horizons was equal to 0 (the horizon was absent). It would be reasonable to assign zero thickness values at the neighborhood. But the kriged estimation of grid point is a weighted average of all nearby samples. So the other observed values showing thicknesses >0 will have an influence that may be significant.

The other problem is that the plot was characterized by the complexity of relief, the abundance of depressions occupied by lakes and peatlands, and the presence of bedrock outcrops, where the soil (and soil horizons) was absent. So it was impossible to predict the thickness of soil horizons there.

We used indicator kriging to delineate the areas of the presence of horizons. Additional "virtual" points with zero values located in the coordinates of lakes, rock outcrops, and other non-soil bodies were added to the data obtained from sampling points. Then we used an additional indicator variable:

$$I(x, \text{soil type}) = 1, \text{ if } z(x) = \text{soil type}$$
$$= 0, \text{ if } z(x) = \text{other soil type or non-soil body}$$

where z(x) was the observed categorical realization at location x. Then we generated probability maps of the presence of soil horizons. Precise estimation of the distribution of soil horizons using indicator kriging enabled delineation of the zones, where the probability of the presence of horizons was higher than the probability of their absence.

At the last stage we compared and combined two maps (one – the map of soil horizon thickness, another – the probability map of the presence of soil horizon). At every grid point, where a defined probability threshold (50% in our case) was not reached, the corresponding grid point containing the estimated thickness was set to zero.

So, smoothing effects occurring around zero thickness investigation sites can be reduced with the help of a probability criterion.

This study was supported by the Russian Foundation for Basic Research (project no. 09-04-00027).

SPATIAL DIFFERENTIATION OF ZINC AND COPPER CONCENTRATIONS IN THE SURFACE LAYERS OF FOREST SOILS IN THE AREA OF KARKONOSZE NATIONAL PARK, POLAND

K. Szopka, A. Karczewska, C. Kabała, P.Jezierski, A.Bogacz

Institute of Soil Sciences and Environmental Protection, Wrocław University of Environmental and Life Sciences, Wroclaw, Poland, katarzyna.szopka@up.wroc.pl, +4871/3205637

Introduction

Soil cover plays a crucial role in determining and governing the quality of forest ecosystems. Potentially harmful factors, such as enhanced concentrations of pollutants, may negatively affect forest productivity and biological balance. From among various soil pollutants, heavy metals are of particular importance, as they may accumulate in soil and remain there for a long time. Some heavy metals, if present in low concentrations, may act as micronutrients, but in the high amounts they are always toxic to biota. This is why metal concentrations in soils should be examined, in particular in the most susceptible ecosystems, such as mountain forest exposed to severe climatic conditions.

This study presents a differentiation of zinc and copper concentrations in soils in the area of Karkonosze National Park (The Giant Mts., SW Poland), which in the last decades of 20th century experienced a distressing processes of forest decline, obviously caused by many coinciding biotic and abiotic factors. Possible toxicity of heavy metals was listed as one of executive factors of forest damage. Therefore, a system of environmental monitoring, has been established in the Park. Soil monitoring is included as an integral part of environmental research. Soil examination within the system involves the analysis of basic soil properties, such as pH and organic matter content, and the concentrations of selected contaminants: sulphur and heavy metals, including zinc and copper [Karczewska et al. 2006].

Total concentrations and distribution of heavy metals in soils depend on natural factors: parent rock composition, biological accumulation and the effects of soil forming processes, as well as on several external factors, such as input of air-borne pollutants and changing chemistry of dry and wet precipitation [Kabała 1998].

The main aim of this study was to analyse spatial differentiation of copper and zinc concentrations in the surface layers of soils in the area of Giant Mts., as related to selected soil properties, topographic factors and other parameters.

Material and methods

The system of environmental monitoring in the Karkonosze National Park consists of 630 sites in a forested zone, and 230 sites in subalpine zone, arranged in a 200x300 m grid [Karczewska et al. 2006]. This study was carried out in the forested zone, where soil samples were collected from 476 of 630 monitoring sites. Soil material was sampled with split-tube samplers from the surface soil layers, i.e. from the depths 0-10 cm and 10-20 cm.

Additionally, the samples of organic layer (a forest floor) were collected if only the forest floor was present in particular site.

Soil samples were dried and homogenised prior to analyses. Laboratory analyses involved basic soil properties: grain size distribution, organic matter content (determined by oxidometric method) and soil reaction pH_{KCl} (determined electrometrically in 1mol/dm^3 KCl. Total concentrations of Cu and Zn were determined by AAS after microwave digestion of samples with "diverse" aqua regia (concentrated HNO_3 + HCl, 3+1).

All statistical calculations were made using a software Statistica 8. The maps of spatial distribution of Zn and Cu in soils were created with a software MapInfo. Further geostatistical analyses will be made with a software Surfer.

Results

Concentrations of Zn in forest floor ranged from 23 to 111 mg/kg (with the mean value of 46.7 mg/kg), whereas the concentrations of this element in the layers 0-10 cm and 10-20 cm varied in the ranges 10 – 124 mg/kg and 5.6-120 mg/kg, respectively (table 1). The mean values for those layers were lower than the mean in the forest floor: at the level of 33.1 and 26.9 mg/kg. Standard deviation of Zn concentrations were similar in all layers examined, i.e. 15.41 mg/kg for forest floor and 15.32 and 16.09 mg/kg for the layers 0-10 cm and 10-20 cm, respectively. Variability coefficient of Zn concentrations in forest floor (33.0%) was lower than that in the deeper soil layers (46.2% and 59.8 %).

Concentrations of Cu in soils ranged from 1.3 to 85 mg/kg, and were the highest in the forest floor, and decreased together with the depth. In the deeper layers, Cu concentrations remained below 53.5 mg/kg. The mean values of Cu concentrations in the sequence of layers: forest floor, 0-10 cm and 10-20 cm, were: 20.85, 13.79, and 8.15 mg/kg. The values of standard deviation showed the same trends as that for Zn, described above (table 1).

Table 1:. *Statistical parameters characterizing variability of Cu and Zn concentrations in soils.*

Parameter	Unit	Zn			Cu		
		Forest floor	0-10 cm	10-20 cm	Forest floor	0-10 cm	10-20 cm
Minimum	mg/kg	23	10	5,6	6	2	1,3
Maximum		111.5	124	120	85	53.5	48.5
Mean		46.74	33.2	26.94	20.85	13.79	8.15
Median		42.95	30	22.2	20	12.9	6.5
Standard deviation		15.42	15.32	16.09	7.78	7.33	5.86
Variability coefficient	%	33.0	46.2	59.8	37.3	53.1	71.8
No of samples	-	280	459	452	279	459	453

Table 2: *Correlation coefficients between the concentrations of Cu and Zn and site altitude and the parameters characterizing soil properties. The cases of significant correlations at p=0.05 are marked with asterisks.*

	Zn			Cu		
	Forest floor	0-10 cm	10-20 cm	Forest floor	0-10 cm	10-20 cm
Altitude, m asl	-0.029	-0.065	-0.120	0.060	0.169*	0.132*
Organic matter, %	0.047	0.218*	0.195*	0.057	0.299*	0.139*
Zn in forest floor, mg/kg	x	-	-	0.160*	-	-
Zn 0-10 cm	-	x	-	-	0.401*	-
Zn 0-10 cm	-	-	x	-	-	-0.006

Correlation analysis based on determination of Pearson correlation coefficient, indicated that the concentrations of Zn and Cu in the layers 0-10 cm and 10-20 cm were linearly dependent on the content of organic matter in particular soil layer (table 2). Such relationships were described by other authors who examined the distribution of metals in mountain soils [Donisa et al. 2000, Wang et al. 2009]. Positive correlation was also found between the concentrations of Zn and Cu in forest floor and their concentrations in the layer 0-10 cm (table 2).

Statistical analysis did not prove any simple correlations between the site altitude and total concentrations of copper and zinc in soils (table 2). High values of standard deviation, described above, confirm that there are some other factors, beside those examined previously, that determine the concentrations of Zn and Cu in the surface layers of forest soils in the area of the Karkonosze National Park. The example of spatial distribution of Cu in forest litter is shown in the map (figure 1). Further analysis will be needed to closer examine the spatial variability of Zn and Cu concentrations in soils and to identify the main factors governing the distribution of Zn and Cu in the soils examined. Geostatistical approach will be helpful in that analysis.

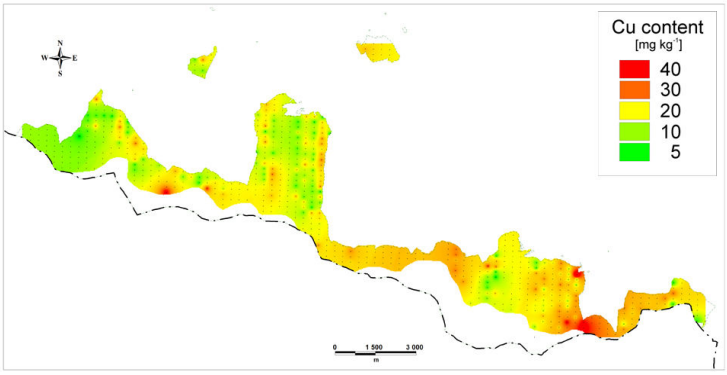


Figure. 1. *Spatial distribution of Cu concentrations in the forest litter in the Karkonosze NP*

References

Donisa C., Mocanu R., Steinnes E., Vasu A. (2000). Heavy metal pollution by atmospheric transport in natural soils from the northern part of Eastern Carpatians. Water, Air and Soil Pollution 120: 347-358

Kabała C. (1998). Trace elements in soils of the Izerskie Mts. ZN AR we Wrocławiu, Rolnictwo LXXIII, 347: 95-106 (in Polish)

Karczewska A., Szopka K., Kabała C., Bogacz A. (2006): Zinc and lead in forest soils of Karkonosze National Park. Polish J. Environ. Stud. 15, 2a: 336-342

Wang X., Cheng G., Zhong X., Li M. (2009). Trace elements in sub-alpine forest soils on the eastern edge of the Tibetan Plateau, China. Environ. Geol. (2009) 58: 635-643.

SPATIAL DEPENDENCE BETWEEN SOIL CARBON AND NITROGEN IN PURE BEECH FOREST STAND IN NORTH OF IRAN

G. H. Zahedi Amiri, S. Ahmadi

Department of Forestry, Faculty of Natural Resources, University of Tehran/ Iran
E-mail address: ghavamza@ut.ac.ir

Abstract:
This study was carried out in research forest at northern part of Iran. In the present study, a spatial estimation theory known as geostatistics was applied by variogram and Kriging interpolation for the analysis of soil carbon and nitrogen in a pure beech forest. Soil organic matter is a key component of any terrestrial ecosystem, and any variation in its amount and composition has important effects on many of the processes that occur within the system. The consequences of a doubling of the global atmospheric CO_2 concentration may be seen in different ranges of the spatial scale. Investigations of heterogeneity in natural systems have indicated substantial variability, even at scales of less than one meter. In fact, the values of a variable at sites that are close together are more similar than those further apart. In this study, 126 rectangular squire plots have been chosen systematically for soil analysis. The spatial analysis of the carbon percentage indicated a well defined structure of the variogram for organic and top-mineral soil depths. The soil nitrogen showed also a similar pattern as found for carbon in relation to the depth variable. The ordinary kriging interpolation showed that the carbon accumulation in top-mineral soil depth in beech forest was estimated to be 124.5 ton ha^{-1}, while in the organic layer was calculated to be only 0.54 ton ha^{-1}. An increase of the rate of nitrogen storage in the mineral layers in beech forest created a favorable soil conditions for rich vegetation species. The impact of tree species and soil characteristics was the most important factors influencing the spatial variability of carbon and nitrogen storage in the forest floor and mineral soil layers. This research indicated that the spatial dependence between carbon and nitrogen could be presented a sampling strategy for soil study in natural forests and silviculture closed to the nature for sustainable forest management.

Introduction:
Forest ecosystems represent an important carbon pool and have functioned as a net source of CO_2 in recent history (Burchel & Kursten, 1992). Soil organic matter is a key component of any terrestrial ecosystem (Evens et.,al, 2006). The loss of organic matter from soils following disturbance is an important source of CO2 for the atmosphere. The consequences of a doubling of the global atmospheric CO2 concentration may be seen in different ranges of the spatial scale (Gaten, et al., 2007). Investigations of heterogeneity in natural systems have indicated substantial variability, even at scales of less than one meter. In fact, the values of a variable at sites that are close together are more similar than those further apart. In the present study, a spatial estimation theory known as geostatistics was applied by variogram and

Kriging interpolation for the analysis of soil nitrogen and carbon in beech forest in north of Iran..

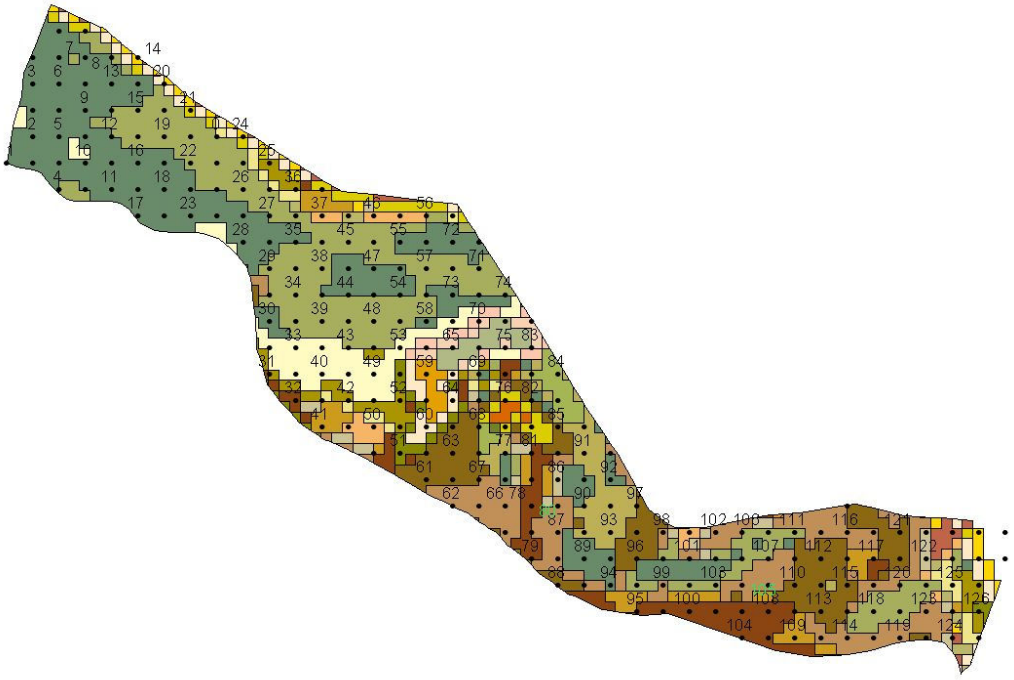


Figure 1: *Sampling plan of research area for soil carbon and nitrogen analysis*

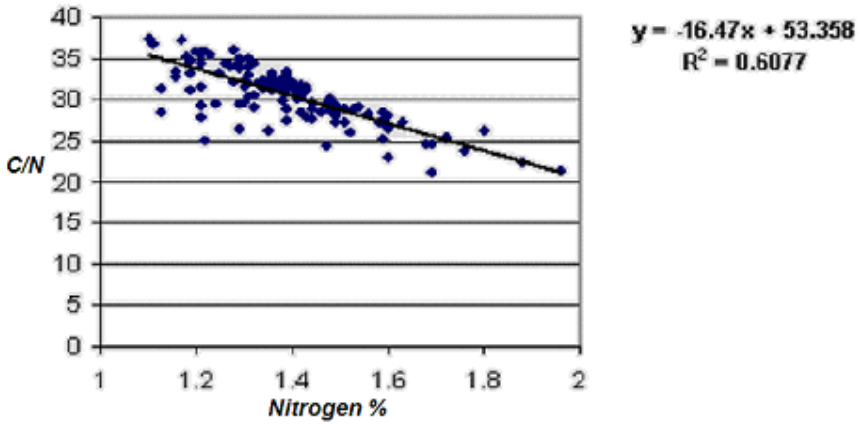


Figure 2: *Correlation between Nitrogen percentage and C/N (mineralization ratio) in organic layers (LFH) in beech forest*

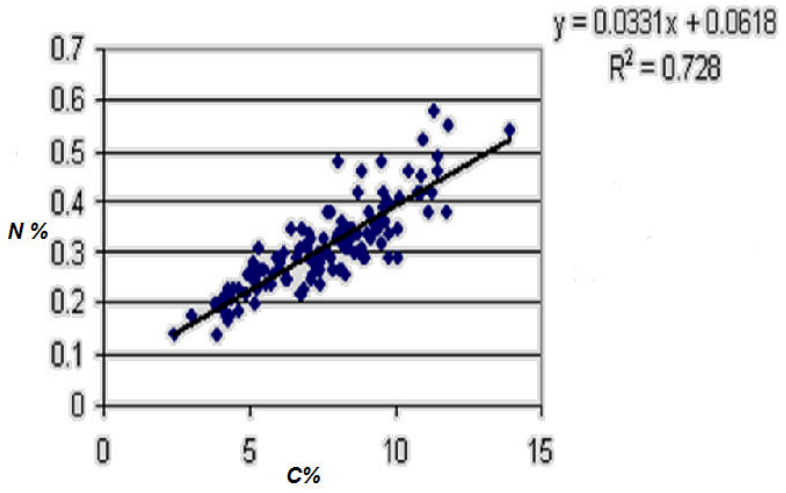


Figure 3: *Correlation between carbon and nitrogen percentage in top mineral soil depth in beech forest*

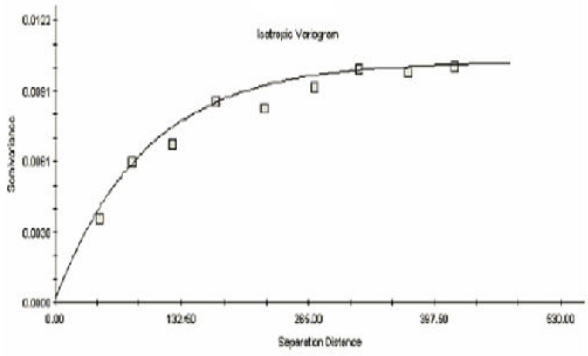


Figure 4: *Experimental variogram (ponts) and exponential fitted model (line) of the nitrogen percentage in the organic soil depth (LFH) at beech forest*