Pedometrics 2005
Frontiers in Pedometrics

Biannual Meeting of Commission 1.5 Pedometrics, Division 1
of the International Union of Soil Science (IUSS)

September 12-14, 2005

Naples Beach Hotel & Golf Club

Naples, Florida, USA

UNIVERSITY OF FLORIDA
IFAS

Project # 0503

Welcome to the Pedometrics 2005 - *Frontiers in Pedometrics* Meeting organized by the Soil and Water Science Department, Institute of Food and Agricultural Sciences (IFAS), University of Florida (UF), Gainesville, Florida. This is the biannual meeting of Commission 1.5 Pedometrics, Division 1 of the International Union of Soil Sciences.

Conference topics include digital soil mapping techniques, geostatistics, mixed quantitative modeling, soil and remote sensing to describe the distribution and genesis of soils. Other topics are focused on understanding spatial relationships between multiple soil and environmental properties, scaling, and soil sampling theory. A variety of spatial scales are covered ranging from field, hillslope to soil regions.

During this conference a total of 63 presentations - 3 keynote talks, 37 oral presentations and 23 poster presentations will be presented by scientists/researchers from the following countries: Australia, Belgium, Brazil, Canada, China, Croatia, Czech Republic, Denmark, France, Germany, Italy, Kenya, Netherlands, Russia, Spain, United Kingdom, and USA.

An optional pre-conference workshop "Quantitative Visible and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Characterization" (instructor: Dr. David Brown, Montana State University, Bozeman, MT) and a post-conference field trip to the Everglades (instructor and guide: Dr. Mark Clark, University of Florida/IFAS) are offered.

Best Pedometrics Paper (2003 & 2004) awards and a student award will be selected at the conference. The best presentations will be selected by the program committee and published in a Special Issue of Geoderma.

Special thanks are given to the scientific program committee members who reviewed submitted abstracts. The assistance of the University of Florida/IFAS, Office of Conferences and support staff, especially Ms. Sharon Borneman who handled most of the conference arrangements and registration, are also gratefully acknowledged.

S. Grunwald

Faculty Organizer,
Sabine Grunwald

Table of Contents

Program Committee

Gerard Heuvelink
ALTERRA and Laboratory of Soil Science and Geology, Wageningen University and Research Centre, the Netherlands

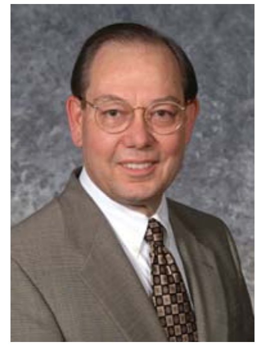
David Brown
Dept. of Land Resource & Environmental Sciences, Montana State University, USA

Pierre Goovaerts
PGeostat, LLC & Biomedware Inc., USA

Budiman Minasny
Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Australia

Sabine Grunwald
Soil and Water Science Department, University of Florida/IFAS, USA

Keynote Speakers



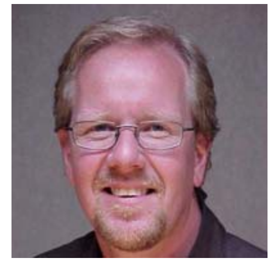
Jimmy Cheek has served as the University of Florida's Senior Vice President for Agriculture and Natural Resources since January 2005. In this role, Cheek administers the Institute of Food and Agricultural Sciences (IFAS) which includes the College of Agricultural and Life Sciences, the School of Natural Resources and the Environment, the School of Forest Resources and Conservation, elements of the College of Veterinary Medicine, the Florida Agricultural Experiment Station (with 13 research and education centers throughout Florida), and the Florida Cooperative Extension Service with offices in each of the state's 67 counties.

Dr. Cheek came to the University of Florida in 1975 as an assistant professor of Agricultural Education and Communication. He served for six years as Dean of the College of Agricultural and Life Sciences (CALS) from 1999-2004 and assistant dean of the college from 1992-99. His numerous awards include Teacher of the Year for CALS, Alpha Zeta Professor of the Year, and the University's first Faculty Superior Accomplishment Award. He is a Fellow of the American Association for Agricultural Education (AAAE) and the North American Association of Colleges and Teachers of Agriculture (NACTA). NACTA awarded him the Ensminger-Interstate Distinguished Teaching Award and named him the Distinguished Educator in 2005.

Dr. Cheek earned his B.S. degree and his Ph.D. from Texas A&M University. He received his master's degree from Lamar University. A native of Texas, he is married to Ileen and they have two adult children, Jennifer and Jeff.

Jay Bell is Professor in the Department of Soil, Water, and Climate at the University of Minnesota. He received his BS and MS Degrees in Agronomy from Virginia Polytechnic Institute and State University where he spent 4 years as a Research Associate conducting research on coal surface-mine reclamation in the Southern Appalachians. He received his Ph.D. in Agronomy from The Pennsylvania State University in 1990 and joined the faculty at the University of Minnesota in 1991. He teaches courses in Soil Genesis and Landscape Relations, Wetland Soils, Introduction to Environmental Science, Freshman Seminars, and conducts soil-geomorphology field tours. Dr. Bell is also the director of the Environmental Science undergraduate program which currently has about 90 students. His research program focuses on understanding soil genetic processes from both a quantitative and a landscape perspective and he as worked with U.S. National Cooperative Soil Survey. His current research projects are investigating soil hydrology, the genesis, morphology, and biogeochemical processes of transitional soils from uplands to wetlands, digital terrain analysis techniques, and quantitative modeling spatial patterns of soil properties. He has expertise with multi-scale databases for geographic information system applications with particular emphasis on the development, interpretation, and use of soils, topography, and land-use information. Dr. Bell is an editor-in-chief of Geoderma and has worked internationally in Morocco and Australia for extended periods of time. He was also chair of the International Soil Resource Assessment Conference held in Minneapolis, Minnesota in 1999.

Dr. J. C. Bell
Department of Soil, Water, and Climate
439 Borlaug Hall
University of Minnesota
St. Paul, MN 55108-6028
Email: jaybell@umn.edu
Phone: 612-625-6703



Harold M. van Es is a Professor of Soil and Water Management at Cornell University, holding a position in extension, research and teaching in the Department of Crop and Soil Sciences. He is a native of the Netherlands and received degrees from North Carolina State University (Ph.D.), Iowa State University (M.S.), and the University of Amsterdam (B.S.). He directs the Cornell Computational Agriculture Initiative, serves as Director of Graduate Studies, and teaches courses in Soil and Water Management and Applications of Space-Time Statistics. He published over 50 refereed scientific journal articles and co-authored a book titled *Building Soils for Better Crops.* Dr. van Es' research focuses on precision management of crop inputs (esp. nitrogen), soil health, and soil statistics. Related to the latter, he has focused on improved experimental design, variability structures in field studies, and the application of data mining methods.

Dr. Harold van Es
Department of Crop and Soil Sciences
Cornell University, Ithaca, NY 14853-1901, USA
Email: hmv1@cornell.edu

Marc Van Meirvenne graduated as an agricultural engineer at the Ghent University, Faculty of Bioscience Engineering, Belgium, where he obtained his PhD on the use of geostatistics in soil science in 1991. Since 1993 he lecturers in GIS and spatial statistics.

Since the foundation of the working group on Pedometrics, he became an active member. He participated in all Pedometrics conferences and most workshops. Between 1998 and 2002 he was chairman of the working group on Pedometrics and he was chairman of the local organization committee of the "Pedometrics 2001" conference held in Ghent. He was co-author of a paper which was awarded the "best paper in Pedometrics" prize in 1998.

His current research interests remain linked with Pedometrics. Topics include the application of Pedometrics in tropical environments, the use of soil sensors to provide auxiliary soil information and environmental applications of geostatistics (delineation and sampling of soil pollution).

Dr. Ir. Marc Van Meirvenne
Dept. Soil Management and Soil Care
Faculty of Bioscience Engineering
Ghent University
Coupure 653, 9000 Gent, Belgium
Phone: + 32 (0)9 264 6056
Fax: + 32 (0)9 264 6247
Email : Marc.Vanmeirvenne@ugent.be

Program Agenda

Abstract page numbers are indicated at the end of listings when applicable [example: "…(p. 2)"]

Friday & Saturday, September 9-10, 2005

Pre-Conference Workshop, University of Florida, Gainesville, Florida

Sunday, September 11, 2005

5:00pm-7:00pm	**Registration / poster set-up / welcome social**

Monday, September 12, 2005

7:30am-5:30pm	**Registration Open**
8:00am-8:15am	**Opening & Welcome – *Sabine Grunwald*,** Soil and Water Science Department, University of Florida/IFAS
8:15am-8:30am	**Welcome Address – *Jimmy Cheek***, Senior Vice President for Agriculture and Natural Resources, University of Florida/IFAS
8:30am-9:10am	**Keynote Talk: Pedometrics in Transition: From Too Few to Too Many Data? – *Marc van Meirvenne***, *Liesbet Cockx* and *Udayakantha Vitharana* .... (p. 81)

SESSION ONE: DIGITAL SOIL MAPPING I (MODERATOR: TOMISLAV HENGL)

9:10am-9:30am	**Optimization of Sample Locations for Universal Kriging of Environmental Variables – *Dick J. Brus***, *Gerard B. M. Heuvelink* and *J. J. de Gruijter* .... (p. 12)
9:30am-9:50am	**Wavelet Analysis of Categorical Soil Variables, Some Approaches Based on Indicators – *R. Murray Lark*** .... (p. 46)
9:50am-10:10am	**A Wavelet Variogram for Two-dimensional Data – *Edward H. Bosch***, *Antonio Paz González* and *José García Vivas* .... (p. 7)
10:10am-10:40am	**BREAK**
10:40am-11:00am	**Modeling Uncertain Categorical Soil Maps Using a Markov Random Field Approach – *Gerard B. M. Heuvelink*** and *Dick J. Brus* .... (p. 37)
11:00am-11:20am	**A Spatially-Explicit Mantel Test Framework to Investigate Relationships among Soil, Water, Landscape and Vegetative Properties – *Gregory L. Bruland***, *Sabine Grunwald*, *K. Ramesh Reddy*, *Todd Z. Osborne* and *Susan Newman* .... (p. 10)
11:20am-11:40am	**Application of Fuzzy Logic in Digital Groundwater Mapping – *Jaap J. de Gruijter***, *T. Hoogland*, *G. B. M. Heuvelink* .... (p. 19)
11:40am-12:00pm	**A Fuzzy Based Method for Spatial Modeling of Complex Soil-Landscape Relationships – *Ruprecht Herbst*,** *Jürgen Lamp* and *Felix Thiemeyer* .... (p. 32)

Monday, September 12, 2005 (continued)

12:00pm-1:50pm	**LUNCH ON OWN**

SESSION TWO: GEOSTATISTICS (MODERATOR: ALEX MCBRATNEY)

Tuesday, September 13, 2005

SESSION THREE: SOIL SENSING (MODERATOR: RON CORSTANJE)

Tuesday, September 13, 2005 (continued)

Tuesday, September 13, 2005 (continued)

Wednesday, September 14, 2005

Thursday, September 15, 2005

Poster Session Directory

Abstract page numbers are indicated at the end of listings [example: "…(p. 2)"]

Poster No.

Poster
No.

Abstracts

Listed alphabetically by presenting author and abstract title.
Presenting authors appear in **bold**.

Soil Attribute Prediction and Spatial Trends – A Comparative Study on Algorithms and Attributes

T. Behrens and *T. Scholten*
Institute of Geography, Department of Physical Geography and Soil Science, University of Jena, Germany

Knowledge Discovery in Databases or Data Mining has become an important and widely used approach for soil attribute prediction (McBratney et al. 2003). Anyhow, compared to interpolation techniques these approaches generally work local on tabular data, with no spatial component or relationship in the prediction approach. Hence a spatial trend is not directly considered, but known to be important for the distribution of almost any soil attribute and all scales.

To overcome this restriction and to incorporate the spatial domain we used two different approaches: On the one hand the spatial coordinates can be used directly by adding two columns - one for the x- and one the y-coordinate of each sample to derive a 'global' spatial trend. On the other hand, somehow related to interpolation, we computed the distance from each sample to all other samples. For the subsequent prediction distance grids were used in both cases. Hence, in addition to all other predictor variables, i.e. terrain attributes, land cover and parent material in our example, it is possible to incorporate the spatial domain in data mining approaches and thus within complex pedo-transfer functions to reveal the underlying spatial trend and derive improved predictions.

Based on this we compared different algorithms like Multivariate Adaptive Regression Splines (MARS), Regression Trees and Artificial Neural Networks. In addition, concerning sparse datasets, these data mining approaches are unstable predictors, which does not mean that they are weak predictors, but that small changes in the training dataset can cause significant changes in prediction accuracy. To overcome the stability problems we used bagging, an ensemble approach where multiple predictions based on bootstrap replicates of the training dataset are aggregated trying to realize stable predictions with high accuracies or in the words of Breiman (1996) "making a silk purse out of a sow's ear, especially if the sow's ear is twitchy".

All Data Mining approaches, with and without bagging, each with and without the different spatial components were compared and tested on different soil attributes in this study.

References:

Breiman, L., 1996. Bagging Predictors. Machine Learning 24(2), pp. 123-140.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, pp. 3-52.

Contact Information: Thorsten Behrens, Institute of Geography, Department of Physical Geography and Soil Science, University of Jena, Loebdergraben 32, D-07743 Jena, Germany. Phone: +49/(0)3641/948829, Fax: +49/(0)3641/948812, Email: thorsten.behrens@uni-jena.de

Dynamic Soil Mapping: Adding the Temporal Dimension

Jay Bell
Department of Soil, Water, and Climate; University of Minnesota; St. Paul. MN. USA

Soils are often represented as static entities in most soil resource assessments. In reality certain soil characteristics important for land-use management can change considerably through time in response to variations in climate. As such, the dynamic nature of soils is often ignored or minimized as we use traditional soil maps to represent spatial, but not temporal variations. We will demonstrate examples of dynamic soil-landscape modeling for recently glaciated landscapes. Spatial patterns of areas with similar hydrologic responses were defined using quantitative soil-landscape modeling techniques and digital terrain analysis. We have collected hydrologic monitoring data for the site over multiple years and used terrain attributes to extrapolate temporal changes in depth to soil saturation to the three-dimensional landscape. By using medical visualization software, we created animations of the annual dynamics of soil hydrology from both a surface perspective and by slicing the landscape along transects to visualize depth to saturation dynamics along landscape cross-sections. These animations can be run through time to visualize soil temporal and spatial variations simultaneously. By recognizing the dynamic nature of soils in mapping, we are able to provide valuable information for the use and management of soils as well as advance our understanding of processes of soil genesis. Temporal and spatial changes in soil chemical and biological processes related to soil properties such as carbon dynamics and gas fluxes are the next frontier in pedometrics research.

Contact Information: Jay Bell, Department of Soil, Water, and Climate, 439 Borlaug Hall, 1991 Upper Buford Circle, University of Minnesota, St. Paul. MN. USA, Phone: 612-625-6703, Email: jaybell@umn.edu

Forest Soil Acidification Assessment Using Principal Component Analysis and Geostatistics

***Lubos Boruvka**, Lenka Mladkova, Vit Penizek* and *Ondrej Drabek*
Department of Soil Science and Geology, Czech University of Agriculture, Prague, Czech Republic

Soil acidification and consequent Al release is a problem particularly under forest cover in mountainous areas of the Czech Republic. It is controlled by a number of factors, like acid deposition, forest type, parent rock, altitude, etc. The Jizera Mountains region presents an area heavily influenced by acidification and forest decline. Though the acid immissions decreased significantly in recent years, the ecosystem stays still threatened by acidificants accumulated in soil. In our previous paper, we studied the relationships between soil characteristics by means of multivariate statistics (Mladkova et al., 2004). Spatial distribution of individual soil characteristics in the forest floor and the influence of stand factors were shown in Boruvka et al. (2005). This paper focused on the effect of stand factors on spatial distribution of soil characteristics using a combination of principal component analysis and geostatistics.

A set of 98 sampling sites was described. Each site corresponds approximately to an area of 2 ha. Altitude, forest type, soil unit, and grass cover were recorded. Data on liming in the past were amended. Samples were collected from surface organic (O) and inner mineral (B) horizons. The analyses included active and exchangeable soil pH (pH_{H2O} and pH_{KCl}, respectively), total content of C, N, and S, pseudototal content of Ca and Mg (after aqua regia digestion), and the ratio of absorbances of soil sodium pyrophosphate extract at the wavelengths of 400 and 600 nm as the indication of humus quality (A_{400}/A_{600}). Moreover, concentrations of two Al forms were determined, namely exchangeable (in KCl extract; Al_{KCl}) and organically bound (in $Na_4P_2O_7$ extract; Al_{KCl} was deducted; Al_{org}). For details see Mladkova et al. (2004).

Principal component analysis was performed based on correlation matrix of soil characteristics, separately for the O and B horizons. The aim was to reduce the number of variables. In each case, four principal components (PC) were selected, explaining 84.6 and 81.8 % of total variability. They were rotated using varimax rotation. In the O horizon, the first PC (33.0 % of variability) showed important positive correlation (>0.5) with C, N, and S contents, and negative correlation with Mg content. The second PC (21.0 %) showed important correlation with both types of pH and with Al_{org}. The third PC (16.8 %) showed positive correlation with Ca content and negative correlation with Al_{KCl}. The fourth PC for the O horizon (13.8 %) showed positive correlation with Al_{org} and A_{400}/A_{600}. In case of B horizon, the first PC (27.6 %) showed also positive correlation with C, N, and S content. The second PC (24.1 %) showed positive correlation with both types of pH and negative correlation with Al_{KCl}. The third PC (15.38 %) showed correlation with Ca and Mg. The fourth PC for B horizon (14.8 %) showed positive correlation with Al_{org} and A_{400}/A_{600}.

Variograms were calculated for all PC score values. Altitude was described by variogram. Indicator variograms were used for forest type (0 - spruce forest, 1 – beech forest), soil unit (0 – Cambisols, 1 – Podzols), grass (*Calamagrostis villosa*) cover and liming. Crossvariograms between PC scores and stand factors were calculated. Spatial relationship with stand factors was assessed. Altitude influenced strongly spatial distribution of most PCs (except PC 3 for B horizons). Forest type influenced the scores of PC 1, 3, and 4 for the O horizon, and to a lesser extent of PC 1, 2 and 3 for the B horizon. PC 3 for the O horizon was influenced mainly by liming. PC 4 for the O horizon was affected by grass cover. Soil unit effects were more apparent for the B horizon than for the O horizon. The effect of parent rock was not studied because it is rather homogeneous in the area, formed mainly by granites. Kriged maps of PC scores were created to document the spatial distribution.

In summary, it was shown that surface horizons are more sensitive to external influence (acid deposition, liming, grass expansion) and their spatial variation is stronger. In the mineral horizons, the effect of pedogenetic processes is more important. The effect of stand factors on Al behaviour is complex and often indirect, mediated for example by organic matter or soil reaction. It is difficult to clearly distinguish the effects of the particular factors. Nevertheless, used combination of pedometrical methods provided a concise information about spatial variation and interrelation between soil characteristics and moreover about the effect of stand factors.

Acknowledgement: The study was supported by grants No. 1G57073 of the Ministry of Agriculture and No. 526/05/0613 of the Czech Science Foundation. The data were obtained in the frame of research project No. QC 1250 supported by the Czech Ministry of Agriculture.

References:

Mladkova, L., Boruvka, L., Drabek, O. (2004): Distribution of labile aluminium forms in soils of Jizera Mountains region. Plant, Soil and Environment 50 (8): 346-351.

Boruvka, L., Mladkova, L., Drabek, O., Vasat, R. (2005): Factors of spatial distribution of forest floor properties in the Jizera Mountains. Plant, Soil and Environment 51 (10): 447-455.

Contact Information: Lubos Boruvka, Department of Soil Science and Geology, Czech University of Agriculture in Prague, CZ-165 21 Prague 6 – Suchdol, Czech Republic, Phone: +420 224 382 751, Fax: +420 234 381 836, Email: Boruvka@af.czu.cz

A Wavelet Variogram for Two-Dimensional Data

Edward H. Bosch[1], *Antonio Paz González*[2] and *José García Vivas*[3]
[1]National Geospatial-Intelligence Agency, Reston, Virginia, USA
[2]Facultad de Ciencias, Universidad de La Coruña, Coruña, España
[3]Instituto de Física, Universidade Federal da Bahia, Salvador, Bahia, Brasil

Assuming the stationary conditions of a variable $z(k)$, the semivariogram given by $\hat{\gamma}(t) = \frac{1}{2(n-t)}\sum_{i=1}^{n-t}\left(z(i) - z(i+t)\right)^2$ in Geostatistics is used for the Kriging process. However, if the data z(i) are not stationary, the curve generated by $\gamma(t)$ can provide an indication of the non-stationarity of the variable $z(i)$ and Kriging in turn can not be performed.

Wavelets have been used for a variety of different applications including signal smoothing, compression, dimension reduction, interpolation, identifying points of discontinuity in a signal and for removing trend in a signal. The wavelet transform essentially splits a signal into a set of low and high frequency components via a pair of low and high frequency filters. The low frequency filter smoothes a signal while the high frequency filter provides information of a signal's high oscillating terms.

Bosch, Oliver and Webster presented in the paper titled, "Wavelets and the Generalization of the Variogram", Mathematical Geology, Vol. 36, No. 2, February 2004, the wavelet variogram as a tool that may be utilized to obtain variance information of a regularly spaced one-dimensional transect by means of comparing more than just pairs of samples. This was attained by means of Daubechies' orthogonal wavelet filters. It was noted in such paper that the graphs generated by the ordinary and wavelet variograms of random data are very similar. Furthermore, it was shown that the wavelet variogram effectively removed higher-order global trend (linear, quadratic, etc.) added to random data while its curve was very similar to that of the ordinary variogram of the residuals. The wavelet variogram also was shown in some cases to be very useful in determining the order of the trend of the transect in question. It also was shown that the wavelet variogram of order η (the number of vanishing moments) is related to the Generalized Variogram of order k.

In this work we extend the definition of the wavelet variogram of a one-dimensional transect to that of a wavelet variogram of a regularly spaced two-dimensional surface. The extension is defined as follows. Note that in the paper mentioned above, the ordinary semivariogram was expressed in terms of matrices by means of the following formula:

$\hat{\gamma}(t) = \frac{1}{(n-t)}\mathbf{z}^{\mathrm{T}}\mathbf{B}_{\mathrm{t}}^{\mathrm{T}}\mathbf{B}_{\mathrm{t}}\mathbf{z}$, where $\mathbf{z} = \left[z(1) \quad ... \quad z(n)\right]^{\mathrm{T}}$ is the transect, t is the lag parameter and $\mathbf{B}_{\mathrm{t}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \underbrace{0...0}_{t-1} & -1 & 0 & 0 & ... & 0 \\ 0 & 1 & \underbrace{0...0}_{t-1} & -1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 0 & 1 & \underbrace{0...0}_{t-1} & -1 \end{bmatrix}$. Note that for $1/(n\text{-}t)$ to be defined, it is required that n-t>0 which implies that the lag $t<n$. In fact, the largest value that t may obtain is n-1. However, it is usually not allowed to attain this value. Therefore, $\mathbf{B}_t$ is an element of the set of real matrices M with n-t rows and n columns. That is, $\mathbf{B}_t \in M(n-t,n)$.

Now, for a two-dimensional dataset $\mathbf{Z}_{(m,n)} \in M(m,n)$ on a regular grid, we can compute the semivariogram of $\mathbf{Z}_{(m,n)}$, <u>in the horizontal and vertical directions</u>, as follows: $\hat{\gamma}(t) = \frac{1}{(m-t)(n-t)}\sum_{i=1}^{m-t}\sum_{j=1}^{n-t}\mathbf{C}(i,j)^2$, where $\mathbf{C} \equiv \mathbf{B}_{(m\text{-}t,m)}\mathbf{Z}_{(m,n)}\left(\mathbf{B}_{(n\text{-}t,n)}\right)^{\mathrm{T}} \in M(m-t,n-t)$ and, $\mathbf{B}_{(m-t,m)} \in M(m-t,m)$ and $\mathbf{B}_{(n-t,n)} \in M(n-t,n)$ have the same structure as that of the matrix $\mathbf{B}_t$ above. Note that these matrices calculate the differences in the vertical and horizontal directions respectively and if m=n, then the matrix $\mathbf{B}_{(n\text{-}t,n)}$ is used for both directions. Finally, the sum of the square of the elements of the matrix $\mathbf{C}$, scaled by 1/[(m-t)(n-t)], generate the semivariogram for each value of the lag parameter t. The wavelet variogram is simply obtained by replacing the values $1/\sqrt{2}$ and $-1/\sqrt{2}$ (the discrete Haar wavelet) in the matrices $\mathbf{B}_{(m\text{-}t,m)}$ and $\mathbf{B}_{(n\text{-}t,n)}$ above with those of a wavelet filter and by changing the upper limits of the sums to that of a term that depends on the length of the filter, t, m and n. The wavelet variogram also may be obtained by means of the convolution of the data with a two dimensional filter (outer product of the wavelet filter elements). However, the former definition provides a more intuitive understanding of the results than that of the latter definition.

An interesting property of this technique is that since its implementation depends on a separable filter (vertical and horizontal directions), complicated polynomial trend can be removed from the data by a wavelet with a small number of vanishing moments. This is in sharp contrast to that of a wavelet in one dimension. Therefore, the wavelet variogram of such data can be very similar to that of the ordinary variogram of the residuals even when the only directions considered are the vertical and horizontal directions. Furthermore, the results are obtained very fast.

We show by means of several examples that the two-dimensional wavelet variogram of random data, with trend of different orders, is similar to that of the ordinary variogram of the residuals. We also have applied it to field measurements and have obtained very encouraging results. It is also possible in some cases to infer the data's order of the trend.

<u>Contact Information</u>: NGA 12310 Sunrise Valley Drive, DN11, Reston, Virginia, 20191-3449 USA, Phone: 703-735-3859, Fax: 703-735-3960, Email: edward.h.bosch@nga.mil

Mapping Soil Organic C, Texture and Clay Mineralogy in a Ugandan Dambo Using Digital Terrain Modeling and Proximal VNIR Diffuse Reflectance Spectroscopy

David J. Brown
Montana State University - Bozeman

There have been few detailed examinations of soil variability in dambos—seasonally saturated, channelless valley floors—which serve as important hydrologic, fertility and carbon reservoirs throughout East and Southern Africa. In this study, we augered a total of 193 profiles across all landscape positions in a 2nd order watershed, sampled at regular depth intervals to 2.7 m and classified 74 profiles as less-well-drained based upon spectroradiometer-determined Munsell® hue and chroma values at a depth of 50-60 cm. For these yellower and greyer profiles, soil organic C (SOC) and texture were determined for all surface samples (0-10 cm) with selected profile subsurface samples also analyzed. Using visible and near infrared (VNIR) spectroscopy, a 4183-sample soil-spectral library, digital terrain modeling (DTM), and empirical non-linear depth functions it was possible to predict SOC and clay % both spatially and vertically. Clay mineralogy was estimated using a 3372-sample VNIR soil-spectral library and boosted regression trees, with montmorillonite and kaolinite reference values in ordinal units of relative x-ray diffraction (XRD) peak intensity (0-5). Ordinal class probabilities were linearly interpolated and aggregated at fixed depth intervals to compute profile mineralogical indices which where then related to terrain parameters using spatial regression techniques. Combining digital terrain modeling for landscape characterization and VNIR spectroscopy for soil characterization supported mapping of within-dambo soil variability at a resolution not possible using traditional mapping and soil analysis techniques.

Contact Information: David J. Brown, Asst. Professor of Landscape Pedology, Montana State University, Dept. of Land Resources & Environmental Sciences, 334 Leon Johnson Hall, P.O. Box 173120, Bozeman, MT 59717-3120, Phone: 406-994-3724, Fax: 406-994-3933, Email: djbrown@montana.edu

A Spatially-Explicit Mantel Test Framework To Investigate Relationships Among Soil, Water, Landscape and Vegetative Properties

***G. L. Bruland*[1]**, *S. Grunwald*[1], *K.R. Reddy*[1], *Todd Z. Osborne*[1] and *Susan Newman*[2]
[1] Soil and Water Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA
[2] Everglades Division, South Florida Water Management District, West Palm Beach, FL, USA

Environmental data characteristically exhibit spatial autocorrelation and covariation among variables. This has rendered attempts to establish causality between predictor and response variables quite difficult as autocorrelation and covariation violate the assumptions of many traditional statistical analyses. The Mantel test offers a solution to this problem by explicitly accounting for spatial autocorrelation and covariation. While these tests have been used in various fields of life sciences, they have not been widely applied to pedometrics. In essence, the Mantel test is a regression in which the variables themselves are distance or dissimilarity matrices summarizing pairwise relationships among sample locations. The test evaluates the significance of the correlation between matrices of pairwise attribute dissimilarities by evaluation of the results from repeated randomization. If the correlations generated by randomization are as strong as the original data, then there is little or no correlation between the matrices, while if the correlations generated by randomization are weaker than the original data, then there is a strong correlation between the matrices. After each randomization, the Mantel Z statistic is calculated. The Z statistic from the original data is compared to the distribution of the Z statistics from the randomization for a significance test. The standardized Mantel statistic (r) is calculated as the Pearson correlation coefficient between the two matrices. This is accomplished by standardizing Z by the variances of the two matrices. Unlike Pearson correlation coefficients, however, Mantel r values do not have to be large in order to be statistically significant.

The flexibility and power of Mantel tests result from the various ways in which the correlations can be framed. As they use dissimilarity matrices, Mantel tests can be applied to different types of variables. By combining geographic coordinates into a distance matrix, Mantel tests can be used to evaluate the relationships between distances in geographic space versus distances in attribute space. Partial Mantel tests can also be conducted in which the relationship between matrices A and B can be assessed while controlling for the effect of a matrix C. Finally, pure-partial Mantel tests can be conducted in which the effects of matrices C, D, etc. can be removed, before testing the correlation between matrix A and matrix B. As the Mantel r is averaged over all distances, Mantel tests cannot discover changes in the pattern of correlations at different distances. The Mantel correlogram overcomes this problem by partitioning the correlation analysis into a series of discrete distance classes. The result of this analysis is a Mantel correlogram which is analogous to a semivariogram but performed on a dissimilarity matrix.

In this study we used the Mantel suite of analyses to analyze the spatial structure and relationships of geographic location, distance from water control structures (WCS), water depth, water chemistry, soil chemistry, and vegetative community composition from three different hydrologic units (HUs) of the Greater Everglades Ecosystem (GEE), Water Conservation area (WCA) 2A, WCA-3, and the Everglades National Park (ENP). One hundred eleven sites were sampled in WCA-2A, 363 sites were sampled in WCA-3, and 302 sites were sampled in the ENP. At each site, a suite of environmental variables were collected. First, water depth was recorded and a water sample was collected and analyzed for total phosphorus, total Kjeldahl nitrogen, and dissolved organic carbon. A sample of the detrital floc material as well as a sample of the upper 0-10 cm of the soil profile were collected and analyzed for 11 physical and chemical soil properties including bulk density, total phosphorus, and total nitrogen. The vegetative community was sampled by estimating the percent cover of species present in a small plot surrounding the sites. As the environmental variables were in different measurement units, the data was standardized into *z* scores. The vegetative data were arcsin-square root transformed and standardized by species maximum values. We used ArcGIS to calculate the distance from each sampling site to the nearest inflow WCS. The geographic and environmental data were then converted into distance and dissimilarity matrices. Simple Mantel tests were used to investigate correlations between the water, floc, soil, and vegetative community matrices with geographic coordinates and distance from WCS. Partial Mantel tests were used to derive the correlations between water, floc and soil variables and vegetation accounting for geographic coordinates and distance from WCS. Pure-partial Mantel correlations were calculated between environmental variables and vegetation accounting for all other variables. We also calculated Mantel correlograms for the water, floc, soil, and vegetative community from each of the three HUs to investigate changes in correlation structure at different spatial scales.

Our results indicated considerable differences in the Mantel correlations for the three HUs. While water depth had significant pure-partial Mantel correlations with vegetation in all three HUs, distance from WCS had significant pure-partial correlations with vegetation in WCA-2A and the ENP, but not in WCA-3. Total P in the surface water and floc had significant pure-partial correlation with vegetation in WCA-2A, while TP in the soil had significant pure-partial Mantel correlations with vegetation in WCA-3 and the ENP. There was a significant pure-partial relationship of vegetation with geographic coordinates in each HU, however this correlation was strongest in the ENP. This indicated that other unmeasured spatial factors such as hydrologic flow paths, fire, etc. were important drivers of ecosystem dynamics in the GEE. Mantel correlograms also indicated nonlinear spatial correlation structure for many of the environmental variables. Spatial correlations in WCA-2A generally occurred at distances < 10 km, whereas spatial correlations in WCA-3 and the ENP occurred at distances > 10 km. The suite of Mantel analyses show great promise for elucidating relationships between geographic distance, environmental data, and community composition at various spatial scales and across a variety of landscapes.

Contact Information: Greg Bruland, Soil and Water Science Department, IFAS, University of Florida, 2169 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA, Phone: 352-392-1951 ext. 210, Fax: 352-392-3902, Email: gbruland@ifas.ufl.edu

Optimization of Sample Locations for Universal Kriging of Environmental Variables

Dick J. Brus, *Gerard B.M. Heuvelink* and *J. J. de Gruijter*
Soil Science Centre, Wageningen University and Research Centre, Wageningen, the Netherlands

For large parts of the earth, the availability of full coverage information on the environment is growing steadily, think for instance of remote and proximal sensing images and digital elevation models. It is a big challenge to use the information contained in these images in mapping environmental variables about which we are sparsely informed. This can be done by modelling the relation between the variable to be mapped (the target variable) and the prior, auxiliary variables, for instance by (generalized) regression. To calibrate the regression model, observations on the target variable are needed. This calibrated model can then be used to obtain predictions of the target variable everywhere. In some situations these model predictions can be improved by interpolating the model residuals, and adding these interpolated residuals to the preliminary model predictions. In order to profit from this second step, the residuals must be spatially dependent, and the density of the observations (residuals) much be large enough. In this case, the observations are used twice, 1. to calibrate the model, and 2. to interpolate the residuals.

This paper deals with the situation where we want to collect (additional) observations on the variable to be mapped with the aid of the prior, auxiliary information. In general such observations are expensive due to labour-intensive fieldwork or costly laboratory analyses. Efficient use of the available resources for sampling is therefore of great importance. Once we have decided on the number of sample locations, the question raises where to observe the target variable. If the observations are used both for modelling and for interpolation, then this problem is not trivial, since the two uses impose conflicting requirements on the sample locations. Estimation of the relation between the target variable and the auxiliary information profits from a large spread of the observations in feature space, while spatial interpolation of the residuals gains from a uniform spreading of the observations in geographic space. As yet, it is unclear how these two requirements should be weighed and how the weighing depends on the characteristics of a particular case.

Little work has been done in this field. Hengl et al. (2003) predicted soil attributes from Digital Elevation Models and remote sensing images by regression-kriging, and proposed for this mapping method an `equal range' design. In this procedure the study region is stratified on the basis of the frequency distribution of the auxiliary variables. Stratification limits are set at equal distances in feature space. From each stratum an equal number of sample points is selected randomly, thus ensuring that the entire sample has a uniform spacing in feature space. Many samples are generated in this way, and the one with the best spatial coverage is retained.

In this paper a new method is presented and evaluated. In this method the spatial distribution of the target variable is described by the sum of a deterministic trend, modelled by a linear regression on the auxiliary variables, and a realization of a stochastic, spatially dependent residual, i.e., by the universal kriging model. Given the auxiliary variables and the variogram of the residuals, the prediction-error variance is a function of the sample locations and the associated values of the predictors only, i.e., the variance is independent of the target-variable values at the sample locations. In this case the problem boils down to searching for the sample locations with the minimum value for the spatially averaged or maximum prediction-error variance. The optimal sample locations are searched for by spatial simulated annealing (van Groenigen and Stein, 1998).

In a first test of the method, Heuvelink et al. (2005) optimized the locations of 4, 9 and 16 points in a square area with no trend, a trend that is linear in the x-coordinate and a trend that is quadratic in the x-coordinate. They found that taking trend estimation into account had a marked effect on the optimized sample locations, especially when spatial autocorrelation of the residual is weak.

In this study the method is tested to select sample locations for mapping groundwater dynamics with the method developed by Finke et al. (2004). The core of their method are multiple linear regression models
for groundwater dynamics at points with spatially exhaustive, auxiliary variables predictors derived from a DEM, topographic map and the existing water table class map.

References:

Finke, P.A., Brus, D.J., Bierkens, M.F.P., Hoogland, T., Knotters, M. and de Vries, F., 2004. Mapping groundwater dynamics using multiple sources of exhaustive high resolution data, Geoderma 123, p. 23-39.

Hengl, T., Rossiter, D.G. and Stein, A., 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Australian Journal of Soil Research 41, p. 1403-1422.

Heuvelink, G.B.A., Brus, D.J. and de Gruijter, J.J, 2005. Optimisation of sample configurations for digital soil mapping with universal kriging. Pedometrics Conference on Digital Soil Mapping, 2004, Montpellier, France.

Van Groenigen, J.W., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. Journal of Environmental Quality 27, 1078-1086.

Contact Information: Dick J. Brus, Soil Science Centre, Wageningen University and Research Centre, P.O. Box 47, 6700 AA, Wageningen, the Netherlands, Phone: +31-317-474237, Email: dick.brus@wur.nl

Digital Agroecosystem Mapping

Mathieu Chevalier[1], **Florence Carré**[1], *Raja Chakir*[2], *Caroline Godard*[2], *Christine Le Bas*[3] and *Pierre-Alain Jayet*[2]
[1]EnviCare, Paris, France
[2]UMR Economie Publique, INRA, Grignon, France
[3]Unité INFOSOL, INRA, Orleans, France

The context of the study is to support the European Commission in assessing quantitative and qualitative socio-economic and environmental impacts of decoupling of direct payments on agricultural production, markets and land use on EU. This project integrates soil sciences, agronomy and economics. One of the milestones of the assessment, which is the purpose of the study, is to map the agricultural productivity of Europe, existing at the European administrative level (scale of about 1:5 M). This study mainly concerns disaggregation.

Agricultural productivity depends on soil and terrain characteristics, climatic constraints, human behavior and management, and norms (Waltman et al., 1999). Then, two major factors can explain productivity: biophysical factors and socio-economic factors. Biophysical factors represent the potentials for agricultural productivity whereas economic factors are the constraints. We define agroecozones as geographic areas that share similar biophysical characteristics for crop production (Bailey, 1996; Omernik, 1993; Waltman et al., 1999; Liu and Samal, 2002) and agroecosystems as the integration of human factor to agroecozones.

Two steps are necessary to model agroecosystems in Europe: the first one consists in modeling potential crop production from soil fertility by a determistic approach. The second one is the agroecosystem mapping integrating human factor information, deriving from European Commission statistics, to the agroecozones.

In order to model and to map potential crop production we use:

- the European soil map at the scale of 1:1 M with associated pedotransfer rules;
- a phenological database of Europe at the resolution of 50 km for seven crops;
- the European landuse coverage at 250 m resolution;
- the digital elevation model of Europe at the resolution of 1 km

The STICS model (Brisson et al., 2002) is run considering landuse to predict crop yields from soil information, terrain attributes, climatic and phenologic data. As we focus on a potential yield, the management practices data are considered as optimum and standardized. The analysis is done at the finest resolution (250 m).

Thus, each pixel of 250 m resolution contains the estimation of yield for seven possible crops in Europe (cereals, pea, potato, sugar beet…) including meadows.

To map final crop allocation, we use the information of average structure and productivity of agricultural systems at the European level and the previous results. Each administrative unit is characterized by absolute surface of agricultural systems. The allocation of these production systems at 250 m resolution consists in optimizing the distribution of surfaces under constraints, at the admistrative level, and considering the connexity of crops between pixels. The optimization corresponds to a Bayesian maximum entropy approach (Christakos, 1990). Finally, each pixel contains potential crop yields and suitability for each agricultural system as a biophysical and conditional optimised probability.

Validation can be done with local statistics on crop production existing for the same time period (1997). In France, agricultural statistics exist at the resolution of 500 km² (cantonal level). Then, the 250 m results have to be aggregated at the cantonal level and the absolute crop production surfaces are compared.

The method used for mapping agroecosystems from digital data combines determistic and econometric approach. Multi-expert knowledge is required in order to define:

- *the final map object*. The relevant questions were: do we need to map agroecozones or agroecosystems? Which information do we have to work on? ;
- *the resolution of analysis*. The question was: do we need to work at largest or finest resolution?

Some problems about taxonomy and scale have to be solved concerning:

- landuse classes which don't have correspondence with crops;
- multi-scale data: some attributes are very detailed, the others are not. This can introduced biases in the two step approach.

Finally, this issue combining multi-expert knowledge with multi-scale data is a good example of final applications of pedometrics since it deals with functional soil mapping. Relevant problem associated to this application is mainly validation since it is difficult to find information existing at the same scale, with the same structure and at the same period.

References:

Bailey, R.G., 1996. Delineation of ecosystem regions. Environment Management 7, 363-373.

Brisson N., C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussière, Y.M. Cabidoche, P. Cellier, P. Debaeke, J.P. Gaudillère, F. Maraux, F. B. Seguin, H. Sinoquet, 2002. An overview of the crop model STICS. European Journal of Agronomy 18, 309-332

Christakos, G. 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. Mathematical Geology 22, 763-777

Liu, M., Samal, A., 2002. A fuzzy clustering approach to delineate agroecozones. Ecological Modeling 149, 215-228.

Omernik, J.M., 1993. Ecoregions : a spatial framework for environmental management. USEPA, Environmental Research Laboratory, Corvallis, OR.

Waltman, W.J., Mortensen, D.A., Cassman; K.G., Nelson, L.A., Specht, J.E., Sinclair, H.R., Waltman, S.W., Narumalani, S., Meachant, J.W., 1999. Agroecozones of Nebraska. Agronomy abstracts, Annual meetings, ASA-CSAA-SSSA, Salt lake City, UT.

Contact Information: Florence Carre, EnviCare, 71 avenue Victor Hugo, 75016 Paris, France
Phone: 0033661731962 ; Fax: 0033142781962, Email: Florence.Carre@envi-care.com

Using Digital Terrain Modeling For Estimation of Soil Properties

Annamaria Castrignanò[1], *Roberto Comolli*[2], *Nicola Lopez*[1], *Gabriele Buttafuoco*[3] and *Alessandra Castrignanò*[2]
[1]CRA-Istituto Sperimentale Agronomico, Bari, Italy
[2]Dipartimento di Scienze dell'Ambiente e del Territorio, Milano, Italy
[3]CNR-ISAFOM, Rende - CS, Italy

Assessment and prediction of the spatial distribution of soil properties are important elements for sustainable land management. A number of generic and robust spatial interpolation techniques have been proposed in the past. Geostatistics is generally preferred because it allows to take into account spatial correlation between observations to predict attribute values at unsampled locations. Digital modeling has become a set of quantitative methods to analyse the landsurface and the relationships between topography and soil properties. The early digital terrain model (DTM)-based predictions of soil properties were built upon simple statistical relationships between soil and topographic attributes at each point of landscape (Odeh et al., 1994). However, accurate prediction of a soil property at each point of landscape is difficult because of the high spatial variability of soil properties. In the last decade, a number of "hybrid" interpolation techniques, which combine kriging with exhaustive secondary information have been developed and tested. A multivariate extension of kriging, known as cokriging, has also been used for merging sparsely observations of the primary attribute with secondary attributes that are more densely sampled (Goovaerts, 2000; McBratney et al., 2000; Hengl et al., 2004).

The objective of this paper was to show as the prediction of soil properties can be improved if correlated secondary information, such as a digital elevation model (DEM), is used. pH and organic matter (OM) data were interpolated using two types of techniques: 1) method (ordinary kriging, OK) which uses only soil data; and 2) algorithms which combine soil data with a DTM (linear regression, simple kriging with varying local means, SKlm; kriging with external drift, KED, in the scope of IRF-*k* theory; and colocated ordinary cokriging, OCK). Prediction performances of the different algorithms were then compared using cross-validation.

The study site (1.5 ha) is located in northern Italy (Val Chiavenna), most of land is pasture at mean altitude of 2000 m. Soil samples were collected at 110 locations and the main physical and chemical properties were determined. Here, we'll refer only on pH measured in water and on organic matter (g kg^{-1}). An irregular DEM of the study site based on 467 points was constructed with a laser distance system linked with an electronic theodolite. The irregular DEM was converted into a regular one using lognormal kriging. We calculated digital models of: slope gradient (G, °), slope aspect (A, °), vertical curvature (k_v, m^{-1}), horizontal curvature (k_h, m^{-1}), mean curvature (H, m^{-1}) and accumulation curvature (K_a, m^{-2}) by the method of Evans (1980). However, only elevation and slope will be considered, because they are the only topographic parameters, which have a significant correlation with pH and OM. A straightforward approach of estimating soil property as a function of the colocated elevation and slope is using linear relation. OK amounts at estimating the unknown soil properties at the unsampled location as a linear combination of neighbouring observations of the only primary variable (Goovaerts, 1997). SKlm consists in replacing the known stationary mean with varying means derived from the secondary information using a linear regression. For KED in the scope of the IRF-k theory of non stationary statistics (Chilès and Delfiner, 1999), the basic hypothesis is that the expectation of the primary variable can be written as a linear combination of secondary variables. An alternative way of integrating secondary information in primary variable modeling is multi-colocated cokriging, where the contribution of secondary variable to kriging estimate relies on the cross-correlation between the two variables. The modified version uses the secondary variable at the target location and at all the locations where the primary variable is defined. The corresponding maps describe the distributions of the soil properties as affected by topographic parameters, however the main shortcoming of this type of regression is that the spatial correlation of the pH and OM observations is not taken into account. Variogram model used in OK included 3 basic structures for pH: nugget, spherical (range of 16.72 m), spherical (range 83.21 m); and one basic structure for OM: spherical (range 24.49 m). The corresponding maps look different from the previous ones, because they do not take into account the secondary information deriving from topography. Unlike the OK, SKlm replace the known stationary mean with varying means derived from the linear regression estimated above. Compared with the OK maps the impact of topography is quite evident. To apply KED in the scope of IRF-*k* theory, we determined the optimal drift, including an intercept and elevation for both primary variables, and the best combination of basic structures for generalised covariance resulting in a nugget effect for both primary variables. The resulting maps for pH and OM, using only elevation as external drift, look quite variable. The main difference between the maps obtained through regression and the ones estimated using KED is that in the latter case the relationship between soil property and elevation is locally estimated, which causes more erratic estimates. In OCK a linear model of coregionalization was fitted to the 6 variograms (both direct and cross) of the 3 variables, pH, OM and elevation, using the same set of basic models (here a nugget effect, a spherical model (range 21.77 m) and a Bessel-k model (scale 21.77 m, parameter 2). Unlike the three previous techniques, the details of the elevation map do not appear so evident in the cokriging map, which looks more similar to the OK map. The performances of the kriging interpolators were evaluated using cross-validation and for the comparison criterion mean and variance of the standardised error were used. As the variance of the residuals (0.74 for pH and 2773.84 for OM) was calculated for the linear regression model fitted using all 110 observations, the prediction error would tend to be underestimated for this method, which was then excluded from the comparison. The results show that the largest predictions errors were generally obtained with OK and SKlm, whereas KED and OCK performed at the best for pH and OM, respectively.

References:

Chilès J.P., Delfiner P., 1999. Geostatistics: Modeling Spatial Uncertainty, Wiley & Sons, N.Y.

Evans, I.S., 1980. An integrated system of terrain analysis and slope mapping. Zeitschrift für Geomorphologie Supplementband 36, 274-295.

Goovaerts P., 1997. Geostatistics for Natural Resources Evaluation. OUP, 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, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75-93.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97, 293-327.

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.

Contact Information: Annamaria Castrignano, CRA - Istituto Sperimentale Agronomico, Via C. Ulpiani, 5 Bari, Italy 70125, Phone: 3.90805E+11, Fax: 3.90805E+11, Email: annamaria.castrignano@tin.it

To Detect the Breakdown of Assumptions of Statistical Stationarity in the Soil Variation of a Complex Landscape

Ron Corstanje[1], *R. Murray Lark*[1] and *Sabine Grunwald*[2]
[1]Environmetrics Group, Biomathematics & Biostatistics, Rothamsted Research, Harpenden, Hertfordshire, UK
[2]GIS Laboratory, Soil & Water Science Department, University of Florida, Fl, USA

Stationarity refers to a random function and is not the property of data. It is a property of the random field that we assume is realized in our data. Nonetheless, it is possible to test the plausibility of the assumption for a given dataset as has been done with wavelet analysis. The essential idea is to obtain a model for the random function under the assumption of intrinsic stationarity. If the global dataset can be regarded as a realization of a Gaussian process (perhaps after transformation) then the global variogram is sufficient for this purpose. By using locally estimated variograms, we then define some statistic of homogeneity of the spatial variation of the data. This allows us to obtain a sample distribution for this statistic, under the null hypothesis of intrinsic stationarity, by generating multiple realizations of the postulated random function at the original sample points and recomputing the statistic for each realization.

A considerable sampling effort in the Everglades has generated a significant dataset at 1,328 sites for a large area in southern Florida encompassing the Everglades National Park, Big Cypress National Preserve and the Water Conservation Areas 1, 2, and 3 (courtesy of Dr. K. Ramesh Reddy, Wetlands Biogeochemistry Lab, SWS, UF and Dr. Susan Newman, South Florida Water Management District). The entire extent of this area is 8,248 km^2. The soil properties measured included soil nutrient content (total nitrogen, total phosphorus and total inorganic phosphorus, total carbon), soil metals (Fe, Al, Mg, Ca) and basic soil physical properties such as bulk density and ash content. Previous geostatistical analysis on this dataset has reflected the hydrological units and boundaries within the system and a significant amount of work has been accomplished in understanding and refining predictions within these units. The size and coverage of this dataset makes it suitable for a detailed study to test to what degree the assumption of stationarity, specifically second order stationarity, actually holds over such a landscape.

We estimated the global variogram of pre-selected soil properties. We then estimated local variograms from the data by using a moving window and made an initial evaluation of the homogeneity of these local variograms from selected variogram characteristics. Next, we assessed how much variation in these statistics is consistent with sampling fluctuations when the data actually are intrinsically stationary by using Monte Carlo methods. We specified the global variogram, and generated an unconditional stochastic simulation of a Gaussian random field with this variogram at all sample sites. The same estimation algorithm was then used to estimate the local variogram parameters within a moving window, allowing us to estimate the standard deviation of these local variogram parameters. We repeated this process with multiple independent realizations of the same random function, generating empirical sampling distributions of the desired statistics, constructing an approximate test of the deviation from homogeneity. These tests will allow us to evaluate to what degree the assumption of stationarity actually holds over the Everglades landscape

Contact Information: Ron Corstanje, Environmetrics Group, Bioinformatics and Biomathematics Division, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, United Kingdom, Phone: + 44 (0)1582 763133 Ext 2402, Fax: Fax: + 44 (0)1582 760981, Email: ron.corstanje@bbsrc.ac.uk

Application of Fuzzy Logic in Digital Groundwater Mapping

J. J. de Gruijter*, T. Hoogland* and *G. B. M. Heuvelink*
Environmental Sciences Group, Wageningen UR, the Netherlands

A central issue in digital mapping is how to use ancillary data and expert knowledge for spatial prediction of the variable of interest. Linear regression models have often been used for this purpose, but the underlying assumption of linearity is usually not realistic, possibly causing an undesirable amount of model-bias. To avoid this, several options for non-linear modeling are available, e.g., generalized linear models, non-linear regression models, regression trees and neural networks. Such models are purely empirical in the sense that, once their structure is chosen, they are determined by the data. As a consequence, they may represent the available data well, however, they have limited possibilities for employing expert knowledge: they are *black-box* models. A well-known shortcoming of such models is that they are of little use for `what-if' analyses. The classical way to accommodate expert knowledge is by theoretical process models: *white-box* models describing a process explicitly by (differential) equations. However, models of this type are often non-existent for the variable of interest, or they demand input data that are not available. A remedy might be provided by application of fuzzy logic, a well-established methodology founded on fuzzy sets. Fuzzy logic represents a compromise between the two extremes mentioned above, leading to so-called *`grey-box'* models.

To explore this idea out, we have developed a spatial prediction method based on fuzzy logic, the core of which is a set of fuzzy rules. These fuzzy rules can be formulated through knowledge elicitation from experts, but they can also be derived by fuzzy cluster-analysis (k-means) of the data. The latter leads to a fuzzy partition of the space spanned by the ancillary variables, and each fuzzy set of this partition represents the antecedent of a fuzzy rule. The consequent of each rule is a predicted value for the variable of interest. The final prediction is then the average of the predictions from the individual rules, weighted with the memberships in the antecedent sets. (The result is a Tagagi-Sugeno singleton fuzzy model.)

The proposed method is applied in a case-study of groundwater mapping, and evaluated by comparing predicted values with values measured on an independent random sample of test points. Special attention is given to transformation of the attribute space prior to fuzzy cluster-analysis, the number of rules, fine-tuning of the model, introduction of expert knowledge, and generalization to a Multiple-Input-Multiple-Output model.

References:

Hellendoorn, H. and D. Driankov, 1997. *Fuzzy Model Identification: Selected Approaches.* Springer-Verlag, Berlin, 1997.

McBratney, A.B. and J.J. de Gruijter, 1992. *A continuum approach to soil classification by modified fuzzy k-means with extragrades*. J. Soil Sc., 43: 159-175.

Contact Information: J. J. de Gruijter, Alterra, P.O.Box 47, 6700 AA Wageningen, the Netherlands, Phone: +31.317.474238, Email: jaap.degruijter@wur.nl

Estimating Soil Carbon Stocks at the Field Level Using GIS and Geostatistics

L. Delisle[1], *R. Yost*[1], *P.C.S. Traore*[2], *M. Doumbia*[3], *A. Ballo*[3] and *K. Traore*[3]
[1]Dept. of Tropical Plant and Soil Sciences, University of Hawai'i at Manoa, Honolulu, Hawai'i
[2]ICRISAT, Bamako, Mali
[3]Institut d'Economie Rurale, Bamako, Mali

We have applied a geospatial approach to the problem of estimating soil carbon tonnages, the associated uncertainty of estimates, and changes in carbon over time, for several small, irregularly-shaped farm regions in Mali, West Africa. The farms are participating in a project, Soil Management Collaborative Research Program (SM-CRSP), to improve soil fertility and productivity through implementation of ridge tillage, a management practice designed to conserve soil moisture and reduce soil erosion by reducing run-off from rainfall. Increases in carbon, critically important in these low-carbon soils of the greater Sahel region, are expected to occur as a result of increases in soil moisture and reduction of soil erosion. Samples were collected on an irregularly-spaced sampling scheme at several time increments over the years 2000-2004 and analyzed for total soil organic carbon. Between 16 to 75 samples were collected at each of two soil depths (0-20 and 20-40 cm) at each time increment. Sampling positions and the boundaries of the fields were recorded with a DGPS. The areas of farm regions were later calculated automatically within GIS software (ESRI, ArcGIS 9) and soil volume calculations were based on these planar surface areas. Global estimates of carbon were based on the mean values of the carbon surface maps predicted by ordinary kriging. Estimates of the precision of estimates are currently being developed by application of the sequential simulation method. Several secondary variables, including elevation and individual bands of Quickbird remotely-sensed multispectral imagery, have been included in the analysis through cokriging and co-simulation. A major challenge has been the modeling of carbon variograms for sparse sample datasets. Similarities observed in experimental variogram shapes across farms, depths, and time, as well as across secondary variables has allowed variogram model parameters to be inferred for the smaller datasets. Predicted organic carbon values range from about 0.2 to 0.6 % for the surface depth (0-20 cm) in a typical farm in the village of Sougoumba (about 9-10 tonnes carbon per hectare). Increase in carbon in the 0-20 cm depth for one farm over the years 2000-2004 was about 12% and was spatially variable. Elevation data, measured only at sampling locations, appears so far not to improve the estimates (as judged from cross-validation statistics). Information from remotely-sensed imagery appears not strongly-correlated with carbon but may improve estimation after stratification.

<u>Contact Information</u>: Laura Delisle, Dept. of Tropical Plant and Soil Sciences, University of Hawai'i at Manoa, 3190 Maile Way, Honolulu, Hawaii, USA, 96822, Phone: 808-521-7091, Email: delisle@hawaii.edu

Spectral Reflectance Measurements for Predicting Soil Organic Carbon Content at Regional Scale

Y. Fouad[1]*, H. Aïchi*[1]*, C. Walter*[1] and *R. A. Viscarra-Rossel*[2]

[1]Agrocampus Rennes-UMR SAS, Rennes, France

[2]Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Ressources, University of Sydney, NSW, Australia

Soil organic carbon (SOC) content represents a key parameter in evaluating the quality of soils (Krishnan *et al.*, 1980; Ben Dor *et al.*, 1995) and is variable in both space and time (Mathieu *et al.*, 1998; Batchily *et al.*, 2003). This parameter has practical significance, inasmuch as information on organic matter content is critical to the evaluation of soil productivity and should be included as a factor in determining the levels of *N* fertilisation and herbicide application (Henderson *et al.*, 1992). Knowledge of the within-field spatial variability of SOC content is of major benefit, particularly as regards the context of soils in Western France (the Armoricain Massif landform). SOC variability is considerable at the regional scale, from 1% to 6% (Walter *et al.*, 1997) yet also appears to be significant at the local scale due to both land use differences and pedological variations. Moreover, the reallocation of land over recent decades has served to increase the level of within-field soil variability as small fields have been merged together. The soil reflectance measurement is a useful tool for characterising this variability. The literature does in fact indicate that soil parameters like OC content, mineralogy, soil moisture, texture and surface roughness all influence the reflection from soils (Bédidi *et al.*, 1992; Dalal and Henry, 1986; Houssa et al., 1996). A strong correlation between soil colour and OC content has been widely documented (see for example Viscarra et al., 2003), with darker soils generally containing more organic carbon than lighter ones. Numerous laboratory studies have measured soil reflectance, in an attempt to quantify this relationship. Moreover, the predictive quality of these models diminishes as their range of application gets extended to geographical areas not incorporated during their calibration stage. It thus seems necessary to develop a set of regional models specific to each pedo-climatic context (Ingleby and Crowe, 2001).

The great interest of soil spectroscopy is that one can simultaneously assess a number of soil properties (OC, Fe, CEC, pH,…). However, in this present work we only deal with SOC predictions from spectral measurements. And the mean goal was to show the methodology we carried out to calibrate a prediction model in Brittany pedo-climatic context. Of course, once calibrated the model may then be used for field-scale predictions of SOC content in Brittany. Thus, the particular objective of the present work, based on laboratory measurements, has been to develop, in Brittany context, a prediction model of soil organic carbon content based on reflectance within the visible and near-infrared spectral ranges.

We constituted a spectral database by collecting 64 A-horizon soil samples (0-20 cm) from various sites in Brittany's Armorican Massif, in order to cover the spatial variability of the regional range of soil types. The samples were then analysed for OC, Fe, N, CEC, pH and texture using conventional laboratory analytical techniques (AFNOR, 1996). We thus performed a Principal Component Analysis (PCA) in order both to explore the samples' chemical and physical characteristic variability and to group samples into separate classes. Afterwards, the samples were separated into two distinct data sets: one for calibration and the other for validation. Spectral measurements, ranging from 400 to 950nm, were conducted on soil samples that have been oven-dried, grounded and sieved at 400µm. A Partial Least Squares regression algorithm for a single variable (PLSR1) was then used to predict soil carbon content from reflectance of overall spectral range. The model calibrated by the mean of PLS regression predicts correctly organic carbon content with R^2 = 0.91 and RMSE = 0.36% values in cross validation. Model validation led to less precise, yet still satisfactory, predictions with R^2= 0.83 and RMSE = 0.46%. The model proved to be valid and robust over the range 0.92%- 5.17% of organic carbon content but, it cannot be extrapolated to predict OC content beyond this range for which it was calibrated; in this case, predictions proved to be poor. In order to derive a more exhaustive model, the soil database would need to be expanded to make it as representative as possible of the heterogeneity of soils within the target region.

References:

AFNOR. 1996 – Qualité des sols. Recueil de normes françaises.

Batchily, A.K., Post, D.F. and Breckenfeld, D.J. 2003. Spectral reflectance and soil morphology characteristics of Santa Rita experimental range soils. USDA Forest Service Proceedings RMRS-P 30 175-182.

Bédidi, A., Cervelle, B., Madeira, J. and Pouget, M. 1992. Moisture effects on visible spectral characteristics of lateritic soils. Soil Science Society of America Journal 153 129-141

Ben-Dor, E. and Banin, A. 1995. Near-Infrred analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal 59 364-372.

Dalal, R.C. and Henry, R.J. 1986. Simultaneous determination of moisture, organic carbon and total nitrogen by near infra red reflectance spectrophotometry. Soil Science Society of America Journal 50 120-123.

Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Scott, D.E. and Coster, D.C. 1992. High dimensional reflectance analysis of soil organic matter. Soil Science Society Am. J. 56 865-872.

Houssa, R., Pion, J.C., Yéssou, H. 1996. Effects of granulometric and mineralogical composition on spectral reflectance of soils in a sahelian area. ISPRS Journal of Photogrammetry and Remote Sensing 51 284-298.

Ingleby, H.R. and Crowe, T.G. 2001. Neural network models for predicting organic matter content in Saskatchewan soils. Canadian Biosystems Engineering 43 7.1-7.5.

Krishnan, P., Alexander, J.D., Butler, B.J. and Hummel, J.W. 1980. Reflectance technique for predicting soil organic matter. Soil Science Society of America Journal 44 1282-1285.

Mathieu, R., Pouget, M., Cervelle, B. and Escadafal, R. 1998. Relationships between satellite-based radiometric Indices simulated using laboratory reflectance data and typic Soil Color of an Arid Environment. Remote sensing of environment 66 17-28.

Viscarra Rossel, R.A., Walter, C. and Fouad, Y. 2003. Assessement of two reflectance techniques for the quantification of the within field spatial variability of soil organic carbon. In : ECPA 2003, the 4th European Conference on Precision Agriculture, J. Stafford and A. Werner (Eds), ATB and ZALF, Germany. pp. 697-703.

Walter, C., Schvartz, C., Claudot, B., Bouedo, T. and Aurousseau, P. 1997 Synthèse nationale des analyses de terre réalisées entre 1990 et 1994 (National synthesis of soil analysis between 1990 and 1994). Etude et Gestion des Sols 4 205–220.

Contact Information: Youssef Fouad, Agrocampus Rennes, UMR INRA/Agrocampus Sol-Agronomie-Spatialisation, 65 rue de Saint Brieuc – CS 84215, 35042 Rennes Cedex, France,
Phone: 33 (0) 2 23 48 55 56, Fax: 33 (0) 2 23 48 56 99, Email: fouad@agrocampus-rennes.fr

Remote Access Soil Proxy (RASP) Modeling Technique for Wilderness Areas

Bruce E. Frazier[1], *Toby Rodgers*[2], *Crystal Briggs*[2] and *Alan Busacca*[1]
[1]Crop and Soil Sciences Dept., Washington State University, Pullman, WA, USA
[2]USDA-NRCS Soil Survey, Mount Vernon, WA, USA

Wilderness areas in Washington State have historically been excluded from soil resource inventories due to the huge investment of time and resources required to map them. Yet, they are ideal places to study soil formation and distribution under natural conditions and the data are important for planning decisions. Computer-based models are an efficient way to predict occurrence of soils. Models lessen the need for intensive field transecting and offer a cost-efficient alternative to traditional cartographic techniques. We have used ArcGIS software to develop the RASP modeling process. Soil pedon data are collected from dominant landscape facets that are accessible by trails. Soil formation is modeled using surrogates for the soil-forming factors.

Two research projects were conducted involving Washington State University (WSU), United States Department of Agriculture Natural Resources Conservation Service (USDA NRCS), Forest Service (USFS), and National Park Service (NPS). These projects explored the feasibility of a GIS-based soil distribution model. The first project (Rodgers, 2000) with assistance from the USFS was to map the Pasayten and Sawtooth Wilderness areas (~ 600,000 acres) of the Okanogan-Wenatchee National Forest, Washington, at a level of detail consistent with an Order IV survey (Soil Survey Staff, 1993). Taxonomic classifications of soils in this Order IV survey were kept to the subgroup or great group level of Soil Taxonomy (Soil Survey Staff, 2003). Ultimately, the product from this research was incorporated into a correlated soil survey for the Okanogan-Wenatchee National Forest. The second project (Briggs, 2004) with assistance from NPS was to map Thunder Creek watershed (~ 74,000 acres) in North Cascades National Park, Washington. Taxonomic classifications of soils in this Order IV survey were also kept to the subgroup or great group level of Soil Taxonomy.

Mapping wilderness areas with the RASP model allows for execution and capture of repeatable logic, calculations, and delineations that can be improved as more information about the soils and soil-landscape patterns is acquired. It also remedies a long standing criticism of traditional surveys where information is lost due to the difficulty of transferring the soil surveyor's tacit knowledge into the soil survey (Hudson, 1992). With the RASP model, all mapping steps are documented and the rationale behind each step of model development is supported by soil-landscape patterns determined through fieldwork, laboratory data, and frequency distributions.

Soil-landscape associations can be modeled with a GIS using accessory information traditionally used in soil surveys and now available in digital form including vegetative patterns, geomorphology, digital elevation models (DEM), and DEM derived attributes such as slope gradient and wetness index. The RASP model has utilized a 30 m DEM and can be adapted to take advantage of higher resolution DEM's and other digital data if available. Adaptability of the RASP model is a key point that is difficult to overemphasize. A single iteration of the model is not necessarily the final version of the model. As a survey crew moves into a new area, the model can be adjusted to mirror changes in the local conditions that affect soil-landscape relationships. Surveyors can also fine-tune the procedural details of an existing model as soil-landscape relationships in an area are better understood.

To create a soil distribution map using the RASP model, it is necessary to understand the pedogenic processes and the soil-forming factors that control them. We found landscape stability, parent material, and vegetation to be important factors of soil formation. Analysis of soil occurrence with various combinations of soil-forming factors allowed us to document the geographic distribution of each soil type. We selected digital data layers to serve as proxy indicators for the soil-forming factors and we used critical thresholds and combinations of these proxies to assign map unit complexes to a final soil map. A landform map was used to proxy parent material and landform stability; a vegetation layer created from remotely sensed data was used to proxy for vegetation and climate. Primary and secondary terrain attributes calculated from a DEM were incorporated into the model to represent slope, aspect and soil wetness.

As a brief explanation of how the RASP model operates, we can examine the constructs behind the map unit complex of Typic Udivitrands and Andic Haplorthods. At locations where Typic Udivitrands and Andic Haplorthods are mapped, the RASP model has extracted a combination of the digital proxies that satisfy the following criteria: the landform is a valley wall, the soil moisture regime is udic, the soil temperature regime is frigid, the overstory is coniferous, and the wetness index value is less than 11. Each cell in the output raster meeting the above criteria is then assigned a map unit number corresponding to the Typic Udivitrands-Andic Haplorthods complex. The rationale for the various combinations of proxies that result in a final output raster of soil distribution is based on field observations of the soil-forming factors and the resultant soils.

References:

Briggs, C.A.D. 2004. GIS-based mapping of soil distribution in Thunder Creek watershed, North Cascades National Park, Washington. MS thesis. Washington State Univ., Pullman, WA. 164 pp.

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56:836-841.

Rodgers, T.M. 2000. Modeling soils of the Sawtooth and Pasayten wilderness areas with a GIS. MS thesis. Washington State Univ., Pullman, WA. 110 pp. www.wsu.edu/~trodgers/TobyThesisPDF.pdf

Soil Survey Staff. 1993. Soil Survey Manual. USDA-NRCS, Agricultural Handbook no. 18. U.S. Govt. Print. Office. Washington, DC. 503 pp.

Soil Survey Staff. 2003. Keys to soil taxonomy. 9th ed. USDA-NRCS. U.S. Gov. Print. Office, Washington, DC. 332pp.

Contact Information: B.E. Frazier, Crop and Soil Sciences, PO Box 646420, Washington State University, Pullman, WA 99164-6420, USA, Phone: 509-335-2381, Fax: 509-335-8674, Email: bfrazier@wsu.edu

Exploring the Spatial Non-stationarity of Relationships among Soil Properties Using Geographically-Weighted Regression

P. Goovaerts
BioMedware Inc., Ann Arbor, MI, USA

Analysis of the correlation between soil attributes is traditionally performed using global or "aspatial" regression, with the implicit assumption that the relationship between variables is constant across the study area. This assumption is likely unrealistic for large areas that display substantial geographic variation in landscape, geology and land uses. In such cases, the geographic variation in the set of variables is too complex to be captured by a single set of correlation or regression coefficients. Ignoring the non-stationarity of relationships in space might lead one to conclude that there is no correlation between soil attributes, which affects description of patterns of correlation and prediction using multivariate regression techniques, including residual kriging.

One straightforward way to account for the non-stationarity of relationships in space would be to conduct the regression within specific regions, but it requires an *a priori* and subjective delineation of these regions. Geographically weighted regression (Fotheringham et al., 2002) avoids this problem by performing the regression within local windows centered on target locations, which may correspond to grid nodes or sampled locations. Within each window, the observations are weighted according to their proximity to the centre of the window. Local regression coefficients and associated statistics (i.e. proportion of variance explained, correlation coefficients) can then be mapped to visualize how the explanatory power of secondary variables changes spatially. The technique, which bears similarities with the geostatistical approach of kriging with an external drift, has mainly been used in social science and in climatology, for example to study the local relationship between altitude and rainfall.

The aim of this paper is to describe the technique of geographically weighted regression and discuss its implementation both for description and prediction purposes. Analysis of several datasets illustrates the ability of the local regression to reveal significant correlations that can even change sign across the region depending, for example, on the presence of different soil types. Blending these local correlations in a global or aspatial approach often leads one to conclude that these soil attributes are non-correlated. Maps of local correlation coefficients are thus very useful to highlight regions of interest where the local pattern of correlation departs from the global one. Application of geographically weighted regression to spatial components estimated by kriging analysis also allows one to explore the concept of non-stationary scale-dependent correlation patterns.

Current implementations of geographically weighted regression ignore the spatial correlation of residuals. Empirical studies show that the autocorrelation of regression residuals is stronger for a global model than local ones, which suggests that part of the residual spatial correlation is caused by the application of a global model to non-stationary processes. In geographically weighted regression, this non-stationary is accounted directly through the regression coefficients instead of being left out in the residuals. Yet, incorporating the residual correlation might improve the prediction, and this issue is investigated in an empirical study.

References:

Fotheringham, A. S. et al, Geographically weighted regression: the analysis of spatially varying relationships, Chichester, Wiley, 2002.

Contact Information: Pierre Goovaerts, Biomedware, Inc. 516 North State Street, Ann Arbor, MI 48104, USA, Phone: 734-913-1098 ext. 8, Fax: 734-913-2201, Email: goovaerts@biomedware.com

Independent Validation of Soil Predictions - The Act of Testing the Truth?

S. Grunwald[1], *G.L. Bruland[1]* and *P. Goovaerts[2]* and *S. A. Bloom[1]*
[1]Soil and Water Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA
[2]BioMedware Inc., Ann Arbor, MI, USA

Validation of models that predict physical, chemical, biological and morphological soil properties is essential to assess the reliability and accuracy of predictions from such models. It is routinely employed in many pedometric applications to assess the quality of predictions and to evaluate the performance of different prediction methods. Validation, also known as jackknife in the geostatistical literature, has been hailed superior to cross-validation because it uses an independent dataset to verify predictions. However, unbiased assessment of model performance is confounded by the fact that exhaustive observations of soils are rarely available and that the validation results might be an artifact of the validation method, in particular the choice of a validation set, rather than a representation of reality.

A standard validation procedure entails subdividing observations into model and validation subsets using a random number generator that provides an independent selection mechanism. Validation projects in soil science face the following constraints: (i) Soil sampling is labor intensive and costly; often the number of observations within a domain provide insufficient data for model validation; (ii) Random selection of validation samples might not be representative of the variability within the study domain, in particular, if datasets are small; (iii) Random selection of validation samples might be biased towards specific geographic regions resulting in clustered/clumped validation subsets; (iv) An unique split into model and validation datasets ignores the impact of sampling fluctuations on prediction performances; and (v) Splitting observations into model and validation datasets based on one vs. multiple properties might provide different results.

Our objective was to quantify how the selection of a validation dataset influences the results of the validation exercise. We compared the impact of choosing independent validation schemes versus targeted validation schemes.

We used a soil dataset collected in the Everglades National Park and Big Cypress National Preserve in south Florida that comprised 399 observations representing a suite of physico-chemical properties including total phosphorus (TP) at the 0-10 cm depth. This soil-landscape domain is a naturally oligotrophic wetland system that is phosphorus limited. A vegetation map was used to delineate strata representing different ecological zones. Different schemes were considered to subdivide the observations into model and validation subsets: (i) random selection; (ii) stratified random selection using vegetation classes as strata; (iii) Latin-Hypercube selection based on factor scores derived from a principal component analysis of the multivariate dataset.

To account for the influence of the size of the validation set (V) versus model set (M), the following nine different combinations were considered: (i) M 90% - V 10%; (ii) M 85% - V 15%; (iii) M 80% - V 20%; (iv) M 75% - V 25%; (v) M 70 % - V 30%; (vi) M 65% - V 35%; (vii) M 60% - V 40%; (viii) M 55% - V 45%; and (ix) M 50% - V 50%.

For each validation scheme and M-V split combination, one hundred random subsets were selected, the semivariogram was computed and modeled automatically, and then TP was predicted using ordinary kriging. Numerous error metrics were used to evaluate model performance, including the mean prediction error, root mean square prediction error, and G coefficient.

Results showed large variations among validation schemes, indicating that the selection of a validation method impacts model performance. Large fluctuations were also found among the 100 realizations of any given validation scheme, which emphasizes the lack of reliability of common procedures where the validation is performed only once. Targeted validation methods better reproduced the soil variability across the study domain than independent validation methods. This effect becomes stronger as the size of the selected validation subset increases. These findings have implications for the assessment of model performance. For example, excellent or poor model performance might be an artifact of the selected validation method rather than the actual representation of the accuracy and reliability of a predictive soil model. This study indicates a need for the development of robust alternative strategies for validating predictive models.

Acknowledgement: Data were collected and analyzed by the South Florida Water Management District (SFWMD) and the Wetland Biogeochemistry Laboratory, Soil and Water Science Department, University of Florida (K.R. Reddy, S. Newman, T.Z. Osborne, and Y. Wang).

Contact Information: Sabine Grunwald, Soil and Water Science Department, IFAS, University of Florida, 2169 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA, Phone: 352-392-1951 ext. 204, Fax: 352-392-3902, Email: SGrunwald@ifas.ufl.edu

Analysis of Root Zone Soil Moisture using Optical Satellite Imagery

Sung-ho Hong[1], *J. M. H. Hendrickx*[1], ***J. B. B. Harrison***[1] and *B. Borchers*[2]
[1]Dept. Earth & Environ. Sci., New Mexico Tech, Socorro, NM
[2]Dept. Mathematics, New Mexico Tech, Socorro, NM

The status of root zone soil moisture plays a crucial role in the water exchange processes between the land surface and the atmosphere. Since root zone soil moisture is a dynamic variable subject to rapid changes in time and space, accurate and detailed information on the distribution of soil moisture is difficult to obtain. Ground based methods for the measurement of spatial and temporal changes in root zone soil moisture require much time and effort. Existing remote sensing methods use microwaves to measure soil moisture near the soil surface (0-10 cm). In this study, the Surface Energy Balance Algorithm for Land (SEBAL) is applied to a series of LandSat TM optical images covering the Sevilleta National Wildlife Refuge (SNWR) in New Mexico to determine the regional distribution of the evaporative fraction. From this, soil moisture conditions are derived using an empirical relationship between evaporative fraction and root zone soil moisture. The first objective of this study is to investigate how optical satellite imagery can be used for the mapping of regional root zone soil moisture conditions. The second objective is to examine the effect of different types of soil, vegetation and amount/distribution of precipitation on soil moisture. In addition to field validation of soil moisture product, the relative importance of soil and vegetation properties will be evaluated in the context of point and spatial distribution of soil moisture. Finally, the limitations of the model and possible extensions will be addressed.

Contact Information: Bruce Harrison, Dept. Earth & Environ. Sci., New Mexico Tech, 801 Leroy Place, Socorro, NM 87801, USA, Phone: (505)835-5864, Email: bruce@nmt.edu

Methods to Interpolate Soil-Classes from Profile Observations: Lessons from Iran

Tomislav Hengl[1], *Norair Toomanian*[2], *Ahmad Jalalian*[2] and *Hossein Khademi*[2]
[1]AGIS centre, Faculty of Agriculture, University of J.J. Strossamyer, Osijek, Croatia
[2]Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, Iran

The paper gives a systematic comparison of the state-of-the-art spatial interpolation techniques that can be used to interpolate soil-classes from point i.e. profile observations. Five distinct approaches are considered: (A) pure classification of auxiliary maps (maximum likelihood method); (B) continuous classification followed by interpolation of multiple memberships; (C) interpolation using multinomial logistic regression; (D) classification based on heuristic rules (terrain segmentation) and (E) classification of interpolated diagnostic soil properties following the reduced classification keys.

In the case of pure classification method (A), the classification is straight forward: a user needs to build up a list of continuous raster maps and then use supervised classification of feature space by using the profile observations. This is equivalent to classification of remote sensing images for land cover classification. Examples of this approach are given by Dobos et al. (2000) and Hengl and Rossiter (2004). A disadvantage of this method is that it reflects the relationship between the auxiliary predictors and soil-classes in a naïve way, which can then lead to artificial patterns in the final prediction map.

In the case of continuous classification followed by interpolation of memberships method (B) each membership is mapped separately so that they can be used to inspect confusion between different classes and derive the most probable class at each grid node. This method was first time suggested by De Gruijter et al. (1997). Here no auxiliary predictors are needed, which also means that the natural patterns of relief and parent material are ignored.

Memberships can also be produced by correlating the soil-classes with auxiliary maps using the multinomial logistic regression (method C). The advantage of this technique is that the both continuous and discrete auxiliary predictors can be used, so that the final prediction maps will very much follow the pattern of the relief. An example how to map soil-types using multinomial logistic regression on terrain parameters is given by Bailey et al. (2003).

As alternative to pure statistical techniques, McMillan et al. (2003) uses heuristic rules and following large soil survey experience to set up classification rules (method D). In their approach, soil-class maps can be built directly by segmenting digital elevation model of a study area. Although this technique is not computationally demanding it relies heavily on the interpreters' knowledge and can lead to serious over and under-estimations.

The last method (E) to map soil-classes relies on cheap (often descriptive) soil diagnostic properties and high quality auxiliary maps. In the case of the most recent FAO classification (WRB), a simplified classification rules can be setup for a study area. This gives a limited number of diagnostic properties (e.g. gleying properties, occurrence of calcic horizon, soil color etc), which are then interpolated using some powerful interpolation technique such as regression-kriging. Finally, each pixel can be classified using the classification rules exactly as in the WRB classification keys. This method is a hybrid of the previously listed techniques and is the most adjusted to the classification system used.

Listed methods have been compared using the multiscale soil profile data from Iran consisting of: (a) the National database with 4250 soil profiles and (b) regional scale study in the Zayandeh-rud valley consisting of 199 profile observations. SRTM digital elevation model, mean annual MODIS NDVI imagery and geological maps have been used as auxiliary predictors. Several aspects of each interpolation technique have been evaluated: spatial accuracy of method, accessibility of software-tools; amount of data processing needed and sensitivity of a technique to the quality of input data. Key advantages and limitations of each technique have been finally emphasized and some pragmatic tips given.

References:

Bailey, N., Clements, T., Lee, J.T. and Thompson, S., 2003. Modeling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. Journal of Environmental Management, 67(4): 395-407.

Carré, F. and Girard, M.C., 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma, 110(3-4): 241-263.

De Gruijter, J.J., Walvoort, D.J.J. and van Gaans, P.F.M., 1997. Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma, 77(2-4): 169-195.

Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L. and Helt, T., 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma, 97(3-4): 367-391.

Hengl, T. and Rossiter, D.G., 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Sci. Soc. Am. J., 67(5): 1810-1822.

Hengl, T., 2003. Pedometric mapping: bridging the gaps between conventional and pedometric approaches. PhD thesis, University of Wageningen, Enschede, 214 pp.

MacMillan, R.A., Martin, T.C., Earle, T.J. and McNabb, D.H., 2003. Automated analysis and classification of landforms using high-resolution digital elevation data: applications and issues. Can. J. Remote Sensing, 29(5): 592-606.

Contact Information: T. Hengl, AGIS centre, Faculty of Agriculture in Osijek, Trg Sv. Trojstva 3, 31000 Osijek, Croatia, Phone: 385-31-224288, Fax: 385-31-207018, Email: hengl@pfos.hr

A Fuzzy Based Method for Spatial Modeling of Complex Soil-Landscape Relationships

***Ruprecht Herbst*[1], *Jürgen Lamp*[2] and *Felix Thiemeyer*[3]
[1]Department of Precision Agriculture, Humboldt Universität zu Berlin, Berlin, Germany
[2]Department of Soil Informatics, C.A. Universität zu Kiel, Schleswig-Holstein, Germany
[3]Farmer, Batzlower Dorfstraße, 15377 Märkische Höhe, Brandenburg

Introduction

The main problem of a broad acceptance in precision farming in practice is to acquire, assess and manage heterogeneity in the field. This is the basis for variable rate application with the idea to reduce factor costs as well as to optimize yield potential. For the detection of spatial heterogeneity point observations (soil- and nutrient probing) and spatial pre-information can be used, which can be derived from sensor data (for example yield maps, electrical conductivity, multispectral data) and further data sources (digital elevation model, soil maps, remote sensing). In this research we focused on the decision making process. Very often the capturing of heterogeneity is a cost intensive procedure. The modeling process we used in this research is a fuzzy based analysing method to assess the soil/landscape continuum by multi criteria evaluation (MCE). A combination of this qualitative-quantitative approach has the advantage that prediction models can be locally adapted in different soil scapes and under various data quality situations.

Investigation Area and Methods

In 2004 we started with the implementation of Precision Farming Technology on a 500 ha crop farm in East Germany. The main soil texture on the training field with a size of 77 ha is glacial sand, partly up to 2m thick and partly loamy glacial residues in the subsoil. Main soil types are Colluvia and Luvisols. The relief is characterized by an average slope length of 250 m. The precipitation is 500 mm/year. A yield monitoring system was installed and yield maps were compiled for almost 400 ha.

The decision making process to predict complex functions in their spatial expendability is very important for the operative usage of variable rate technology. In this research we used a fuzzy modeling process to describe the soil continuum based on local soil-factor relationships (Ameskamp, 1997, Lagacherie & Voltz, 2000; Mc Bratney et. al., 2000). The fuzzy logic was developed by Zadeh 1965 and instead of strictly distinguishing between objects that belong to a given set and those that do not, fuzzy sets allow grades of membership expressed by real numbers between 0 and 1. During the second step in the multi criteria evaluation a weighting strategy to assess the importance of the factors is included. The weighting process is used to develop a set of relative weights for a group of factors in a multi-criteria evaluation. The procedure by which the weights are produced follows the logic developed by T. Saaty (Saaty, 1977). The factors and their resulting weights are used as input for weighted linear combination.

Results

This research focused on the soil born yield potential as an example. A yield map from 2004, a digital elevation model, a map of the electrical conductivity and the available water capacity were used as model inputs. With one exception, the mean yield for which we took a linear membership function, we analysed by a sigmoidal membership functions. For this landscape

with sandy soils and a lack of water during the vegetation period, especially the water balance influencing the yield potential. Therefore, the rules for the evaluation process are determined by deductive estimation of evapotranspiration. For example, north orientated slopes are much better weighted for the yield than the other areas. To estimate the soil born yield potential, the fuzzyficated input maps were used and the relative weight of each factor to the others were determined by an evaluation matrix based on the analytical hierarchical process.
This uncertain modeling process describes the continuum of soils much better than crisp methods. Further on, we can use constraints which means, that areas inside of fields can be excluded in the modeling process. This might be for example bog areas which are characterized by special dynamics and have to be individually interpreted for management decisions.

Conclusion

Strong heterogeneity and independent distribution of soil and nutrient parameters (Herbst & Lamp, 2004) make the need for application of decision support systems necessary. The multi criteria evaluation described above can be used for predicting single parameters as well as complex functions in their spatial expendability. This modeling process can be individually influenced by the membership functions and in the analytical hierarchical process. In addition the modeling can take during the growing season the change of the natural conditions into account, and this can be considered in the evaluation process as well. In 2004 and 2005 we used this modeling process successful for variable fertilisation strategies on the Thiemeyer farm.

References:

Ameskamp, M. 1997. Three-Dimensional Rule-Based Continuous Soil Modeling. Dissertation Kiel, Institut für Informatik und praktische Mathematik, Number 9701

Herbst, R. & Lamp, J. 2004. Strategies from Point to Spatial Information to Determine the Soil Continuum in Precision Agriculture, In: Proceedings of the 4th international conference on Precision Agriculture, Minneapolis, in press

Lagacherie, P. & Voltz, M. 2000. Predicting soil properties over a region using sampling information from mapped reference area and digital elevation data. Geoderma, Vol. 97, Issue 3, p. 187-208

McBratney, A. B.; Odeh, I.; Bishop, T.; Dunbar, M.; Shatar, T. 2000. An overview of pedometric techniques for use in soil survey. Geoderma Vol. 97, p. 293-327

Saaty, T.L. 1977. A Scaling Method for Priorities in Hierarchical Structures. J. Math. Psychology, 15, p. 234-281

Zadeh, L. 1965. Fuzzy Sets. Information and control. 8, p. 338-353

Contact Information: Ruprecht Herbst, Department of Precision Agriculture, Humboldt Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany, Phone: 0049/30/2093-6402, Fax: 0049/30/2093-6294,
Email: r.herbst@agrar.hu-berlin.de

Integrating Geoelectrical Sensor Data for Detailed Surveys Of Soil Bodies

J. Lamp[1], *M. Graeber*[2] and ***R. Herbst***[3]
[1]Pflanzenern.&Bodenk., C.A.Universitaet, Kiel, Germany
[2]GeoServe Ltd., Luetzowstr. 3 , Kiel, Germany
[3]Precision Agriculture, Humboldt Universitaet, Berlin, Germany

Introduction

Precision Farming (PF) and competitive land uses, esp. for drinking water supply, ask for very detailed, but cost-effective surveys which predict soil parameters at different depths for several variable rate (VRT) or conventional farm applications. Remote or proximal sensing of spectral re- and emissions in many electromagnetic wavelength bands may detect and map soil surface parameter, efficiently (Viscarra Rossel & McBratney 1998, Corg-mapping on-the-go: Reimer 2003). But the ease of a non-destructive access to soil, subsoil and deeper zones diminishes with depth, rapidly. Geophysicians use the ellipsoidal streamlines of electrical currents which are injected from the surface by steel electrodes into soil bodies. Pairs of voltage electrodes at varying distances (from <1dm to >100m) along transects measure potentials and – resolving Ohm's law - specific electrical conductivities (EC in mS/m). The current is conducted preferably on moist, ion-loaded clay surfaces of soils, therefore the EC values are highly correlated to an most important soil feature, the clay content. But other factors may also influence the signals (esp. soil moisture and temperature, see Durlesser 1999). Efficient devices working on soil surfaces are available which use either indirectly electromagnetic induction (Geonics EM38) and capacity methods or directly spiked electrode wheels (Veris, ARP-Geocarta). All kinds can be used to collect soil indicator values at high time and areal rates very efficiently (eg. EM38: ~2 s, ~50 values/ha, >100 ha/day) along the tramlines of cultivated land (often 18-24m width). While the EM38 measures relative values of bulked apparent ECa's from one soil range 0 to ~2m only, the direct devices are able to sense absolute potentials and conductivies from several soil depth ranges (Dabas&Tabbagh 2003).

Study objects and methods

Mixed soil samples in plastic boxes were used in the laboratory to measure EC values by electrodes in two configurations (Wenner, mini-SoilRover) and to estimate correlations to varying clay contents (1 – 80 %), moisture states (dry to wet) and texture strata (eg. sand on loam or v.v.). These results en-miniature assisted the interpretation of field results from two geo-electrical sensors, but also "SoilRover" augerings in different north-german fields were used to calibrate sensor signals as follows.

<u>EM38, an inductive soil EC sensor:</u> A sending coil induces currents by a 14.6 kHz electromagnetic wave into soils and their bulked conductivities are measured by a secondary wave. The Geonic's device is mounted on a PVC shield 5m behind the all terrain vehicle (ATV "SoilRover", equipped with dGPS and laptop data logger). Durlesser investigated backgrounds, Herbst (2002) and Reimer (2003) showed some "pros and cons" for PF applications (correlations to mapped grain yields; handling of noises like moisture). Due to high performance, this technique is increasingly asked for by farmers to produce digital soil texture maps.

<u>"Pluripol", a direct pluri-depth EC sensor:</u> The Geoserve company constructs the EC measure box, assisted to configure the "Pluripol" carrier of electrodes and pre-evaluated the multiple EC

values. Two current electrodes and – at varying distances from 40cm to >4m apart – up to five voltage electrode pairs can be mounted on the Pluripol. Again, the EC values are georeferenced by dGPS and stored on the bord laptop. By data inversion the integrative EC signals are transferred to absolute estimates of usually four intergrading soil zones (A: ~0-30; B: ~30-90; C: ~90-150; D: ~150–>250cm). Database and GIS programs help to manage, interpolate and visualize EC maps. Though depth information is much better resolved by this EC technique, the interpretation of results is often multivalent and must be calibrated.

"SoilRover", an ATV for 1.5-3m soil core augering: A hydraulic auger mounted on a Land-rover Defender 110 draws 1.5m soil cores (standard, extendable to 3m), automatically. The dGPS-geoferenced soil data are stored and managed by an Access program "Solum" on the mobile laptop. This allows a high performance of soil profile inventories (>50/day), but the total efficiency of the survey is mainly given by well defined sites of calibrating profiles: the multiple EC data are regionalized and grouped to select most representative positions.

Evaluation, modeling and results

In spite of many progresses in computer soil data handling, solid modeling of soil bodies is still a complex subject and a domain of human reasoning by experienced soil surveyors. But the modeling can be assisted by several pedometrical operations.

a) Error pre-treatment of biased EC recordings: direct multipol devices often show low EC values caused by contact problems of electrodes, esp. in dry surface soils. Density curves of neighbouring values are used to detect and exclude biased data.

b) Regional cluster analyses define groups of similar EC profiles and most representative candidates for core sampling.

c) Supervised classification, starting with selected soil auger profiles, can be used to classify and discriminate soil groups on maps.

d) Three-D Soil Modeling by a Continuous Rule-based System of Ameskamp (TCRS, 1999) may help to predict and map soil bodies, if fuzzy horizons change with texture layers.

By selected soil transects (2D) and soil bodies (3D) in north-german soilscapes which mainly derive from young (Weichselian) and old (Saalian) loamy tills, often with sand covers and inclusions, we demonstrate the performance and potential of both EC approaches. While the inductive technique (EM38) is faster, the multi-depth access to soil layers and horizons is a clear advantage of devices like the Pluripol. But essentially, the EC results from both methods must be calibrated and interpreted by the profile data of standard or extended augerings.

Literature:

Ameskamp, M. (1997): Three-Dimensional Rule-Based Continuous Soil Modeling. Dissertation, Inst.Informatik, Universitaet Kiel, Germany

Dabas, M. & A. Tabbagh (2003): Comparison of EMI and DC methods used in soil mapping. Precision Agriculture 121-127, 4th ECPA congress, Berlin (ISBN 9076998213)

Durlesser, H. (1999): Bestimmung der Variation bodenphysikalischer Parameter in Raum und Zeit mit elektromagnetischen Induktionsverfahren. Dissertation, Shaker Verlag

Herbst, R. (2002): Bodenschätzung, geoelektrische Sondierung und pedostatistische Modelierungen als Basis von digitalen Hof-Bodenkarten im Präzisen Landbau. Dissertation, Schriftenr. 60 Inst. Pflanzenern.&Bodenk., Universitaet Kiel, Germany

Reimer, G. (2003): Spektrale Naherkundung und Ertragskartierung als Basis von digitalen Hof-Bodenkarten im Präzisen Landbau. Dissertation, Schriftenr. 61 Inst. Pflanzenern.& Bodenk., Universitaet Kiel, Germany

Viscarra Rossel, R.A. & A.B.McBratney (1998): Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay & water content. Geoderma 85,19-39

Contact Information: Ruprecht Herbst, Department of Precision Agriculture, Humboldt Universität zu Berlin, Philippstraße 13, 10115 Berlin, Germany, Phone: 0049/30/2093-6402, Fax: 0049/30/2093-6294,
Email: r.herbst@agrar.hu-berlin.de

Modeling Uncertain Categorical Soil Maps Using a Markov Random Field Approach

Gerard B. M. Heuvelink and *Dick J. Brus*
Soil Science Centre, Wageningen University and Research Centre, The Netherlands

Soil maps are not perfect. There are many reasons that make that a soil map differs from the reality that it aims to represent, the most important ones being measurement, interpretation, classification, generalisation and interpolation errors. The errors in soil maps will propagate to the results of analyses that use soil map information as input. Therefore, it is important that the accuracy of soil maps is quantified, because this allows that the accuracy of the results of the analyses can be assessed and that measures can be taken when the accuracy is below acceptable standards.

The uncertainty about soil maps that results from the various error sources can be modeled using statistical approaches. Geostatistical theory offers a rich instrumentation to model the uncertainty about spatially distributed uncertain soil attributes (Goovaerts, 2001), but the theory is biased towards methods that apply to variables that are measured on a continuous, numerical scale. The instrumentation is not that rich for variables that are measured on a categorical scale. Geostatisticians typically rely on indicator geostatistics to model the uncertainty in categorical spatial variables. Indicator geostatistics was introduced to soil science by Bierkens and Burrough (1993), and has since then been often applied. Although indicator geostatistics appears to work satisfactorily in practice, there are a number of theoretical problems about it that call for the development of alternative approaches (Moyeed and Papritz, 2002). One promising alternative that makes use of the Bayesian Maximum Entropy principle, has already drawn the attention of soil scientists (D'Or and Bogaert, 2004; Brus et al., 2005). In this presentation we explore the usefulness of the Markov random field approach to characterise the uncertainty in categorical soil maps. This approach has a solid statistical basis (Besag, 1974) and has been successfully applied in landcover mapping (e.g. Sarkar et al., 2005) and geology (Norberg et al., 2002).

The key simplification made in the Markov random field approach is that the global probability distribution of a spatially distributed variable can be reduced to the product of local distributions, by assuming that the full conditional distribution of the variable y at location i depends only on the values of y in a neighbourhood ∂_i around i:

$$p(y_i \mid y_j, j \neq i) = p(y_i \mid y_j, j \in \partial_i)$$

The simplification means that *given* the value of y in the neighbourhood, there is no additional information contained in values of y outside the neighbourhood. This has a direct analogy in time series analysis, where the Markov property entails that *given* the present, past and future are independent.

Major problems with real-world application of the Markov random field methodology are estimation of the parameters that characterise the conditional probabilities and computational complexity. The latter problem can be overcome by assuming unidirectional dependencies (Li et al., 2004), but this brings along many unwanted artefacts.

In this presentation we explore the potentials and difficulties associated with application of the (multi-directional) Markov random field approach in soil science, among others by means of an application of the methodology to a case study taken from the 1:50,000 Dutch general-purpose soil map.

References:

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). Journal of the Royal Statistical Society B 36: 192-236.

Bierkens, M.F.P. and P.A. Burrough, 1993. The indicator approach to categorical soil data. I. Theory. Journal of Soil Science 44: 361-368.

Brus, D.J., G.B.M. Heuvelink and P. Bogaert, 2005. Bayesian Maximum Entropy simulation of soil classes using hard and soft data. Abstract book Pedometrics 2005.

D' Or, D. and P. Bogaert, 2004. Spatial prediction of categorical variables with the Bayesian Maximum Entropy approach: the Ooypolder case study. European Journal of Soil Science 55: 763-775.

Goovaerts, P., 2001. Geostatistical modeling of uncertainty in soil science. Geoderma 103: 3-26.

Li, W., C Zhang, J.E. Burt, A.-Xing Zhu and J. Feyen, 2004. Two-dimensional Markov chain simulation of soil type spatial distribution. Soil Science Society of America Journal 68: 1479-1490.

Moyeed, R.A. and A. Papritz, 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. Mathematical Geology 34: 365-386.

Norberg, T., L. Rosen, A. Baran, and S. Baran 2002. On modeling discrete geological structures as Markov random fields. Mathematical Geology 34: 63-77.

Sarkar, A., A. Banerjee, N. Banerjee, S. Brahma, B. Kartikeyan, M. Chakraborty and K.L. Majumder. Landcover classification in MRF context using Dempster-Shafer fusion for multisensor imagery. IEEE Transactions on Image Processing 14: 634-645.

Contact Information: Gerard B. M. Heuvelink. Soil Science Centre, Wageningen University and Research Centre, P.O. Box 37, Wageningen 6700 AA, The Netherlands. Phone: +31 317 482410 Fax: +31 317 482419, Email: gerard.heuvelink@wur.nl

The Effect of Parent Material and Topography on the Scale of Variation in Soil Properties

R. Kerry[1] and *M. A. Oliver*[2]
1Department of Geography, Brigham Young University, Provo, Utah, USA
2Department of Soil Science, University of Reading, Reading, England

Information on how the soil varies spatially is required for several soil science applications, such as land management, engineering and modeling the transport of nutrients and pollutants through the soil. Geostatistical methods are known to provide accurate maps of soil properties, provided that the sampling is appropriate. However, to implement these methods properly can be prohibitively expensive. Traditional soil surveys used expert knowledge about how the soil was likely to vary based on parent material, geology and topography to map it as discontinuous classes. Can such knowledge be quantified and used to map the soil in a continuous way as embodied in the geostatistical approach?

McBratney & Pringle (1999) suggested that average variograms could be used to guide soil sampling and so avoid the need for a reconnaissance survey. Kerry & Oliver (2004) showed that this was possible if the average variograms were computed to take into account parent material and topography. Furthermore, Kerry & Oliver (2002) showed that when no other variograms were available, average and standardized average variograms could be used to krige sparsely sampled soil properties that vary in a similar way (i.e., similar nugget:sill ratio and range). Kerry (2003) suggests that average variograms could be used more generally for kriging when databases of variograms for a range of soil properties are established that are specific to parent material and topography. The validity of this concept depends on there being a plausible explanation, based on soil science theory, as to why one might expect the variogram of a given soil property to vary according to the soil forming and topographic situations. Here we present a possible explanation for differences in the variogram range in relation to parent material and topography.

The variogram range essentially represents the average size of patches with similar values of a soil property. Exhaustive variogram analyses of several soil properties, and associated ancillary data, were undertaken at four field sites on different parent materials that embraced several topographic units. The results showed that variograms of topsoil properties and ancillary data had longer ranges where the soil was more clayey and the landscape was fairly level, such as plateau and valley areas. We suggest that this is associated with the hydraulic conductivity of the soil. Specifically, we argue that water and therefore nutrients move more slowly through soil where small particle size predominates or in the more level areas where the potential energy is weak and cannot hasten the movement of water through the soil or over the surface. Where there is slower horizontal movement of water and minerals through the soil the spatial variation is more continuous which leads to larger areas where properties have similar values and a longer variogram range. Where horizontal movement is slow, the water has time to move vertically through the soil and this could result in the stronger development of soil horizons in level areas. The soil on slopes shows less continuity in the variation; the micro variation reduces continuity and leads to shorter variogram ranges. To assess the validity of these explanations we analyzed our data in several ways and linked the findings to soil processes.

Summary statistics for a range of soil properties varied according to parent material and landscape units, and three-dimensional plots of these properties showed how they vary with topography. Moving correlations between soil properties and elevation calculated with a window size of 5 by 5 sites show that the relations vary according to landscape position. Indeed, a relationship that is positive on the slope may be negative in the valley area and vice versa. Even small-scale features, such as small dry valleys that dissect some of the slopes, were evident in plots of the moving correlation.

Changes in the soil with depth were investigated from plots of resistivity measurements along short transects of 20 m length at an interval of 1 m to give effective depths of inquiry of 25 cm, 75 cm, 127 cm and 185 cm. Large resistivities are generally associated with light textured materials, and changes in soil texture with depth could be detected. The changes in resisitivity with depth were well defined for plateau and valley areas, but not for the slopes. This supports the suggestion that vertical movements of solutes etc. predominate in the more level areas. Variogram analysis of the resistivity data shows that even at this fine resolution variogram ranges are longer for transects on the plateau and in the valley than for the slopes.

The evidence from the analyses described above supports our explanation for differences in variogram range with parent material and topography. There is a need for further research to assess its plausibility more generally. If the use of average variograms could be generalized further and their application underpinned by soil science theory they could be used readily for kriging. This could reduce the costs for those wanting to produce soil contour maps for a range of soil science applications. Databases with variograms of a range of soil properties from different parent materials and topographic units need to be developed.

References:

Kerry, R. 2003. *Determining the effect of parent material and topography on the structure of spatial variation in soil properties for precision agriculture*. Ph.D., University of Reading, Reading, England.

Kerry, R. & Oliver, M.A. 2002. A comparison of kriged predictions using average variograms of soil properties and standardized average variograms of ancillary data. In: *IAMG 2002, 8th Annual Conference of the International Association for Mathematical Geology* (ed. U. Bayer, H. Burger and W. Skala), pp. 93-98.

Kerry, R. & Oliver, M.A. 2004. Average variograms to guide soil sampling for land management. *The International Journal of Applied Earth Observation and Geoinformation*, **5**, 307-325

McBratney, A.B. & Pringle, M.J. 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture*, **1**, 125-152.

Contact Information: Ruth Kerry, Brigham Young University, Geography Department, 690 SWKT, Provo, Utah, 84602, USA, Phone: (801)422-4311, Fax: (801)375-5159, Email: ruth_kerry@byu.edu

Mapping Soil Structure using Ranked Observations and Indicator Kriging

***R. Kerry*[1] and *M. A. Oliver*[2]
[1]Department of Geography, Brigham Young University, Provo, Utah, USA
[2]Department of Soil Science, University of Reading, Reading, England

Structure is an important physical feature of the soil that affects the nature and distribution of pores, which hold water, air and allow roots to penetrate. It is also indirectly associated with much of the activity in the soil in terms of water movement, oxygen, microorganisms and nutrient uptake. A soil without any obvious organisation of its components is known as apedal. This can be single grain where the mineral material is almost surrounded by a continuous pore phase (usually sandy soil), or massive where the mineral material is continuous and the pores are discontinuous (usually clayey soil). This suggests that information about the soil texture needs to be recorded where soil structure is observed. Accurate maps of top- and sub-soil structure are desirable for a wide range of models that aim to predict erosion, solute transport, or flow of water through the soil. Such maps would also be useful to land managers when deciding how to apply nutrients and pesticides in a site-specific way, to target sub-soiling, and to stabilize soil structure.

Typically, soil structure is inferred from bulk density (Bartoli *et al*., 2005) or penetrometer resistance measurements (Munkholm *et al.,* 2003), but the results of both methods are affected by many properties of the soil that vary spatially; in particular the moisture content at the time of sampling, soil texture, and the presence of stones. Recently, soil resistivity and conductivity surveys have been used to infer soil structure, but as for bulk density and penetrometer resistance there are problems in interpreting the results from these methods in different places.

We suggest a return to traditional methods for observing soil structure and to use geostatistics to map these observations. Several authors have used indicator kriging to map nominal data such as water table class (Bierkens & Burrough, 1993), and notions of soil quality (Smith *et al*. 1993). Therefore, it seemed an appropriate approach for mapping soil structure.

The top- (0–15 cm) and sub-soil (20–30cm) were sampled by auger at the intersections of 20-m and 30-m grids at four field sites where the soil had developed on different parent materials (two of the sites were very stony). Observations of soil structure were made using the descriptive system of Hodgson (1974), which takes into account the degree of ped development and type of ped. The observations were then ranked in three different ways. One used Hodgson's (1976) structure charts for three textural groups that show which type and degree of ped development would be considered good, moderate, poor or very poor structure. The second used Peerlkamp's (1967) descriptions of structure with ranks of 1-9 for clay and loams and sandy soils; the higher ranks represent poorer soil structure. Finally, the structural observations were ranked for clayey and sandy soils into three classes (good, medium or poor) and seven classes, where each different type of structural observation e.g. weakly developed crumb structure or strongly developed blocky structure had its own rank (Kerry, 2003).

Indicator and the sum of indicator variograms (SIV) were computed and modeled for each method of ranking. Cross-validation was done using the SIV for each ranking system and the results were used to determine which method of ranking would be the most suitable for kriging. The individual indicators (presence or absence data) were then kriged with the appropriate variogram to predict the probability of encountering soil with the structure represented by that indicator. The predictions were mapped and such maps could be used to indicate where sub-soiling or mulching was needed to improve soil structure. In addition the predictions could be incorporated into models of soil erosion, and water and solute transport through soil.

Observations of soil structure are reasonably straightforward to make for a single observer. The nature of soil texture, i.e. clayey, sandy or silty, at the surface and at depth could be assessed at the same time by hand texturing. This means that in fields or areas where the soil texture varies spatially and with depth, separate maps of the structure for the clayey and sandy areas could be created. The observations are not affected by the soil moisture at the time of sampling or a large stone content. In addition, samples could be obtained automatically using a vehicle-mounted soil corer and observations made later in a laboratory. The relative merits of the different ranking systems at each site will be discussed and a protocol for mapping that incorporates economy in the number of observations made will be presented.

References:

Bartoli, F., Genevois-Gomendy, V., Royer, J.J., Niquet, S., Vivier, H. & Grayson, R. 2005. A multiscale study of silty soil structure. *European Journal of Soil Science*, **56**, 207-233.

Bierkens, M.F.P. & Burrough, P.A. 1993. The indicator approach to categorical soil data.II. Application to mapping and land use suitability analysis. *Journal of Soil Science*, **44**, 369-381.

Hodgson, J.M. 1974. *Soil sampling and soil description* Clarendon Press, Oxford.

Hodgson, J.M. 1976. *Soil survey field handbook. Technical monograph No. 5* Soil Survey of England and Wales, Harpenden.

Kerry, R. 2003. *Determining the effect of parent material and topography on the structure of spatial variation in soil properties for precision agriculture*. Ph.D., University of Reading, Reading, England.

Munkholm, L.J., Schjonning, P., Rasmussen, K.J. & Tanderup, K. 2003. Spatial and temporal effects of direct drilling on soil structure in the seedling environment. *Soil & Tillage Research*, **71**, 163-173.

Peerlkamp, P.K. 1967. *West European methods for soil structure determination II.*

Smith, J.L., Halvorson, J.J. & Papendick, R.I. 1993. Using multiple-variable indicator kriging for evaluating soil quality. *Soil Science Society of America Journal*, **57**, 743-749.

Contact Information: Ruth Kerry, Brigham Young University, Geography Department, 690 SWKT, Provo, Utah, 84602, USA, Phone: (801) 422-4311, Fax: (801) 375-5159, Email: ruth_kerry@byu.edu

Spatial and Temporal Changes of Crack Formation of a Vertisol in the Texas Gulf Coast Prairie

Andrea S. Kishné[1], *Cristine L. S. Morgan*[1] and *Wesley L. Miller*[2]
[1]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA
[2]USDA-Natural Resources Conservation Service, Victoria, TX, USA

During drying process of shrinking-swelling clay soils, formation of cracks affects the surface hydrology of a landscape dramatically because the cracks facilitate rapid transport of water and pollutants into the soil and/or groundwater. A small plot of a Laewest clay (0-1% slopes) in Victoria County, Texas Gulf Coast Prairie was studied over 10 years to characterize, analyze and model the changes of cracks related to precipitation and microtopography.

A 100 m^2 area in native rangeland was selected for study. From 1989 to 1998, surface crack width and length were measured and plotted within 5 cm accuracy on forty-two diagrams with a 1 inch to 1 m scale. When each diagram was made, representative crack depth, water table depth and soil moisture were measured. Precipitation was also recorded. Microtopography was measured on a 0.5 m x 2 m grid using a survey quality laser level. A soil characterization pit was opened at 24 m from the study area to describe and sample representative areas of a microhigh and microlow.

The crack diagrams were scanned, and were digitized in ArcView. The analysis of spatial autocorrelation provides the range of independent samples and the basis for analysis of variance of width and depth measured on the field. A model to estimate crack depth as a function of crack width will be presented. Crack volume and density per unit area will be analyzed over time within categories of microtopography and in relation to precipitation, evapotranspiration and groundwater. Results will be used for parameterizing a preferential flow model in the future.

<u>Contact Information</u>: Andrea S. Kishné**,** Department of Soil and Crop Sciences, Texas A&M University, 370 Olsen Blvd., College Station, TX 77843-2474, USA, Phone: 979-845-2541, FAX: 979-845-0456,
Email: akishne@ag.tamu.edu

Modeling of Regional Soil Nitrate-Nitrogen Patterns Using a Mixed Geospatial Modeling Approach

S. Lamsal, *S. Grunwald, G.L. Bruland, C.M. Bliss* and *N.B. Comerford*
Soil and Water Science Department, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA

The Santa Fe River, which drains into the Suwannee River and the Gulf of Mexico, has been contributing to elevated loads of nitrate-nitrogen in the ground, spring and surface waters. Mixed land use includes pine plantation (32%), wetlands (16%), upland forest (15%), improved pasture (14%), urban (9%), forest regeneration (6%), crops (5%), rangeland (4%) and a variety of specialized land uses such as tree groves, dairies, and feeding operations. The soils are predominantly sandy with loamy to clayey argillic and spodic horizons. The area is characterized by karst terrain which promotes the many solution basins that form the Santa Fe River Watershed (SFRW) (3,585 km^2).

Our objectives were to describe regional, seasonal, and geospatial patterns of soil nitrate-nitrogen across the SFRW. We used soil data from the Soil Survey Geographic Database and land use data derived from Landsat satellite imagery to delineate soil-land use categories (SLC) using ArcGIS. The SLC were used as strata in a stratified random sampling design to select soil sampling sites proportional to the aerial extent of individual SLC. In addition we targeted "high-risk" SLC expected to contribute to high nitrogen loads. Soil samples were collected as composites, proportional to the support size, from four depth increments (0 to 30, 30 to 60, 60 to 120 and 120 to 180 cm) during Sept. 2003, Jan. 2004 and May 2004 and analyzed for KCl-extractable nitrate-nitrogen content. The nitrate-nitrogen values from different depths were used to calculate profile averages for each sampling site. We adopted a hybrid geospatial modeling approach that combined sparsely measured soil nitrate-nitrogen observations collected in three seasons with dense auxiliary environmental datasets to predict nitrate-nitrogen within the SFRW. A suite of 79 soil-landscape and environmental variables was assembled for each sampling site using local and focal (neighborhood) geospatial methods. We developed season-specific regional nitrate-nitrogen models representing "wet/end-of cropping season" (Sept.), "dry winter season" (Jan.) and "dry spring season" (May), respectively.

A classification tree model for the Sept. 2003 event generated four nitrate-nitrogen classes: A ($\leq$ 0.01 $\mu g\ g^{-1}$), B (0.01 – 0.23 $\mu g\ g^{-1}$), C (0.23 – 0.56 $\mu g\ g^{-1}$) and D ($\geq$ 0.56 $\mu g\ g^{-1}$). The four classes were upscaled to the watershed scale using the following predictor variables: land use, physiographic division, and soil order. This tree-based trend model was the best estimate of soil nitrate-nitrogen representing the fall season. Regression trees were used to develop trend models of nitrate-nitrogen for January and May 2004. We identified the predictor variables - reflectance band 4, land use and available water content low - for the January event; and land use, focal compound topographic index, soil component name, and focal reflectance band 7 for the May event. The tree-based trend models were used to predict soil nitrate-nitrogen across the SFRW for January and May 2004. Although the predicted values differed for the January and May events, the spatial distributions of high and low values were similar. The relative error was 0.86 for the January event and 0.51 for the May event. Because the residuals of the January sampling event showed strong spatial dependence with a nugget to sill ratio of < 0.25, we proceeded with semivariogram modeling and ordinary kriging of residuals. The residual autocorrelation was modeled using a spherical semivariogram with a range of 5,000 meters, a partial sill of 9.2 and a nugget of 1.8. The interpolated residual model was added to the trend model to produce the final nitrate-nitrogen distribution map (regression kriging). Adding the kriged residuals to the tree-based model corrected for the over- and under-predictions made by the trend model. This resulted in a lower mean square prediction error (MSE) of 3.04 and higher G coefficient of 82.6 for the regression kriging model when compared to the tree-based trend model with a MSE of 11.32 and G coefficient of 53.9.

The hybrid geospatial upscaling technique used auxiliary environmental variables for trend modeling and accounted for the spatial autocorrelation of the residuals, thereby improving predictions across the watershed. Future research involves identifying short range variability of nitrate-nitrogen in a representative area of the watershed, and incorporating the short range variation to improve predictions at the watershed scale.

Acknowledgement: This project was funded by the USDA Nutrient Science for Improved Watershed Management Program (#2002-00501). We thank Dr. Mark Clark, Dr. Don Graetz, and Dr. Randy Brown for their contributions.

Contact Information: Sanjay Lamsal, Soil and Water Science Department, IFAS, University of Florida, 2169 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA, Phone: 352-392-1951 ext. 233, Fax: 352-392-3902, Email: slamsal@ufl.edu

Wavelet Analysis of Categorical Soil Variables, Some Approaches Based on Indicators

R. Murray Lark
Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, Hertfordshire, United Kingdom

Wavelet transforms have been used for the analysis of continuous soil variables, and have given insight into complex variability that is not consistent with simple assumptions of stationarity in the variance (Lark & Webster, 1999; Redding *et al*., 2003; Si & Farrell, 2004). If the variability of a soil property changes from one part of the landscape to another, and includes transient features, then it is best analysed using the wavelet transform — since this represents variation on a basis of wavelet functions that have a compact support and represent local variation at some scale of generalization.

Many soil data consists of categorical or multistate variables rather than continuous ones. Such variables include soil classes (e.g. the soil series) and properties that may be recorded in the field by experienced surveyors (e.g. profile drainage classes, the structural class of peds or the local topography). These variables can be very useful. They may be quicker to collect than data that require a laboratory analysis, and many observations of such variables have been recorded historically. Further, the identification of some multistate variable, such as profile drainage status, by an experienced soil surveyor, conveys a lot of information on local soil conditions. The variation of soil as represented by categorical variables may therefore be of both practical and scientific interest. However, since the values of these variables, even when they are ordered, do not constitute a mathematical field, we cannot analyse them in the same way as continuous variables.

This problem has been addressed in the context of geostatistics. Bierkens & Burrough (1993) showed how indicator methods can be used to analyse the spatial variation of categorical variables and to predict their value at unsampled sites. In this paper I shall show how complex and non-stationary variation of categorical soil variables can be analysed with the wavelet transform.

It is possible to transform each categorical variable to an indicator and then perform a multiresolution analysis using wavelet basis functions. This can be done in two general ways. The indicator may be treated as a continuous variable (i.e. our raw data take values of 0 or 1, but smooth representations of the data at different scales may take other real values), or the wavelet transform may be constrained so that all smooth approximations and detail components take values of 0 or 1. Both these approaches may be useful for representing the variability of categorical varibles and different scales of generalization.

When our concern is to characterize complex and non-stationary variation we can take a different approach. I shall demonstrate an analysis based on a proposal by Wang & Johnson (2002) in which we find a mapping from a full set of p indicator variables defined on a p-state categorical variable to a new real-valued variable with a continuous spectrum. Following the work of Stoffer *et al*. (1993) on stationary categorical variables, we find a mapping that maximizes some feature of interest — in this case the local wavelet variance.

After describing the method I shall demonstrate its usefulness in soil science by the analysis of a set of observations on a categorical variable, topographic class, observed at sites on a transect across a gilgai landscape in Australia. The basic periodic pattern is shown to have non-stationary features — both transient variations and a change in the distribution of variation between spatial scales. These features are identified by an appropriate inferential method on the wavelet coefficients that allows the hypothesis to be tested that the features of the data represent random fluctuations in the realization of an underlying stationary process. I can then show that a change in the variation of topography that is not consistent with stationarity is also reflected in the spatial variation of a continuous soil property measured on the same transect.

The implications of these results are that an analysis on a rapidly determined categorical variable can allow us to identify features of the variation of the soil that are not consistent with stationarity. As well as giving insight into the spatial variation of the soil, this could also be a basis for subdividing a study region into apparently stationary subregions prior to sampling for continuous variables that are more expensive to determine.

I shall then consider possible developments of this method, including the incorporation of wavelet packet bases into the analysis.

References:

Bierkens, M.F.P. & Burrough, P.A. 1993. The indicator approach to categorical soil data. 1. Theory. *Journal of Soil Science*. 44, 361–368.

Lark R.M. & Webster R. 1999. Analysis and elucidation of soil variation using wavelets. *European Journal of Soil Science*. 50, 185–206.

Redding T.E., Hope G.D., Fortin M.J., Schmidt M.G. & Bailey W.G. 2003. Spatial patterns of soil temperature and moisture across subalpine forest-clearcut edges in the southern interior of British Columbia. *Canadian Journal of Soil Science*. 83, 121–130.

Si B.C. & Farrell R.E. 2004. Scale-dependent relationship between wheat yield and topographic indices: A wavelet approach. *Soil Science Society of America Journal.* 68, 577–587.

Stoffer, D.S., Tyler, D.E. & McDougall, A.J. 1993. Spectral-analysis for categorical time-series — scaling and the spectral envelope. *Biometrika*. 80, 611–622.

Wang, W. & Johnson, D.H. 2002. Computing linear transforms of symbolic signals. *IEEE Transactions on Signal Processing*. 50, 628–634.

Contact Information: R. Murray Lark, Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, Herts, AL5 2JQ, United Kingdom, Phone: +44 (0)1582 763133 , Fax: +44 (0)1582 760981, Email: murray.lark@bbsrc.ac.uk

Pedometricians! Use the REML–E-BLUP for Spatial Prediction!

R. Murray Lark[1], *Brian R. Cullis*[2] and *Sue J. Welham*[1]
[1]Biomathematics and Bioinformatics Division, Rothamsted Research, United Kingdom
[2]New South Wales Agricultural Research Institute, Wagga Wagga, NSW, Australia

Most geostatistical predictions made by pedometricians correspond to the empirical best linear unbiased predictor or E-BLUP. The best linear unbiased predictor (BLUP) is based on the linear mixed model of variation with an uncorrelated error term, one or more spatially dependent random effects and one or more fixed effects (Stein, 1999). If the fixed effect is simply the unknown mean, assumed to be the same everywhere, then the BLUP is ordinary kriging, if the fixed effects consist of polynomial functions of the spatial co-ordinates then the BLUP is universal kriging, UK, (or regression kriging), and if the fixed effects are polynomial functions of some other variable(s) known everywhere (such as remotely sensed data) then the BLUP is kriging with an external drift, KED, (or regression kriging). To obtain the BLUP we need variance parameters for the random effects and error (which will include terms that describe spatial dependence). If we use estimates of these then our prediction is called the empirical BLUP, E-BLUP.

It is when estimating these variance parameters that pedometricians may have problems. In the ordinary kriging case, where the variable is assumed to be intrinsically stationary, the variogram can be estimated from the data directly and modeled with an appropriate parametric function. This is standard practice, although it should be noted that objections have been raised against this procedure as commonly implemented because of the strong auto-correlation between estimates of the variogram at successive lags (Stein, 1999).

Difficulties arise in the UK or KED case. The empirical estimator of the variogram cannot separate the non-stationary mean from the random variation. Two general approaches have been taken to this problem. The first is to assume that the effects of any trend are small at short lags, and that at long lags we can identify pair comparisons for some lag $\mathbf{h}$, $\{z(\mathbf{x}), z(\mathbf{x+h})\}$ that are not affected by the trend (perhaps because the comparison is perpendicular to the trend. This approach is reasonable, but it is not generally applicable. The second method is to estimate the fixed effects first, then to obtain a variogram for the residuals. The problem with this approach is that the fixed effects are nuisance parameters with respect to the variance parameters, and so the latter will be biased. This is true whether we estimate the fixed effects by ordinary least squares or by generalized least squares.

The solution is to estimate the variance parameters by residual maximum likelihood (REML) due to Patterson & Thompson (1971). This solution has existed for some time, but has received surprisingly little attention in soil science. In REML we obtain from our data a new random variable that is independent of a specified set of fixed effects and which has a covariance matrix derived from a known form of covariance function for the errors from the trend. We apply maximum likelihood to estimate the parameters of that function. REML gives unbiased estimates of the variance parameters. These are then used to estimate the fixed effects by generalized least squares.

REML–E-BLUP can be efficiently implemented in practice using the average information algorithm of Gilmour *et al* (1995), although there are hazards if we wish to use the spherical variogram function (Lark & Cullis, 2004).

In this poster we outline the principles of REML–E-BLUP, and illustrate its application to spatial prediction of a soil variable with a pronounced spatial trend.

References:

Gilmour, A.R., Thompson, R. & Cullis, B.R. 1995. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*. 51, 1440–1450.

Lark, R.M. & Cullis, B.R. 2004. Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science*. 55, 799–813.

Patterson, H.D. & Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545–554.

Stein, M.L. 1999. *Interpolation of spatial data: some theory for kriging*. Springer, New York.

Contact Information: R. Murray Lark, Biomathematics and Bioinformatics Division, Rothamsted Research, Harpenden, Herts, AL5 2JQ, United Kingdom, Phone: +44 (0)1582 763133 , Fax: +44 (0)1582 760981, Email: murray.lark@bbsrc.ac.uk

Accounting for Interclass Dependences in Stochastic Simulation of Categorical Soil Variables Using Markov Chain Geostatistics

***Weidong Li*[1]** and *Chuanrong Zhang*[2]
[1] Department of Geography, University of Wisconsin, Madison, WI, USA
[2] Department of Geography and Geology, University of Wisconsin, Whitewater, WI, USA

Interclass dependence is the normal characteristic of multinomial classes of categorical variables. Strong interclass dependence exists among classes of complex categorical soil variables such as soil types. Obviously, geostatistical simulations ignoring interclass dependence do not make full use of the heterogeneity information conveyed by observed data of this kind of categorical variables. Generally, interclass dependence may include cross-correlations, juxtaposition, and directional asymmetry of spatial distribution of classes. Conventional kriging geostatistical methods seem difficult in accounting for interclass dependence of a number of classes.

While the recent developing trend in kriging geostatistics is focusing on incorporating multiple-point statistics from various data sources such as training images (i.e., multiple-point geostatistics) (Journel, 2005), the recent progress in geostatistical development is seeing the emergence of Markov chain geostatistics (MCG), which has been a long-time research topic of the authors. Both approaches have one common purpose – to better imitate the complex spatial structure of discrete variables. MCG uses non-linear Markov chain-based estimators. It deals with many classes simultaneously and has no computational limitation on the number of classes. MCG is free of some difficult issues bothering indicator kriging, such as order relation problems and coregionalization of input parameters. Interclass dependence is naturally incorporated into simulation through cross-transition probabilities. A new spatial heterogeneity measure - transiogram was proposed recently to provide continuous transition probabilities for simulations. Because of these advantages, realizations generated by MCG are highly imitative to the real patterns given a number of observed data. Potentially, MCG can work with any kinds of data, such as points, lines, small areas, and mixtures. Although MCG is still at the early stage of theoretical and methodological development, it has shown exciting features in simulating multinomial classes.

MCG is a new non-kriging geostatistics. The basic idea of this geostatistics is to use Markov chains to perform multidimensional interpolation and simulation. Compared with the covariance-based (or variogram-based) geostatsitics, MCG is transition probability-based. Compared with the kriging-based geostatistics, MCG is Markov chain-based. MCG directly uses Markov chains to accomplish conditional simulation. Some pioneer studies that are contributive to MCG include Carle and Fogg (1997), Elfeki and Dekking (2001), and more significantly, Pickard (1980). Our recent publications can be seen in Li et al. (2004) and Zhang and Li (2005), which show some preliminary results. The basic idea of MCG is that an unknown location is related on its nearest known neighbors in different directions. With a Markov chain moving around in a space, its conditional probability distribution at any unknown point is entirely dependent on its nearest known neighbors in different directions. The interaction between each nearest known neighbor and the unknown location is expressed by a transition probability at the corresponding distance. Therefore, transiograms are the explicit components of the conditional probability function.

Since MCG as applicable techniques (software system) is still in developing, currently we emphasize its special capabilities – conditional simulation of multinomial classes, such as soil types, with interclass dependence. This presentation will demonstrate how this new geostatisitcs works and some simulated results of soil types. The fully developed MCG will work with any discrete variables and data types.

References:

Carle, S.F., and Fogg, G.E., 1997, Modeling spatial variability with one and multidimensional continuous-lag Markov chains. *Math. Geol.* 29:891-918.

Elfeki, A.M., and Dekking, F.M., 2001, A Markov chain model for subsurface characterization: theory and applications. *Math. Geol.* 33: 569-589.

Li, W., Zhang, C., Burt, J.E., Zhu, A.X. and Feyen, J., 2004, Two-dimensional Markov chain simulation of soil type spatial distribution. *Soil Science Society of America Journal* 68, 1479-1490.

Journel, A.G., 2005, Beyond covariance: The advent of multiple-point geostatistics. In: Geostatistics Banff 2004, Series: Quantitative Geology and Geostatistics, vol. 14, O. Leuangthong and C. V. Deutsch (eds.) 2005, XXVIII, p.1167

Pickard, D.K., 1980, Unilateral Markov fields. *Adv. Appl. Probab.* 12: 655-671.

Zhang, C. and Li., W., 2005, Markov chain modeling of multinomial land-cover classes. *GIScience and Remote Sensing* 42: 1-18.

Contact information: Weidong Li, 1563 W. Wildwood Rd., #12, Whitewater, WI 53190, USA.
Email: weidong6616@yahoo.com

Automated Predictive Ecological Mapping in a Forest Region of B.C., Canada, 2001-2005

R. A. MacMillan[1], *D. E. Moon*[2] and *R. A. Coupé*[3]
[1]LandMapper Environmental Solutions Inc., Edmonton, Alberta, Canada
[2]CDT Core Decision Technologies Inc., Richmond, B.C., Canada
[3]Southern Interior Forest Region, B.C. Ministry of Forests, Williams Lake, B.C., Canada

Procedures for automated predictive thematic mapping were developed and applied to project areas totaling more than 3 million hectares of forested land in British Columbia, Canada. The effective scale of mapping was 1:20,000 using data at a grid resolution of 10-25 m.
The methods can be described as a hybrid of automated, semi-automated and manual procedures that develop and apply heuristic, rule-based conceptual models of ecological-landform and soil-landform relationships in a manner similar to the CLORPT or SCORPAN approaches as described by McBratney et al., (2003). The procedures do not use any field sampling to develop or train classification rules. All rules are constructed by examining and deconstructing published field guides that define the required output classes and that document the current expert understanding of the conditions and criteria that control the spatial distribution of the desired output classes. An iterative, trial and error, process is used to develop, apply, evaluate and revise rules. Local expert knowledge is used at each stage to evaluate each new set of output results and to guide refinement of the classification rules. Field sample observations obtained along randomly selected closed traverses are collected following a line-intercept approach and used to assess the accuracy of the final predictions of ecological classification. These field sample observations require a considerable investment of time and effort and represent a significant proportion of the total project costs.

The procedures were implemented using the LandMapR suite of programs that capture and apply expert knowledge to digital input layers using a Semantic Import (SI) Model implementation of Fuzzy Logic. The methods rely heavily on terrain derivatives extracted from available digital elevation models (DEMs). Hydrological analysis of surface flow networks provides the basis for many of the most useful parameters. These regional, hydrological variables are sensitive to variation in the context and scale of terrain features within areas of interest and have proven useful and effective additions to terrain derivatives computed within local windows of fixed dimensions. The primary input has been the BC provincial Terrain Resource Information Management (TRIM) digital elevation model (DEM) surfaced to a regular raster grid. Other input layers include manually interpreted maps of parent material texture, depth and ecological exception classes, manually prepared maps of the spatial distribution of ecological zones of the BC Biogeoclimatic Ecosystem Classification (BEC) system and, to a limited extent, LandSat7 digital satellite imagery.

Application of the procedures progressed from a pilot project through projects to evaluate operational scale-up to full-scale commercial application to millions of hectares. Costs were reduced from a high of $3.50 per ha in the late nineties to less than $0.25 per hectare in 2005. Rates of progress increased from 150,000 ha per person year to more than 2.0 million ha per person year. Independent assessments of map accuracy demonstrated accuracy equivalent to the highest accuracies reported for alternatives, including traditional manual mapping methods. All maps achieved levels of accuracy (66-70%) that were comparable to the agreement achieved among four ecological experts independently classifying the same flagged traverses (65%). This suggests that predictive accuracy is approaching the theoretical limit set by field measurement error.

We conclude that we have formalized and automated many of the concepts and techniques previously used to create thematic maps of ecosystems and soils using manual interpretation of stereo air photos and ancillary data combined with field observations. Our procedures act as replacements for the traditional mapping tools of air photos, stereoscopes and grease pencils. The procedures implement an explicit hierarchy that partitions space into successively smaller and more homogeneous areas comparable to those produced by traditional methods of hierarchical ecological land classification. They produce spatial predictions of ecological or soil classes that mimic or reproduce depictions produced by traditional manual interpretation. Typically our automated predictive maps provide a finer level of spatial detail than is offered by manually prepared maps of comparable scale (1:20,000 to 1:50,000). We have speeded up and lowered the costs associated with using traditional approaches. We have shown that automated landform analysis, combined with both Boolean and fuzzy logic, is able to provide most of the inputs required to capture and apply the concepts of landform control referenced by typical landform-based soil and ecological models and classification systems. We have shown that extensive field sampling is not necessary in order to develop or train effective classification rules but that unbiased field observations are essential in order to assess the accuracy of the final classifications. We have demonstrated that it is possible to produce accurate and cost-effective ecological, landform and soil maps by applying fuzzy and Boolean logic and automated landform analysis procedures to widely available spatial data.

Reference: McBratney, A. B., Mendoca Santos, M. L. and Minasny, B. 2003. On digital soil mapping. Geoderma. 117: 3-52.

Contact Information: R. A. (Bob) MacMillan, LandMapper Environmental Solutions Inc., 7415 118 A Street NW. Edmonton, Alberta, Canada, T5G 1V4, Phone: 780-435-4531, Fax: 780-436-1788,
Email: bobmacm@telusplanet.net

A New Approach to Automated Extraction and Classification of Repeating Landform Types

R. A. MacMillan
LandMapper Environmental Solutions Inc., Edmonton, Alberta, Canada

Most examples of automated predictive mapping of soils, landforms or ecotypes presented in the literature have effectively limited themselves to partitioning hillslopes along a catenary toposequence from crest to trough to delineate and classify segments of the landscape equivalent in concept to landform facets or hillslope elements. Fewer published examples describe efforts to automatically classify repeating types of landforms equivalent in concept to low or high hills, level, undulating or hummocky plains or rolling, ridged or hummocky uplands. Over the period 2001-2005, the LandMapR toolkit was used operationally and commercially for automated predictive mapping of ecological entities at a scale of 1:20,000 for several million hectares of forested land in British Columbia, Canada. Initially these LandMapR procedures also focused almost exclusively on delineating and classifying environmental settings, or landscape situations, that were conceptually equivalent to partitioning a toposequence into landform facets. With increasing experience, it was determined that, depending upon the size, scale and type of landscape under consideration, different knowledge bases and different rule sets were required to partition different landscapes into different environmental settings. The LandMapR procedures were enhanced to permit a pre-classification of large areas of interest into different classes of repeating types of landforms. Each class of repeating landform type differed from other classes in terms of its size, scale, relief and complexity. The new procedures model simulated surface flow for both a normal and an inverted digital elevation model (DEM) to first locate and extract hydrological spatial entities equivalent to individual peaks or individual hillslopes. Individual peak catchments and/or individual hill slopes are used to link groups of grid cells into spatial entities that are then classified into repeating types of landforms of different size, scale, relief, shape and orientation. Each peak catchment exactly and precisely defines one trough to peak to trough wavelength of the landscape and captures one full cycle of topographic variation at the scale of interest. Each hillslope exactly captures one half of a topographic repeat cycle from crest to trough or channel. A wide range of measures of landform morphology is computed for each delineated peak or hillslope. The morphological attributes computed for each cell of each delineated peak or hillslope are summarized to characterize each entity in terms of its frequency distribution of relief, slope gradient, slope length, aspect, curvature, wetness index and other attributes of interest. These statistical distribution functions, along with means, modes, minimums and maximums are used as inputs to procedures that classify each spatial entity into a distinctive class of repeating landform type. Each distinctive class of repeating landform type is used to define a different classification domain within which different sets of rules are used to further partition the landscape into different types and numbers of landform elements, or environmental settings, along a toposequence from crest to trough for each peak or hillslope. This multi-level, hierarchical approach to classifying landscapes into first landform types and then landform elements has proven to be useful and effective. It has increased the ability to customize or tailor rules for recognizing the more detailed landform elements so that they are sensitive to, and adjust to, differences in the size, scale, relief and slope length of the larger landscape of which they are components. Maps of landform types are interesting and useful in their own right and bear a striking resemblance to soil maps at the level of soil associations.

Contact Information: R. A. (Bob) MacMillan, LandMapper Environmental Solutions Inc., 7415 118 A Street NW. Edmonton, Alberta, Canada, T5G 1V4., Phone: 780-435-4531, Fax: 780-436-1788,
Email: bobmacm@telusplanet.net

Using Outputs from Hydrological Flow Modeling as Inputs to Predictive Mapping

R. A. MacMillan
LandMapper Environmental Solutions Inc., Edmonton, Alberta, Canada

Increasingly, research efforts have recognized the benefits of using outputs from hydrological flow modeling as inputs to procedures for predictive modeling of soils, landforms and ecological spatial entities. This paper identifies and describes how commercial predictive mapping efforts undertaken in Canada by LandMapper Environmental Solutions Inc. from 2001-2005 have benefited from calculation and consideration of measures derived from hydrological flow modeling. We describe calculation and use of a wide variety of regional and compound hydrological terrain derivatives, hydrologically derived linear spatial entities such as topologically encoded stream and ridge networks and hydrologically derived aerial spatial entities such as catchments and sub-catchments, hillslopes and peaks. The LandMapR procedures model hydrological flow on both an original digital elevation model (DEM) and on an inverted, mirror image, of the DEM for each area of interest. Modeling surface flow for an inverted DEM offers capabilities to apply all algorithms developed to model simulated surface flow in the normal down slope direction to simulate surface flow in a notional upslope direction. Instead of flowing down to locate and describe pits and channels, flow in an inverted DEM goes upslope to locate and describe peaks and divides (pits and channels in the inverted DEM). Measures of absolute flow distance, elevation change and upslope area in the down slope and upslope directions have been combined to compute highly useful measures of relative landform position. Compound topographic index (or wetness index) has been used as a measure of relative moisture condition and diffuse upslope area as a measure of relative landform position for predictive mapping. Calculations of vertical and horizontal distance to simulated stream channels and to identified depressions or wetlands have been used to establish measures of local and regional hydrological context to aid in recognizing particular classes of ecological or soil spatial entities. Individual peak catchments and/or individual hill slopes have been used to define spatial entities that are classified into repeating types of landforms of different size, scale, relief, shape and orientation. Each peak catchment exactly and precisely defines one trough to peak to trough wavelength of the landscape and captures one full cycle of topographic variation at the scale of interest. Each hillslope exactly captures one half of a topographic repeat cycle from crest to trough or channel. These spatial entities capture repeat cycles in landscapes of interest in a manner that is both more precise and more theoretically meaningful than comparable efforts to compute the dimensions of local landforms using geostatistics and variograms. We conclude that measures of hydrological scale, moisture regime and context extracted from automated analysis of DEMs can provide more relevant and useful inputs to predictive mapping than do the more traditional measures of local surface form and orientation (e.g. slope, aspect and curvatures) computed within local roving windows of fixed dimensions. We anticipate increasing use for predictive mapping of derivatives and spatial entities computed from hydrological analysis of DEMs. We argue for tighter and more explicit linkages between entities mapped to depict the variation in space of soils, landforms or ecotypes and hydrological spatial entities such as depressions, stream networks and the hillslopes and local catchments that flow into them.

Contact Information: R. A. (Bob) MacMillan, LandMapper Environmental Solutions Inc., 7415 118 A Street NW. Edmonton, Alberta, Canada, T5G 1V4, Phone: 780-435-4531, Fax: 780-436-1788,
Email: bobmacm@telusplanet.net

Adaptive Sampling for Automated Soil Mapping

B. P. Marchant and *R. M. Lark*
Biomathematics & Bioinformatics Division, Rothamsted Research, Harpenden, Hertfordshire, England

Introduction

One of the biggest problems for management of within-field variation is obtaining adequate information on the variations of important variables at acceptable cost. If soil properties are to be mapped with adequate precision, then the sampling intensity needed will differ between fields. Therefore farmers and agronomists run the risk of completing a survey then finding that they have substantially over-sampled and so wasted effort, or that they are not able to produce a reasonable map from the our data because they are too sparse. In this talk we describe the development of adaptive methods for designing sample schemes which are suited to the variable being measured despite the fact that we start with little or no information on the spatial variability of the variable.

Rather than deploying all the sample effort in a single designed phase, our adaptive approach splits it up into several phases (the total number of which may not be known in advance). In early phases we are concerned primarily to characterize the spatial variability of the variable of interest, in later phases we focus on sampling to support mapping of the variable by geostatistical interpolation. Any phase is designed using information from past phases (including the known values of the property at previously-sampled sites) and generates information used to decide how to sample next. We hypothesize that, by such adaptive sampling, it should be possible to arrive at an efficient final sampling scheme in the absence of any prior information on variability. However, information from ancillary sources could be used to initiate the process, although if it turns out that the ancillary data is a poor guide in any specific case this will be recognized.

Adaptive Sampling Algorithms

We describe two adaptive sampling algorithms. The first uses spatial simulated annealing (Van Groenigen, 1999) to design an entire survey prior to sampling. This survey is adapted as measurements are collected and analysed so that the final survey is suited to the variable. A variogram must be assumed to design the initial survey. This variogram may be based upon any prior information that is available such as the variograms of previous surveys of the same variable in different locations. The spatial simulated annealing algorithm selects the sampling locations which minimize an expression for the mean squared error in the survey including the effects of variogram uncertainty. The first phase of sampling consists of a fixed number of sample points from the initial survey which provide the most information about the variogram of the variable. The variogram is then estimated from the collected data and the remainder of the survey is redesigned according to this estimated variogram. Another phase of sampling from this design is then collected and analysed and again the remainder of the survey is redesigned. This process continues until enough samples are collected for the mean squared error in the survey to lie within a pre-specified tolerance.

The second adaptive algorithm follows McBratney *et al.* (1981) and splits the survey into a reconnaissance survey which is designed to learn about the variogram of the variable and a regular main survey which is designed for interpolation. The spacing of the main survey is selected according to the variogram estimated from the reconnaissance survey and ensures that the kriging variance of the survey falls with a pre-specified tolerance. The number and location of sample points required to accurately estimate the appropriate sampling interval varies according to the actual variogram. The reconnaissance survey is split into several phases. After each phase the uncertainty associated with the estimate of the required sampling interval (which arises because of uncertainty in the estimated variogram) is assessed within a Bayesian framework. If the estimate is sufficiently precise the reconnaissance survey is halted and the main survey is designed and carried out. Otherwise another phase of the reconnaissance survey is designed within the Bayesian framework in order to minimize the uncertainty associated with the required sampling interval.

Results

The greater efficiency of each of the adaptive sampling approaches compared with conventional sample schemes is demonstrated upon simulated data sets. The size and configuration of an adaptive survey is seen to vary according to the variogram of the simulated data. A field system is implemented for an adaptive survey of moisture content over a field. The system consists of a hand-held PC running the adaptive algorithm connected to a GPS. The results of such a survey are presented.

Conclusions

In general the sampling intensity of soil maps is selected somewhat arbitrarily. The required sampling intensity varies between different variables depending upon quantities such as the variogram which are unknown when sampling commences. Our presentation demonstrates how geostatistical methods may be applied to design sample schemes which are suitable for the particular variable being measured. This removes the risk of a soil scientist completing a survey and either discovering that the sample scheme is insufficient or that costly oversampling has occurred. A number of new approaches of quantifying the uncertainty in a geostatistical survey are described.

References:

McBratney, A.B., Webster, R. and Burgess, T.M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalised variables. *Computers and Geosciences*, **7**, 331–334.

Van Groenigen, J.W. (1999). *Constrained optimisation of spatial sampling*: ITC Publication Series, v.65, Enschede.

Contact Information: Ben Marchant, Biomathematics & Bioinformatics Division, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, England, Phone: +44 (0)1582 763 133, Email: ben.marchant@bbsrc.ac.uk

Modeling In-Situ Soil Profile Evolution

Budiman Minasny[1], ***Alex. B. McBratney***[1] and *Sebastien Salvador Blanes*[2]
[1]Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW 2006, Australia
[2]Faculté des Sciences et Techniques, Parc de Grandmont, Tours, France

The ever increasing environmental problems have created a need for a better understanding of soil-landscape relationships. Quantitative modeling has been proposed and used as means to evaluate the long-term impact of human activities and environmental change on the soil and landscape. While many mechanistic models have been developed to simulate soil processes (eg. transport of water, solute, gas in soil) very few attempts have been made to simulate the development of a soil profile as a whole. Kirkby (1985) described a mechanistic model for soil profile evolution by linking weathering, organic matter decomposition and nutrient cycling. Legros and Pedro (1985) modeled the weathering of particle-size distribution in the profile. Here we present a model simulating in-situ soil profile development considering the physical and chemical weathering of the minerals.

The model is designed as follow: a given thickness of soil material (called layer) is released through physical disintegration of bedrock. For each time step, a layer of regolith is created. The regolith layer is then subjected to physical and chemical weathering. We follow the evolution of each layer which enabled us to calculate the particle size distribution and chemical; composition with time. The soil profile results in the vertical summation of all these layers.

1. Bedrock lowering, release of soil material

The bedrock surface lowering de_i at time step *i* and the resulting increase in soil thickness is represented by an exponential decline with the thickening of soil, as recently reviewed by Minasny and McBratney (1999), using the following equation:

$$de_i = P_0 \exp(-b.h_i)$$

where
de_i: bedrock lowering at time step i (m)
P_0 : potential weathering rate at h=0 (m $year^{-1}$)
b: empirical constant (m^{-1})
h_i: soil thickness at time step i (m)

2. The evolution of the coarse fraction

The evolution of the coarse fraction (>2 mm) of the soil through time is considered as a physical weathering process only. This coarse fraction is composed of an assemblage of rock fragments, *i.e.* pieces of rock that are an assemblage of minerals.

3. The evolution of the fine fraction

The fine fraction (<2 mm) of the soil is considered to weather both physically and chemically in various proportions according to the nature of the primary mineral. For this purpose, we consider that the fraction <2mm of each horizon is represented to occupy 1000 boxes corresponding to particle radii ranging from 1000 to 1μm for each primary mineral particle, that is considered to be spherical in shape. This is similar to the model of Legros and Pedro (1985).

a – Physical weathering

The physical weathering consists in breaking a given mineral particle into smaller particles. The model considers the physical breakdown of the particles into 2 smaller particles only. Physical breakdown is randomly happening for each given box: for each box, a random number is calculated. If this number is larger than a given chosen value of resistance to physical weathering, then the physical breakdown into two smaller pieces of random size occurs. Else, the physical breakdown is not happening.

b – Chemical weathering

The chemical weathering consists in calculating for each layer and each time step the quantity of a given primary mineral that is weathered and whenever it is the case the number of moles of secondary minerals formed according to known chemical weathering pathways. The primary mineral weathering rate in our model is calculated as a function of the reaction rate constant and the surface area of the mineral considered (White et al., 1996):

$$m_t = k_r.S_t$$

m_t: moles of primary silicate mineral dissolved (mole.m^{-2} soil)
k_r: reaction rate constant (mole.cm^{-2} mineral.y^{-1})
S_t: surface area of the mineral, varying as a function of time (cm^2 mineral. m^{-2} soil)

References:

Kirkby, M.J., 1985. A basis for soil profile modeling in a geomorphic context. J. Soil Sci. 36, 97-121.

Legros J.P., Pedro, G., 1985. The causes of particle size distribution in soil profiles derived from crystalline rocks. France. Geoderma 36, 15-25.

Minasny, B., McBratney, A. B., 1999. A rudimentary mechanistic model for soil production and landscape development. Geoderma 90, 3-21.

White, A.F., Blum, A.E., Schulz, M.S., Bullen, T.D., Harden, J.W., Peterson, M.L., 1996. Chemical weathering rates of a soil chronosequence on granitic alluvium: I. Quantification of mineralogical and surface area changes and calculation of primary silicate reaction rates. Geochimica et Cosmochimica Acta 60, 2533-2550.

Contact Information: Alex. McBratney, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney, NSW 2006, Australia,
Phone: +61 2 9351 3214, Fax: +61 2 9351 3706, Email: Alex.McBratney@acss.usyd.edu.au

Random Catena Sampling: For Establishing Soil-landscape Rules for Digital Soil Mapping

Alex. B. McBratney and *Budiman Minasny*
Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW, Australia

Survey methods for soil attributed and classes were presented at the inaugural workshop on digital soil mapping at Montpellier in September 2004. Methods presented were based on geostatistics (Heuvelink et al. 2004) and latin-hypercube sampling (Minasny & McBratney, 2004). During the discussion some of the more experienced field soil surveyors suggested that sampling methods that more explicitly explored soil-landscape relationships were appropriate sampling schemes for establishing mapping rules for digital soil mapping. No formal sampling method was suggested however. We make a first attempt at such a method here.

If we consider soil-landscape relationships to be catenary relationships we can formalise a procedure called *random catena sampling*. Using a digital elevation model, for a *k*-th order stream catchment, with the catchment boundary and streams defined, find a point (pixel) at random, trace all points uphill and downhill (to the stream) from this point using the steepest ascent/descent method, this set of points is a random catena. (All points belong to at least one such catena, points higher in the landscape have a larger probability of belonging to two or more catenas.) An alternative method of defining catenas would be the simple requirement that a neighbouring pixel be uphill or downhill (chosen randomly).

A subset of all possible catenas can be chosen by simple random sampling, stratified random sampling (strata being e.g., lithology or aspect) or Latin hypercube sampling. Positioning of observation locations along the catena seems more open. It would necessarily seem appropriate to sample the highest and lowest position in the transect. Other points could be sampled using equal intervals of horizontal, vertical or across-the ground distance (or some other metric).

An example of a random catena sampling design will be given.

References:

Heuvelink, G., et al. 2004. Opimisation of sample configurations for digital soil mapping with universal kriging. Workshop on Digital Soil Mapping. Montpellier.

Minasny, B & McBratney, A.B. 2004. Latin hypercube sampling as a tool for digital soil mapping. Workshop on Digital Soil Mapping. Montpellier.

Contact Information: Alex. McBratney, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney, NSW 2006, Australia,
Phone: +61 2 9351 3214, Fax: +61 2 9351 3706, Email: Alex.McBratney@acss.usyd.edu.au

Spatial Prediction Using BLUP with Matérn Covariance Function

Budiman Minasny and ***Alex. B. McBratney***
Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW, Australia

Stationarity of the variogram is usually not observed in soil data. A way out can be achieved by dividing the area of interest into sub-zones that are homogenous, each being modeled by different variograms (Goovaerts, 1997). Another solution is removing the trend from data. The trend in spatial data can be modeled either regional geographic trend (internal drift), or ancillary information (external drift). Modeling the trend also allows us to gain knowledge on the processes that gave rise to the spatial variation and improve predictions.

Another attraction is using ancillary information for the purpose of digital soil mapping. This is mainly due to the availability of the environmental data that can be obtained relative cheaply, and is available for the whole area, which is called exhaustive secondary data by Goovaerts (1997). The aim is to remove the trend, and enhance the prediction of soil properties by combining the prediction using environmental data and spatial interpolator.

The general linear spatial model in the presence of exhaustive ancillary information can be written as:

$$z(\mathbf{x}) = \mathbf{m}(\mathbf{x})^T \beta + \varepsilon(\mathbf{x}) \tag{1}$$

where **x** is the vector of spatial coordinates, observation vector $\mathbf{z} = \left[Z(x_1),\ldots,Z(x_n)\right]^T$, **m**(**x**) is the trend function, $\boldsymbol{\beta}$ is parameter vector for the trend, and ε is the error with a mean of zero and covariance structure **K**.

Methods for prediction in the presence of exhaustive ancillary information include:
(a) universal kriging or kriging with internal drift (Webster and Burgess, 1980) ,
(b) kriging with external drift (Goovaerts, 1997).
Universal kriging applied to a special situation where the exhaustive ancillary information is provided solely by spatial position. Both (a) and (b) have the same formulation, the trend and residuals are modeled in a system. There is a 'short cut' sub-optimal version so-called regression kriging, or kriging after detrending, where the trend function and its residuals are modeled separately (Knotters et al., 1995).

In this paper we look at a more statistically sound solution using best linear unbiased predictor (BLUP), where the parameter of the trend function and the covariance function are directly estimated from the data using restricted maximum likelihood method (REML).

Minasny and McBratney (2005) suggested the Matérn function as a general soil variation model for **K** :

$$K_{ij} = c_0\delta_{ij} + c_1\left[\frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{r}\right)^{\nu} K_{\nu}\left(\frac{h}{r}\right)\right] \tag{2}$$

with parameter vector $\boldsymbol{\theta} = [c_0, c_1, r, \nu]$. δ_{ij} is the Kronecker delta, c_0 is the nugget variance, and $c_0 + c_1$ is the sill variance, h is the separation distance, K_ν is a modified Bessel function of the second kind of order ν, Γ is the gamma function, r is the distance or 'range' parameter and ν is the 'smoothness' parameter which allows great flexibility for modeling the local spatial covariance. Minasny and McBratney (2005) showed that the Matérn function can describe the spatial structure of various soil properties.

The unknown parameters for BLUP are the parameters of the covariance function θ and trend function β. The covariance structure **K** refers to residuals, and they can only be obtained after fitting the trend, however we need **K** to fit the trend. Thus we need a method that can predict the parameters accurately and unbiasedly. Methods for estimating θ **and** β include:
(1) The crude, sub-optimal approach – regression kriging (RK), where β is estimated by fitting the trend function using least-squares and calculating the residuals of the trend function. The semivariance of the residual is then estimated using the method of moments and θ is estimated by fitting a covariance function to the empirical variogram.
(2) Maximum likelihood (ML) (Mardia and Marshall, 1984), where an initial β is estimated by least-squares, then θ is estimated by maximizing the log-likelihood of the generalized least squares.
(3) Restricted maximum likelihood (REML) (Stein, 1999), a more robust method is to estimate θ which does not depend on correct estimates of β.

The most statistically sound method is REML, which recently has gained attention for analysis of spatial crop yield data. Thus for prediction using BLUP comprised of two stages:
- estimate θ with REML by maximizing the log-likelihood function
- estimate β using generalised least squares solution with predicted θ.
- apply spatial prediction to unknown sites.

An application of BLUP for spatial prediction of soil properties using the Matérn function will be given.

References:

Burgess, T. M.; Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties. III Changing drift and universal kriging. Journal of Soil Science 31: 505-524.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford Univ. Press, New-York, 483pp.

Knotters, M., Brus, D.J., Oude Voshaar, J.H., 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderm a67, 227-246.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 72, 135–146.

Minasny, B., McBratney, A.B., 2005. The Matérn function as a general model for soil variograms. Geoderma (in press).

Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging, Springer, New York.

Contact Information: Alex. McBratney, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney, NSW 2006, Australia,
Phone: +61 2 9351 3214, Fax: +61 2 9351 3706, Email: Alex.McBratney@acss.usyd.edu.au

Using Remote Sensing Based Evapotranspiration (Et) Map to Assess Soil and Crop Yield Variability

U. Mishra and *D. Clay*
Plant Science Department, South Dakota State University, Brookings, SD, USA

Ever-increasing world population needs more food production. In many dry land situations water is the most limiting factor for crop yield. By understanding water budgets it may be possible to increase field productivity. It has been proposed that Remote Sensing can be used to assess yield and soil variability. The objective of this study was to examine the relationships between remote sensing based Et maps and soil and yield variability of corn (*Zea mays L.*).

The METRIC^{tm} model was used to estimate Et from a Landsat 7 image taken in August 2002. The Et flux was calculated as a residual of surface energy budget equation which is expressed as the energy consumed by the evaporation process. The Et flux is converted into Et by dividing by the latent heat of vaporization and expressed as a depth of water per unit of time. The crop yield was measured with a yield monitor and 92 Soil samples data were available from the study area. The relationship between Et and soil and crop yield variability will be assessed in GIS environment and the results will be presented at the conference.

Contact Information: U. Mishra, Plant science department, South Dakota State University Brookings, SD, 57007, USA, Phone: 605-688-4755, Email:Umakant.mishra@gmail.com

Digital Soil Mapping for a Tradeoff Analysis Application in Kenya

***Alejandra Mora-Vallejo**, Lieven Claessens* and *Jetse Stoorvogel*
Laboratory of Soil Science and Geology, Wageningen University, The Netherlands

Detailed information on land characteristics is an essential input for land use analysis and sustainable land use planning. Since traditional soil mapping is costly and time-consuming, large areas have soil information of poor quality and low resolution. This is especially true in developing countries where lack of infrastructure and expertise are common constraints. In this respect, the use of Digital Soil Mapping (DSM) techniques (McBratney et al., 2003) appears to be an interesting alternative to capture soil spatial variability in a rapid and cost-effective manner by combining observation data, auxiliary information and expert knowledge.

DSM was applied in the Machakos District (Kenya) as an input for the Tradeoff Analysis Model (TOA) (Stoorvogel et al., 2003). TOA is currently being applied for the integrated assessment of mixed farming systems and analysis of their impacts on the environment, particularly regarding soil fertility decline. Since TOA is a spatially explicit land use model, data on the spatial variation in land characteristics is required. However, the Machakos study area is 20,000 km^2 in size and soil data available are limited. Initially, soil inputs for the TOA model were derived from the 1:1,000,000 Exploratory Soil Map of Kenya (Sombroek et al., 1980) in combination with the Fertilizer Use Recommendation Program (MoA, 1987). This soil map describes the whole area based on only seven representative soil profiles. Of course this is an extreme simplification of actual soil variation. Therefore, it was necessary to produce a more detailed and accurate soil map using a more quantitative approach.

Soil spatial prediction was based on the concepts of the soil forming factors equation (Jenny, 1941) and soil-landscape relationships, combined with geo-statistical methods. Jenny's equation states that soil formation is a function of climate, organisms, relief, parent material and time. Capturing the spatial variation of these factors can provide a better understanding of soil variability. Through various analyses we got insight in the variation in parent material, land cover, topography and climate. An earlier analysis on the basis of the SOTER methodology (Engelen and Wen, 1995) yielded a combined geology and landform map on the basis of topographic maps (scale 1:50,000 or 1:100,000). Land cover was derived from satellite imagery on the Africover map. It was used to mask non agricultural areas and to further subdivide the agricultural area into intensive agriculture in the hilly terraced areas near Machakos and extensive agriculture and pastoralism in the lower dry areas from Makueni to the south. In terms of relief, the 90 m resolution DEM was analyzed with the LAPSUS model (Schoorl, 2002) in order to disaggregate soil units according to soil dynamics (erosion, deposition, stability) by means of a mechanistic approach. Finally, only two weather stations are present in the area. Weather data (temperature, precipitation and solar radiation) were interpolated with a mechanistic model for climate interpolation (Baigorria Paz, 2005).

Field work yielded 170 additional soil observations for the DSM exercise. The sampling sites were fairly regularly spread out over the area, while ensuring that each main geologic unit was sampled at least once. Soil observations and chemical analyses were correlated with the factors of soil formation. These correlations were used in a regression-kriging framework (Hengl et al., 2004) to predict soil properties such as soil organic matter and texture. These properties are required as inputs by the crop simulation models incorporated in TOA. The prediction accuracy was evaluated by cross-validation. The results show that the digital soil map is more accurate than the map derived from the 1:1,000,000 soil map.

Finally, the performance of the TOA model was tested for both the coarse initial soil map and the map derived using DSM techniques. The Tradeoff Analysis methodology allows for the ex-ante evaluation of agricultural policies and alternative management strategies through an integrated analysis of tradeoffs between economic and environmental indicators (Stoorvogel et al., 2001; Stoorvogel et al., 2003). TOA combines spatially explicit econometric production models with spatially explicit mechanistic biophysical models, and provides quantitative insight into the complex nature of agricultural systems and is thus beneficial to decision making by farmers and agricultural managers (Stoorvogel et al., 2003). Results show that the quality of the soil map is clearly reflected in the quality of the TOA outcomes and that high resolution data can be obtained for this specific land use analysis.

References:

Baigorria Paz, G.A. 2005. Climate interpolation for land resource and land use studies in mountainous regions. PhD, Wageningen University, Wageningen.

Engelen, V.v., and T.T. Wen. 1995. Global and National Soils and Terrain Digital Databases (SOTER): Procedures Manual, pp. 125. FAO, Rome.

Hengl, T., G.B.M. Heuvelink, and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75-93.

Jenny, H. 1941. Factors of Soil Formation, A System of Quantitative Pedology McGraw-Hill, New York.

McBratney, A.B., M.L. Mendonca Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3-52.

MoA. 1987. Fertilizer Use Recommendation Project (Phase I). Ministry of Agriculture, National Agricultural Laboratories, Nairobi.

Schoorl, J.M. 2002. Addressing the multi-scale lapsus of landscape : multi-scale landscape process modeling to support sustainable land use : a case study for the Lower Guadalhorce valley South Spain. PhD, Wageningen University, Wageningen.

Sombroek, W.G., H.M.H. Braun, and B.J.A. van der Pouw. 1980. Exploratory Soil Map and Agro-Climatic Zone Map of Kenya. Kenyan Soil Survey, MoA, Nairobi.

Stoorvogel, J., J.M. Antle, C.C. Crissman, and W. Bowen. 2001. The Tradeoff Analysis Model Version 3.1: A Policy Decission Support System for Agriculture (User Guide) Laboratory of Soil Science and Geology, Wageningen University, Wageningen.

Stoorvogel, J.J., J.M. Antle, C.C. Crissman, and W. Bowen. 2003. The tradeoff analysis model. Version 3.1: a policy decision support system for agriculture. Wageningen University., Wageningen.

Contact Information: Alejandra Mora-Vallejo. Laboratory of Soil Science and Geology, Wageningen University, P.O. Box 37, Wageningen 6700 AA, The Netherlands. Phone: +31 317 484445 Fax: +31 317 482419, Email: alejandra.moravallejo@wur.nl

Characterizing Soil Clay Content Profiles *In Situ* Using Visible-Near Infrared Spectroscopy

Cristine L. S. Morgan, *Travis Waiser* and *David J. Brown*
[1]Texas A&M University, Soil & Crop Sciences Department, College Station, Texas, USA
[2]Montana State University, Department of Land Resources and Environmental Sciences, Bozeman, Montana, USA

The soil system demonstrates a high degree of heterogeneity both vertically and spatially. Contemporary soil surveyors attempt to capture this variability by describing and characterizing soil profiles and interpolating between sampled locations using implicit or explicit soil-landscape models. However, given the time and cost of profile characterization, relatively few locations are sampled and the thus the accuracy of maps produced is insufficient for many applications. Rapid, inexpensive soil characterization methods are required. Recent research has shown the effectiveness of Visible-Near Infrared Diffuse Reflectance Spectroscopy (VNIR-DRS) to provide a non-destructive rapid prediction of soil physical, chemical, and biological properties of air-dried ground soil samples in the laboratory (Chang et al., 2001; Shepard and Walsh, 2002). However, the capability of VNIR-DRS for *in situ* soil analysis is less explored (Sudduth and Hummel, 1993).

The objective of our research was to evaluate the accuracy and precision of VNIR-derived predictions of soil clay content, organic carbon, and inorganic carbon, using scans of field-moist soil cores simulating *in situ* characterization. Seventy-two cores, from Central Texas, were scanned at field water content for both smeared and unsmeared cores, and air-dried water content for the unsmeared core. Water potential measurements were taken with a Decagon SC10 thermocouple psychrometer at the time of scanning. Once scanned, the soil samples were characterized for particle size analysis and organic and inorganic carbon. The spectral data were processed by averaging on 10 nm intervals and by taking the first and second derivative. To calibrate the model, 70 percent of the cores were chosen randomly. The remaining 30 percent were used for model validation. Partial least squares (PLS) regression analysis was used to model the correlation between soil properties and the first and second derivative of the reflectance. Preliminary results of measured versus predicted clay percentage yielded an r^2 of 0.84 and root mean squared deviation (RMSD) of 6% clay content (Figure 1). The results of this study indicate that VNIR-DRS may be useful in measuring some properties of soil profiles *in situ*, which would allow for quantifying soil properties at high vertical and horizontal resolutions quickly and more cost effectively.

Contact Information: Cristine Morgan, Soil & Crop Sciences Department, Texas A&M University, MS 2474 TAMU, College Station, TX 77843-2474, 979.845.3603, Email: cmorgan@ag.tamu.edu

Landscape Models of Claypan Soil Profile Properties as a Function of Divergence from Clay-Maximum Depth

***D. Brenton Myers*[1]**, *Newell R. Kitchen*[2], *Kenneth A. Sudduth*[2] and *E. John Sadler*[2]
[1]Soil, Environmental, and Atmospheric Sciences, University of Missouri, Columbia, MO, USA
[2]USDA-ARS-Cropping Systems and Water Quality Unit, Columbia, MO, USA

Applications in spatial modeling of landscape geomorphology, crop growth, and hydrology need spatially continuous estimators of soil profile properties. Clay distribution in soil profiles of NRCS-Major Land Resource Area 113 of the Central Claypan Region exhibits genetic control on the distribution of other physical, chemical, and hydrologic soil properties of the profile. A prominent argillic horizon in these soils, known as the claypan, has a peak clay content of 45-65%, occurring just below its upper boundary. This study was conducted to develop predictive models of soil properties in claypan soil landscapes based on the depth of the profile-maximum clay content. Soil samples were taken from profiles distributed across a claypan landscape. Divergence from clay-maximum depth (DCMD) of a soil sample was calculated as its height above or below the depth of the maximum clay content for its source profile. Several soil properties including textural components, Ca, Mg, K, Na, Al, P, organic matter, buffer pH, and calculated values of plant available water capacity and saturated hydraulic conductivity were modeled as both parametric and non-parametric functions of DCMD. Georeferenced ground conductivity measurements were used to spatially predict depth to clay-maximum. These predictions were, in turn, used to develop three-dimensional maps of soil property profiles in agricultural fields based on the DCMD functions. Divergence from clay-maximum depth effectively centered profiles from across claypan landscapes onto a single scale. Both parametric and non-parametric functions accurately modeled the DCMD profile of soil properties. Continuous three-dimensional representations of claypan soil landscapes suitable for crop and hydrology modeling were developed. The DCMD profiles of soil properties were useful for differentiating between surface alterations of soil properties and their genetic properties. These results establish that prominent genetic horizons which exhibit control on profile properties are useful as normalizing features for the spatial prediction of soil property profiles.

Contact Information: D. Brenton Myers, Soil, Environmental and Atmospheric Sciences, University of Missouri, 269 Agricultural Engineering Building, Columbia, MO, 65211, USA, Phone: (573) 446-6856,
Email: myersdb@missouri.edu

Using Optically Stimulated Luminescence Dating for Estimating Soil Formation Rates

Asger Nielsen[1,2]*, Bo Elberling*[1]*, Morten Pejrup*[1] and *Andrew S. Murray*[2]
[1]Institute of Geography, University of Copenhagen, Copenhagen, Denmark
[2]Nordic Laboratory for Luminescence Dating, Risø National Laboratory, Roskilde, Denmark

Using a CLORPT-approach (Jenny, 1941), soil chronosequences (sequences of soils having equal conditions except soil age) allow the study of the effects of soil development through the entire time period covered by the sequence, yielding a rare opportunity to assess the local soil development rates. As the chronosequences contain inherent uncertainties, the dating of the sequences needs to be very precise, otherwise the results will be very questionable.

Previous soil chronosequence studies have applied a wide range of dating methods, most commonly radiocarbon (^{14}C) dating, usually with error ranges less than +/- 10% (Trumbore, 2000). However, ^{14}C dating has several major shortcomings, mainly that the method obviously needs carbon in appreciable amounts, excluding many mineral soils (ibid.). In addition, ^{14}C dating of bulk soil organic material is very unreliable, as the resulting age is roughly equal to the average carbon residence time in the soil and not necessarily equal to the sediment age (ibid.). This forces many studies to apply radiocarbon dating to nearby proxies, such as peat bogs, instead of dating the sequence soils directly.

Based on these experiences, this study dated 28 samples from a Holocene chronosequence (a beach ridge plain located north of Frederikshavn, northern Jutland, Denmark), using quartz-based Optically Stimulated Luminescence (OSL) dating and applying a standard Single Aliquot Regenerative-dose (SAR) protocol (Murray & Wintle, 2000; 2003). The results from this study show that OSL dating is very suitable for dating chronosequences in sandy, coastal marine soils, displaying high internal consistiency and good agreement with external, independent age controls.

Subsequently, the OSL ages were combined with selected soil chemistry data from the chronosequence; thus quantifying the soil development (primarily pH, carbon stocks, carbon residence times and extractable iron/aluminium fractions) on the basis of 120 soil samples collected from 12 soil profiles in four major groups. In all of the profiles, seperate samples were OSL-dated. This allowed the construction of a number of chronofunctions, expressing the soil development parameters mentioned above as a function of time. For instance, the CBD (Sodium citrate – sodium hydrogen bicarbonate – sodium dithionite) extractable aluminium stock increased linearly ($r^2 = 0.99$) with 0.5 kg/m^2/1000 years during the ~4000 years covered by the chronosequence.

Depending on the criteria, fully developed podzols were developed in 2000-4000 years, and proto-podzols were formed within 1700 years. These podzol development times are intermediate compared to podzol development times from other studies; ranging from 350 years (Singleton & Lavkulich, 1987) to 7000 years (Ewing & Nader, 2002). The differences in podzol development times are largely explained by different climates among the various field sites.

Whereas soil formation times have been reported in numerous studies, soil formation times based on OSL dating (dating the quartz grains in the soil *directly* instead of using various proxies) have not been presented in previous studies, to the knowledge of the present authors. Hence, this study aims to demonstrate that OSL dating holds a large, unused potential for aiding soil chronosequence studies by allowing a direct dating of the soils in the chronosequence. This will almost certainly enhance the reliability of the derived chronofunctions.

References:

Ewing, H.A. & Nader, E.A. (2002): Holocene Soil Development on Till and Outwash Inferred from Lake-Sediment Geochemistry in Michigan and Wisconsin. Pp. 234-43 in: Quaternary Research, vol. 57.

Jenny, H. (1941): Factors of Soil Formation. A System of Quantitative Pedology. New York, USA: McGraw-Hill Book Company, Inc.

Murray, A.S. & Wintle, A.G. (2000): Luminescence dating of quartz using an improved single-aliquot regenerative-dose protocol. Pp. 55-73 in: Radiation Measurements, vol.32.

Murray, A.S. & Wintle, A.G. (2003): The single aliquot regenerative dose protocol: potential for improvements in reliability. Pp. 377-81 in: Radiation Measurements, vol. 37.

Singleton, G.A.; Lavkulich, L.M. (1987): A soil chronosequence on beach sands, Vancouver Island, British Columbia. Pp. 795-810 in: Canadian Journal of Soil Science, vol. 67.

Trumbore, S.E. (2000): Radiocarbon Geochronology. Pp. 41-60 in: Noller, J.S., Sowers, J.M., Lettis, W.R. (Eds.), Quaternary Geochronology. American Geophysical Union, Washington DC, USA.

Contact Information: Asger Nielsen, University of Copenhagen, Institute of Geography, Oester Voldgade 10, Copenhagen, DK-1350, Denmark, Phone: (+45) 35 32 41 80, Fax: (+45) 35 32 25 01, Email: asgern@yahoo.dk

Reducing the Cost of Accurate Soil Mapping by Using Maximum Likelihood Variograms for Prediction

M. A. Oliver[1] and *R. Kerry*[2]
[1]Department of Soil Science, University of Reading, Reading, England
[2]Department of Geography, Brigham Young University, Provo, Utah, USA

Many soil science applications require accurate contour maps of several soil properties for example, for land management and for dealing with contaminated sites. Viscarra-Rossel & McBratney (1998) stated that soil contour maps are best produced by grid sampling followed by geostatistical analysis based on computing and modeling the usual method of moments variogram (MoM) and kriging. In many situations, the need for a minimum sample size of 100 sites at an appropriate interval to compute a reliable MoM variogram is either prohibitively expensive or impractical. An appropriate sampling interval is one that resolves the spatial variation at the level of interest, e.g. drainage basin, agricultural field or trial plot. Kerry and Oliver (2003) showed that variograms from certain kinds of ancillary data, such as aerial photographs of bare soil and apparent soil electrical conductivity data, had similar ranges to variograms of the more permanent soil properties, for example particle size distribution, soil depth, etc. Therefore, such ancillary data could be used to determine an appropriate sampling interval and so dispense with the need for a reconnaissance survey. However, if the sampling interval suggested in this way is large in comparison to the extent of the area of interest, only a small sample size would be needed to resolve the spatial variation. This would mean that either the MoM variogram might be a poor indication of the true underlying variogram, or that a larger sample size would be needed regardless of the nature of the spatial variation.

A possible solution to this problem is to use the Maximum Likelihood (ML) variogram to describe the variation and for prediction. Pardo-Igúzquiza (1998) stated that, “a few dozen data may suffice” to estimate the ML variogram. However, he did not provide a rational strategy for deciding how to sample to estimate the ML variogram. To assess the validity of Pardo-Igúzquiza’s statement and different sampling schemes we computed the ML variogram for large and small sets of data and for different sampling configurations. We compared them with those for the associated MoM variograms.

The top-soil was sampled on 20-m and 30-m grids at four study sites on different parent materials in southern England. The percentage clay content of bulked soil samples was determined using laser granulometer methods. The form of MoM and ML variograms calculated using the 20-m and 30-m data (100 or more data points) were compared together with the mean squared errors (MSEs) and mean squared deviation ratios (MSDRs) produced by each during cross-validation. The form of both types of variogram, the MSEs and MSDRs were comparable. Variograms from ancillary data were used to determine an appropriate sampling interval at each site. The original data were sub-sampled to produce data with sampling intervals similar to those suggested by the variograms of ancillary data (a). The data were also sub-sampled to produce data sets with 50 points at the original sampling interval (i.e. in a section of the field), (b). In this way the effect of placing the sampling sites more intensively in one part of the field could be compared with a more general less intensive distribution of sites. Finally, sampling was at an interval similar to that suggested by the variograms of ancillary data and supplemented with sites at half that interval until a sample size of 50 was reached (c). The ML and MoM variograms for each sub-sample of data were estimated and the parameters used for cross-validation. The forms of the ML and MoM variograms of the various sub-sampled data were compared with the MoM variogram computed using 100 or more data points.

As the number of sites in a sub-sample decreased, the form of both MoM and ML variograms became more different from those of MoM variograms computed using 100 or more data points; however, the MoM variograms from the sub-sampled data were always more different from the original variogram. The ML variograms computed from sub-sample (c) were similar to the MoM variograms computed from 100 or more data. Cross-validation showed that those variograms whose model parameters were similar to those of the MoM variogram of 100 or more data were more appropriate for kriging. Lark (2000) has also shown that the ML variograms for one site computed from 60 observations were similar to those for 90 to 120 sites. However, for another set of data the performance of the variogram estimators depended on the nature of the spatial variation. We maintain that provided the 50 sites are at an appropriate interval and are supplemented by others at shorter intervals, the variogram will be a more reliable indication of the true underlying variogram. The additional sites also ensure that more of the short-scale variation is resolved and that the nugget variance is minimized.

The results of our analysis show that an optimal approach to soil mapping would be to: (1) determine an appropriate sampling interval using a variogram from suitable ancillary data, (2) sample the field at that appropriate interval, and (3) take additional samples at random at half that interval to form a data set of 50 points. This approach reduces by half the number of samples needed to estimate a reliable variogram that can be used for both interpretation and prediction. Therefore the costs of producing accurate soil contour maps could be reduced markedly. This is important as funds for soil sampling and subsequent laboratory analyses are limited, yet there is a growing demand for more accurate predictions for environmental and land management initiatives.

References:

Kerry, R. & Oliver, M.A. 2003. Variograms of ancillary data to aid sampling for soil surveys. *Precision Agriculture*, **4**, 253-270.

Lark, R. M. 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science,* **51**, 717-728.

Pardo-Igúzquiza, E. 1998. Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, **30**, 95-107.

Viscarra-Rossel, R.A. & McBratney, A.B. 1998. Soil chemical analytical accuracy and costs: implications from precision agriculture. *Australian Journal of Experimental Agriculture*, **38**, 765-775.

Contact Information: M. A. Oliver, Reading University, Soil Science Department, PO Box 233, Reading Berkshire RG6 6DW, England, Phone: +44 118931 6557, Fax: +44118931 6660, Email: m.a.oliver@reading.ac.uk

The Assessment of Point and Diffuse Soil Pollution from an Urban Geochemical Survey of Sheffield, England

***Barry G. Rawlins*[1]**, *R. Murray Lark*[2], *Kirsten E. O'Donnell*[1], *Andy M. Tye*[1] and *T. Bob Lister*[1]
[1]British Geological Survey, Keyworth, Nottingham, Nottinghamshire, UK
[2]Rothamsted Research, Harpenden, Hertfordshire, UK

There is increasing interest in the quality and management of urban soils, including the implications of elevated concentrations of metals in soil. Data on metal concentrations in urban soil typically show very complex variation. In particular there are spatially continuous variations, arising from parent material and the effects of diffuse pollution. The identification of spatial outliers in the data may give insight into sources and processes of pollution. Once identified, further interpretation of the source of the contamination can be aided by assessment of the soil samples and historical map data. This would include the visual appearance of the soil (such as the presence of any unusual particles or material), and historical map data (including information on land use) that show the history and development of the area. This may provide information on the point processes that have influenced soil chemistry. In addition, using estimates of the geochemical composition of the soil from a survey in adjacent rural and peri-urban areas over the same parent material types, it may be possible to estimate the magnitude of the diffuse contribution to the urban soil after removal of outliers.

A model of soil variability as a continuous background process with superimposed point contamination was applied to 569 measurements of metal concentrations (Cr, Ni and Pb) in the topsoil of Sheffield. The city of Sheffield has a long-history of coal-mining, metal manufacture and processing industries, resulting in both point and diffuse pollution of a range of metals, including Cr, Pb and Ni. Robust estimators of the variogram were shown to be required to describe spatial variation of the metal concentrations at most sampled locations. This is diagnostic of the presence of a contaminating process. Values of the standardised kriging error (SKE) from the cross-validation of each datum were used to identify spatial outliers for each metal. Geochemical maps were prepared by kriging after removing the outliers to estimate the background variation.

Each of the 35 spatial outliers that occurred in gardens had concentrations exceeding their soil guideline value for residential land use with plant uptake, highlighting a potentially significant exposure pathway. The frequent observation of coal and furnace waste at these sites suggests that their dispersal following domestic use and industrial processes respectively represents a significant point contaminant process. There was no evidence for spatial clustering of the point process. However, the spatial outliers of Cr and Ni show a significant association with disturbed sites identified from historical land use maps, in part due to their prevalence in areas of historical steel manufacture. The magnitude of diffuse pollution for each metal in the urban soil was estimated by removing the spatial outliers and comparing robust measures of location with those from a survey of soils developed over the same parent materials in adjacent rural and peri-urban environments. The Winsorized mean Pb concentrations in urban topsoil (203 mg kg^{-1}) were twice the value in the rural environment (101 mg kg^{-1}), highlighting a very substantial diffuse Pb load to urban soils. The equivalent estimated diffuse component in urban soils for Cr and Ni were, respectively, 25 % and 14 % higher than the rural soils. The implications of some of these findings are discussed.

Contact Information: Barry G. Rawlins, British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK, Phone: 44-115-9363140, Fax: 44-115-9363200, Email: bgr@bgs.ac.uk

Use of Soil Survey Information for Determining Soil Hydraulic Properties

Walter J. Rawls[1] and *Yakov A. Pachepsky*[2]
[1]USDA-ARS Hydrology And Remote Sensing Laboratory, Beltsville, MD 20705 USA,
[2]USDA-ARS Environmental Microbial Safety Laboratory, Beltsville, MD 20705 USA

Soil hydraulic properties are a key in modeling hydrologic processes. At present, average soil hydraulic properties developed in the early 1980's according to soil texture are being used in hydrologic models. Laboratory and field methods for determining soil hydraulic properties are time consuming and expensive. Indirect estimation of soil hydraulic parameters from readily available or easily measurable soil data has become critical for hydrologic modeling. The term pedotransfer function is often used to describe equations expressing dependencies of soil water retention and soil hydraulic conductivity on basic soil parameters. Most Pedotransfer function development has primarily been directed toward finding the best measured soil physical properties for determining a particular soil hydraulic property. Usually pedotransfer functions have been developed to incorporate various levels of soil information. Current research issues in pedotransfer function development include (a) searching for better mathematical expressions of pedotransfer function equations, (b) searching for the most affluent basic soil parameters to be used as pedotransfer function inputs, and (c) searching for ways to group soils to have more accurate pedotransfer functions for each separate group. To date no one has developed pedotransfer functions which incorporate the field soil survey information. There are thousands of soil survey descriptions available in the United States and soil survey description are easily made in the field; thus using this information will make the pedotransfer functions more universally usable. The objective of this study was to use regression tree analysis to determine the most important soil survey information for estimating soil water content θ_{33} at matric potential -33 kPa and soil water content θ_{1500} at matric potential -1500 kPa and subsequently using this information to predict the soil hydraulic properties. Pedotransfer functions derived to estimate soil water content θ_{33} at matric potential -33 kPa and soil water content θ_{1500} at matric potential -1500 kPa can be used to determine the water retention curve and saturated hydraulic conductivity.

The U.S. Natural Resource Conservation Service (NRCS) soil characterization data base which includes measured basic physical and chemical properties, full genetic characterization, and soil water content at several given potentials was used for the analysis. The NRCS database includes over 20,000 pedons. Soil survey descriptions include descriptive terminology, class definitions, hierarchical soil groupings, and operations that are applicable to various scales and appropriate to a wide variety of uses. After reviewing the soil survey information, the field determined taxonomic unit, soil texture, horizon, structure, consistency, plasticity, stickiness, and land surface configuration were thought to be the available information which would have the greatest effect on hydraulic soil properties.

Regression tree modeling was applied because it can use both categorical and numerical variables as predictors. Regression tree modeling is an exploratory technique based on uncovering structure in data. The resulting model partitions data first into two groups, then into four groups and so on providing groups as homogenous as possible at each of the levels of partitioning. Each partitioning can be viewed as a branching and the final fit of the model to data looks like a tree with two branches originating from each node. The model can be used visually or easily computerized. Another benefit of regression tree modeling is that grouping field textural classes has an advantage of lesser demand to the accuracy of the field soil texture determination which is misjudged in the field about 50 % of the time.

The analysis of all the three data sets indicated that the field texture classes was the most important property for estimating soil water content, θ_{33} , at matric potential -33 kPa and soil water content, $\theta_{1500,}$ at matric potential -1500 kPa. . When combined with field soil texture classes organic carbon added the largest increase in accuracy in the prediction of θ_{33}; followed by topography for A horizon and taxonomic order. The largest increase in accuracy in the prediction of θ_{1500}. was obtained when both taxonomic order and organic carbon or topography and horizon were used with field soil texture classes. The most important topographic predictors were slope %, slope position and slope shape across the slope. The most important structure class was shape, while plasticity and stickness and dry consistency were more important than structure. Organic carbon was important.

Field texture classes was the most important property for estimating soil water content ,θ_{33} , at matric potential -33 kPa and soil water content, $\theta_{1500,}$ at matric potential -1500 kPa.. Combining field texture classes into groups increased the accuracy. Water retention at -33 kPa is affected by the organic carbon more strongly than water retention at -1500 kPa. Water retention of soils with coarse texture is substantially more sensitive to the amount of organic carbon as compared with fine-textured soils. Using soil structural and consistence parameters along with textural classes provides a small, although significant improvement in accuracy of water retention estimates as compared with estimation from texture alone. Soil structural and consistence parameters can serve as predictors of soil water retention because those parameters reflect soil basic properties that affect soil hydraulic properties. Using the topotextural classes led to a statistically significant but small improvement in the accuracy of the water retention estimates. Use of topotextural information allows pedotransfer functions to transcend scales.

Contact Information: Walter j. Rawls, USDA-ARS Hydrology And Remote Sensing Laboratory, Bldg. 007, Rm. 104, BARC-W, Beltsville, MD 20705 USA, Phone: 301 504-8745, Fax: 301504-8931, Email: wrawls@hydrolab.arsusda.gov

Incorporation of ASTER Satellite Imagery into Multi-Variate Geostatistical Models To Predict Soil Phosphorus

Rosanna G. Rivero[1,2], *Sabine Grunwald*[2] , *Susan Newman*[3], *Todd Z. Osborne*[2] and *K. Ramesh Reddy*[2]

[1]Department of Urban and Regional Planning, University of Florida, Gainesville, Florida, USA
[2]Soil and Water Science Department, University of Florida, Gainesville, Florida, USA
[3]South Florida Water Management District, West Palm Beach, FL

Geostatistical methods to predict soil properties using high-quality geospatial datasets have been used for some time. But it has not been until the last decade, with advances in geographic information systems and remote sensing, that these methods have been improved to incorporate secondary or ancillary environmental variables into the mapping of soil properties. These have been defined as “hybrid interpolation techniques”.

The objective of this study was to compare the model performance of traditional and hybrid modeling techniques to predict soil total phosphorus (TP) in Water Conservation Area-2A (WCA-2A) that is part of the Greater Everglades ecosystem. In this area, human-induced nutrient inputs have caused shifts in water and soil chemistry that have been linked to the expansion of vegetation species such as cattail (*Typha sp.*). Numerous studies have documented the P-enrichment in WCA-2A, as a result of nutrient influx from canals and inflow structures from adjacent areas including the Everglades Agricultural Area; but none of them attempted to characterize the spatial variability and distribution of TP using remote sensing data.

In 2003, soil samples at 111 sites spatially-distributed throughout WCA-2A were collected at 0-10 cm depth and analyzed for a suite of soil physico-chemical properties including TP. We used reflectance values from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) NASA satellite imagery with 15 meter spatial resolution to derive landscape indices (e.g. Normalized Difference Vegetation Index - NDVI) as proxies to map vegetative structures such as tree islands, *Typha* sp. and *Cladium Jamaicense* (sawgrass). We employed log-normal kriging (LNK), multi-Gaussian kriging (MGK) co-kriging (CK) and regression kriging (RK) to predict TP at unsampled locations. The semivariogram of log TP was fitted using an exponential model with a range of 7,470 m, sill of 0.05 and nugget of 0.01. Normal–score transform was used to transform TP values into Gaussian format to conduct MGK. We used heterotopic CK using the TP observations as primary variable that was known at few points and ASTER reflectance data as auxiliary (secondary) variable known everywhere in the study domain. Regression kriging was implemented by mixing regression and kriging methods. First, a regression model was derived relating TP to NDVI. Residuals of TP were derived, kriged and then added back to the trend regression model to generate the final prediction map. To assess model performance we used standard error metrics and the E-coefficient.

Both multi-variate methods, CK and RK, performed better than LNK and MGK. The ancillary environmental data derived from ASTER satellite imagery provided exhaustive coverage to delineate vegetation properties that were correlated with soil TP. Incorporating sparsely sampled soil and dense, pixel-based remote sensing datasets is ideal to improve the prediction of soil properties across wetland ecosystems. Accurate soil predication maps that characterize the current status in nutrient enrichment are a prerequisite to support the ongoing restoration efforts in the Greater Everglades ecosystem.

Acknowledgement: Data were collected and analyzed by the South Florida Water Management District (SFWMD) and the Wetland Biogeochemistry Laboratory, Soil and Water Science Department, University of Florida (K.R. Reddy, S. Newman, T.Z. Osborne, and Y. Wang).

Contact Information: Rosanna Rivero, Department of Soil and Water Science, GIS Laboratory, McCarty Hall A 1184, Soil and Water Science Department, University of Florida, Fax: 352-392-3902,
Phone: 352-392-1951 ext. 233, Email: RGRivero@ifas.ufl.edu

Detecting Residual Pyrite after the Aznalcóllar Mine Spill (SW Spain) Using Electromagnetic Soil Conductivity Data

K. Vanderlinden[1], *R. Ordóñez*[2], *J. V. Giráldez*[3] and *J. L. Muriel*[1]
[1]Organic Farming and Natural Resources Area, IFAPA, CIFA "Las Torres-Tomejil", Alcalá del Río, Spain.
[2]Organic Farming and Natural Resources Area, IFAPA, CIFA "Alameda del Obispo", Córdoba, Spain.
[3]Dept. of Agronomy, University of Córdoba and IAS-CSIC, Córdoba, Spain.

The accurate spatial delineation of polluted soils is essential to successful decision making in relation to remediation activities or risk assessment. The financial cost of the soil data acquisition is often prohibitive for achieving estimates with acceptable uncertainty levels, especially when the polluted areas are highly spatially variable.

The use of electromagnetic (EM) soil sensing techniques might alleviate these inconveniences if there exist some direct or indirect, relationships between soil electrical conductivity and the pollution. EM instruments measure the soil bulk electrical conductivity (EC_b) *in situ* without making any physical contact with the soil and provide a simple means for fast and detailed mapping of EC_b when equipped with a DGPS. The EM38, an electromagnetic induction device that was developed especially for agronomic applications, has been used during the last decades, first for mapping soil salinity, and later on also for the spatial characterization of soil properties within the framework of precision agriculture (Corwin and Lesch, 2003).

In this study we used the EM38 to detect hazardous spots in a 9 ha area, polluted by pyritic mining sludge. The site is situated along the Guadiamar River, approximately 8 km downstream from the Aznalcóllar pyrite mine (SW Spain) that collapsed six years ago and that released approximately 6 Mm^3 of acid water and sludge into the Gauadiamar basin (Grimalt et al., 1999). Shortly after the spillage a cleanup program was set up to remove the toxic residue from the riverbed and banks. Subsequently, after applying limestone, the cleaned soil was tilled to mix the remaining sludge with the topsoil layer.

From 1999 on, several soil chemical properties of the reclaimed land, on which remained a small - but highly spatially variable - quantity of pyritic sludge mixed with the topsoil layer, have been monitored. Currently, despite of the repeated application of lime and sugar beet foam, pH values below 3 are still observed in the polluted spots. This persistent acidification is caused by the oxidation of pyrite (Fe_2S) or other sulfuric minerals and enhances mobility and availability of most of the present metals. Common minerals that result from pyrite weathering under semiarid or arid conditions are sulfate salts like gypsum or sodium sulfate. The use of EC_b data to detect polluted zones in the study area was motivated by the presence of these salts and the comparatively larger soil moisture content of the polluted spots.

The EC_b maps of the area show a spatial pattern, similar to the one observed for the soil chemical parameters. A comparison with the observed data confirms the relationship. Not only the spots with remaining sludge are detected. Also fine and coarse textured zones in the area are accurately delineated. This information is then used to evaluate the local risk associated with the residual pollution.

Finally, the performance of several geostatistical methods is compared for the spatial estimation of pH and Zn, using classified EC_b maps as secondary information (Triantafilis et al., 2001).

References:

Corwin, D.L. and S.M. Lesch. 2003. Application of soil electrical conductivity to Precision Agriculture: Theory, principles, and guidelines. Agron. J. 95:455-471.

Grimalt, J.O., M. Ferrer, and E. Macpherson. 1999. The mine tailing accident in Aznalcóllar. Science of the Total Environment 242:3-11.

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65:869-878.

Contact Information: Karl Vanderlinden, Organic Farming and Natural Resources Area, CIFA "Las Torres-Tomejil", IFAPA, Ctra. Sevilla-Cazalla, km 12,2, 41200 Alcalá del Río, Spain,
Email: karl.vanderlinden.ext@juntadeandalucia.es

Data Analysis of Hyperspectral VNIR Sensing for the Assessment of Soil Variability

Harold M. van Es*, Stephen D. DeGloria, Thomas Owiyo, A. Volkan Bilgili, H. Dean Hively* and *Deborah G. Grantham*
Department of Crop and Soil Sciences, Cornell University, Ithaca, NY, USA.

Reflectance spectroscopy has been adopted as a nondestructive method for material characterization in a wide range of applications. Fundamental features of chemical components (e.g., vibrational bands of H-C, H-N and H-O) often occur in the thermal infrared range, but their overtones (one half, one third of the wavelength) are often expressed in the near infrared (NIR), which is more readily measurable. Hyperspectral soil sensing yields large data sets that can be used to find correlations with measured soil properties. Raw data, as well as first and second-derivatives each provide valuable information that can be analysed separately or combined by traditional statistical methods or data mining techniques. The latter methods are of increasing interest because they allow for the rapid analysis of very large data sets and may provide superior results. In this work, we determined the feasibility of using hyperspectral sensing for rapid assessment of a wide range of soil characteristics in different locations, including New York, Kenya, Turkey and Bangladesh. Various data analysis methods were evaluated and compared.

Contact Information: Harold M. van Es, Department of Crop and Soil Sciences, 1005 Bradfield Hall, Cornell University, Ithaca, NY, 14853-1901, USA. Phone: 607-255-5629, Fax: 607-255-3207, Email: hmv1@cornell.edu.

Spatially-Balanced Experimental Designs for Field Experiments

Harold M. van Es[1], *Carla P. Gomes*[2,3], *Meinolf Sellmann*[2] and *Cindy L. van Es*[3]
[1]Department of Crop and Soil Sciences, Cornell University, Ithaca, NY, USA
[2]Department of Computing and Information Sciences, Cornell University, Ithaca, NY, USA
[3]Department of Applied Economics and Management, Cornell University, Ithaca, NY, USA

Spatial heterogeneity in fields may affect the outcome of experiments. The conventional randomized allocation of treatments to plots can cause imprecision and bias in the presence of trends (including periodicity) and spatial autocorrelation. Agricultural scientists appear to mostly use conventional experimental designs that are susceptible to adverse affects from field variability. The objectives of this research were to (i) quantify the use of different experimental designs in agronomic field experiments, and (ii) develop spatially-balanced designs that are insensitive to the effects of trends and spatial autocorrelation. A review was performed of all research efforts reported in Volumes 93-95 of the Agronomy Journal and the frequency of various experimental designs was determined. The method of simulated annealing was used to develop Spatially-Balanced Complete Block Designs (SBCBD) based on two objective functions: promoting spatial balance among treatment contrasts, and disallowing treatments to occur in the same position in different blocks, when possible. The journal review showed that the vast majority (96.7%) of agronomic field experiments are implemented through Randomized Complete Block Designs. SBCBDs were successfully developed for designs up to 15 treatments and 15 replications. Square SBCBDs were realized as Latin Squares, and perfect spatial balance was obtained when feasible. SBCBD's are simple to implement and provide protection against the adverse effects of spatial heterogeneity, while randomized allocation of treatments still insures against user bias.

Contact Information: Harold M. van Es, Department of Crop and Soil Sciences, 1005 Bradfield Hall, Cornell University, Ithaca, NY, 14853-1901, USA. Phone: 607-255-5629, Fax: 607-255-3207, Email: hmv1@cornell.edu

Pedometrics in Transition: From Too Few to Too Many Data?

Marc Van Meirvenne, *Liesbet Cockx* and *Udayakantha Vitharana*
Department of Soil Management and Soil Care, Ghent University, Gent, Belgium

Most of the pedometrical applications reported in literature and at previous Pedometrics conferences deal with soil variability in space or time. The rational behind this focus is that soil is a continuously varying mantle which is difficult to observe directly, either by eye or through sampling. Hence indirect sources like topography, landscape position and natural vegetation have been used frequently to provide information on expected soil composition. Until recently, the usefulness of these sources was limited by a lack of detail, e.g. contour lines served as a basis for the construction of elevation models but had wide contour intervals, or satellite images had coarse resolutions. Consequently, soil sampling remained an essential primary source of information. But since soil sampling is hard labour and causes large expenses, the objective was to take as few as possible samples. First, only a few representative samples were used to characterize entire soil mapping units, but more recently more intensive sampling strategies became necessary to serve geostatistical mapping purposes. But still, the lack of sufficient hard data needed to be compensated by the “best” possible interpolation methodology, since the fewer the data the more important is the theoretical basis of the interpolation. This explains the success of geostatistics at the first Pedometrics conferences.

At present, after a decade of GIS technology, we face a different situation. Soil properties remain difficult to observe directly through sampling, but other technologies have evolved tremendously. Moreover, GPS technology allows accurate positioning or navigation, even on a centimetre scale, cutting down time investments in the geographic aspects of field work.

The resolution of satellite imagery continuous to decrease, but most of this information is still related to vegetation and therefore not directly usable in soil mapping or monitoring.

Laser altimetry and RTK-GPS provide methods to scan accurately the land surface and map its elevation with centimetre accuracy. Therefore one can expect a strong increase in availability of detailed elevation data and investigations about soil-landscape relationships will benefit strongly from this.

In agronomic applications an increasingly available source of indirect soil information are yield maps. But these data still contain quite some noise and can have a large temporal variability between years since crop growth and yield is influenced by many factors (climate, fertilization,…).

A very interesting technology is soil sensing. Research efforts continue and several sensors have been developed, but to date the most promising are related to measuring soil electrical properties (mostly electrical conductivity or resistivity). These properties can be observed at driving speed and are related to a number of key soil variables like texture, moisture content, organic matter content,... Large volumes of stable soil-relevant observations can be obtained in a short period of time and their availability increases steadily.

As a consequence, the past focus on the development of methods for mapping soil properties with small numbers of data evolve in a need for methods to filter and clean up large volumes of different kinds of sensed data and combine several sources of exhaustive information. The data shortage problem has become a data wealth problem. Evidently data processing techniques must adapt to this new situation.

This contribution aims to analyse these new directions and challenges of Pedometrics. Examples will be taken from soil studies related to organic farming, precision agriculture, forestry, soil pollution, and the investigation of past quaternary environments.

Contact Information: Marc Van Meirvenne, Dept. Soil Management and Soil Care, Faculty of Bioscience Engineering, Ghent University, Coupure 653, 9000 Gent, Belgium, Phone: +32-92646056, Fax: +32-92646247, Email: marc.vanmeirvenne@ugent.be

Using Topographical and Geological Information in Modeling the Spatial Variation of Soil Attributes: A Case Study from Burundi

Anicet Sindayihebura[1], **Marc Van Meirvenne**[1] and *Stanislas Nsabimana*[2]
[1]Department of Soil Management and Soil Care, Gent University, Belgium
[2]Department of Geography, University of Burundi, Burundi

The Study area

The study area is located in Burundi between 29°41’22” and 29°45’16” eastern longitudes and 3°54’24” and 3°56’35” southern latitudes. The area is 14.8 km^2 and is located in the mountainous Central Plateau region, with altitudes ranging from 1800 m to 2100 m. The main geological feature is an alternation between sandstones and shales along the E-W direction. Dyke-like porphyric granite oriented to NNE intrudes the sedimentary rocks. This zone was found to be representative of geological and geomorphological complexities in Burundi. The study area was matched to the extent of two contiguous catchments in order to account for water flow processes (Moore et al., 1993).

Motives

Topography and parent material are among the soil forming factors that can be more easily mapped than soil attributes. The use of this secondary information in quantitative soil mapping is becoming more and more useful due to advances in GIS and geostatistics, but it is still non existent in most developing countries such as Burundi. In this contribution we focus on the quality of a digital elevation model (DEM) and its derivatives and on the impact of geology on the spatial variability of some key soil attributes.

Material and methods

Contour lines were digitized from an 1/50000-scale topographic map and different interpolation methods were used to create a DEM and several derived terrain attributes. For each interpolation method the DEM quality was assessed by checking for artefacts.

Soil sampling was executed as follows: 120 samples were first taken at a regular interval of 350 m, then an additional 83 samples were taken at irregular intervals but keeping in mind to represent the different hillslope elements. The sampling depth was limited to the top 30 cm. Samples were analyzed for pH, OC and 5 textural fractions.

Estimation of topographic attributes

A common way to assess the quality of an interpolated DEM is the inspection of artefacts in its derivatives, like slope angle and contour map (Wilson and Gallant, 2000). Ordinary kriging combined with a moving window operation and drainage enforcement techniques performed best.

Exploratory data analysis

Soil samples taken from bottom valleys showed OC and clay contents which were higher than those on hillsides and were identified to belong to a distinctly different population. Mean values of pH, OC and texture data were calculated for the 3 rock types of the study area. These mean values suggest no difference for pH and OC data, but highlighted differences in soil particle distribution. The influence of geology is also evidenced in experimental directional variograms: from northern to eastern directions the slope of the variogram gradually increased and the range gradually decreased, with a hole effect for eastern directions.

Discussion

Information on relationships between terrain attributes, soil attributes and geology has to be accounted for in cost effective sampling and mapping. The use of DEM and other auxiliary data has showed to be effective elsewhere: Goovaerts (2000) got good estimates of rainfall by incorporating a DEM in simple kriging with varying local means, kriging with an external drift and collocated cokriging. Hengl (2004) showed that prediction of organic carbon and topsoil thickness by means of regression-kriging performed better than ordinary kriging. The target variables were predicted from categorical and numerical attributes on relief and mapping units.

References:

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

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

Moore, I.D., Gessler, P.E., Nielsen, G.A. and Peterson, G.A. (1993): Soil attributes prediction using terrain analysis. Soil. Sci. Soc. Am. J., Vol.57, 443-452.

Wilson, J.P., and Gallant, J.C. (2000): Terrain Analysis: Principles and Applications. John Wiley and Sons, 479p.

Contact Information: Marc Van Meirvenne, Dept. Soil Management and Soil Care, Faculty of Bioscience Engineering, Ghent University, Coupure 653, 9000 Gent, Belgium, Phone: +32-92646056, Fax: +32-92646247, Email: marc.vanmeirvenne@ugent.be

Improved Modeling of NIR and MIR Soil Diffuse Reflectance Spectra Using Wavelet Analysis

R. A. Viscarra-Rossel
Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW, Australia

Soil spectra can often be noisy and the spectral libraries used for multivariate calibration can be large. Often these libraries are used and re-calibrated many times over for different analysis, making computation of the models quite demanding. Wavelet analysis can be used to smooth and compress spectral data. Here, wavelet analysis is implemented as a pre-processing tool for soil diffuse reflectance spectra and as a complement to partial least squares (PLS) regression.

Therefore, the aims are twofold: (i.) to investigate the use of the discrete wavelet transform (DWT) through multi-resolution analysis (MRA) as an effective de-noising and data compression tool for soil spectra in the near infrared (NIR), and mid infrared (MIR) portions of the electromagnetic spectrum and (ii.) to combine the DWT with PLS regression and compare the resulting models and their predictability. The wavelet used was the orthogonal wavelet basis from the Daubechies family with four vanishing moments.

This work may be viewed as a pilot study that demonstrates the advantages of using the DWT through MRA and PLS. These techniques were useful for detecting characteristic features of the spectra, removing the high-frequency stochastic components of soil spectra, thus determining the optimal amount of de-noising required, and compressing the spectral data into more parsimonious data sets for the regression analysis. The data was compressed to less than 6 % of its original size with slight consequence to the models and hence the accuracy of predictions. The wavelet transformed PLS models (W-PLS) illustrated the stability of the PLS technique and the robustness of the complementary use of DWT and PLS. De-noising improved predictions of soil organic carbon (OC). The techniques may prove beneficial for handling large spectral libraries and for real-time on-the-go and on-line sensing applications.

Contact Information: Raphael Viscarra-Rossel, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney NSW 2006, Phone: +61 2 9351 5813, Fax: +61 2 9351 3706, Email: r.rossel@agec.usyd.edu.au,
Web: www.usyd.edu.au/su/agric/acpa/people/rvrossel/

Mid Infra-Red Spectra as Input to a Soil Inference System

R. A. Viscarra-Rossel, *A. B.McBratney* and *B.Minasny*
Australian Centre for precision Agriculture, The University of Sydney, NSW 2006, Australia

Inference is defined as a process of deriving logical conclusions from the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation. McBratney et al. (2002) proposed soil inference systems (SINFERS) as a knowledge base to infer soil properties and populate the soil digital databases. SINFERS takes measurements with a given level of certainty and infers data that is not known with minimal uncertainties by means of logically linked predictive functions. These predictive functions in a non-spatial context are referred as pedotransfer functions. The basic assumption underlying SINFERS is if we know or are able to predict the basic fundamental properties of the soil, we should be able to infer all other physical and chemical properties using pedotransfer functions. Pedotransfer functions (PTFs) relate basic soil properties to other more difficult or expensive to measure soil properties by means of regression and various data mining tools.

The keys to soil inference systems are reliable inputs and the ability to link basic soil information. The most basic and useful sets of properties is particle-size distribution. Clay content has been demonstrated to influence many physical and chemical properties. The inputs to the inference system can be from various sources:

- Soil survey, i.e. from a soil morphological description: field texture, pH, structure, colour.
- Laboratory measurement of soil physical and chemical properties.
- Spectroscopically, where several key physical and chemical properties can be predicted.

The spectroscopy method has been proved to be most efficient and can provide reliable estimates of soil physical properties (particle size distribution) and chemical properties (pH, CEC, Organic C, etc) (Viscarra Rossel et al., 2003).

This paper demonstrates the application of the soil spectra from near-infrared regions as input to the soil inference system. First the basic soil properties are predicted from the spectra using partial least-squares (PLS) (Geladi and Kowalski, 1986). The predicted properties are: sand, silt and clay content, pH, organic C, and CEC. The uncertainty of the prediction model is also quantified using bootstrapped PLS. From the predicted basic soil properties, other more difficult-to-measure properties can be derived. Examples are:

- The water retention curve,
- hydraulic conductivity cahrateristics,
- soil pH buffering capacity.

References:

Geladi, P. and Kowalski, B.R. 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185: 1-17.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109, 41-73.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik L.J. and Skjemstad, J.O. (2005). Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma (in-press)

Contact Information: Raphael Viscarra-Rossel, Australian Centre for Precision Agriculture, The University of Sydney, JRA McMillan Building A05, NSW 2006, Australia. Phone: (61 2) 9351 5813, Fax: (61 2)9351 3706, Email: r.rossel@agec.usyd.edu.au

ParLeS - Executable Software to Perform Partial Least Squares (PLS) Regression with Delete-One-Jackknife Cross Validation

R. A. Viscarra-Rossel
Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW, Australia

ParLeS provides options for, and performs the following tasks: (i.) spectral data linearization through transformations to absorbance and/or Kubelka-Munk optical density units (ii.) spectral corrections, de-trending and smoothing using various methods including multiplicative signal correction (MSC), standard normal variate (SNV), wavelet transform, first and second derivatives, etc., (iii.) mean centre data (iii.) delete-one-jacknife cross validation to determine optimal number of bilenear factors for (iv.) calibration modeling and (v.) prediction of unknowns. The following data can be saved for external analysis and plotting: (i.) pre-processed spectra, (ii.) the cross validation predictions, (iii.) the scores (t), loadings (P), weights (W) and slopes (B) of the regression and the predictions of unknowns.

Partial Least Squares (PLS) regression (also known as PLSR or PLSR1) is a popular modeling technique in chemometrics, econometrics and in industrial applications. It is a data compression technique that is also commonly used in spectral quantitative analysis. Research in science often involves using variables that are easily (or cheaply) measured to explain or predict the behaviour of response variables that are often much more difficult (or expensive) to acquire. When the factors are few in number and are not significantly redundant (or collinear) and have a well understood relationship to the responses, then multiple linear regression (MLR) can be a useful way to turn data into information. However, if these conditions break down, then MLR will not be efficient or appropriate. PLS is a method used to construct predictive models when factors are many and highly collinear, e.g. in reflectance sprectroscopy. The emphasis of PLSR is on predicting the response. However, when used interactively with proper graphics and validation, it also allows the user to attain a good causal insight into the underlying relationships between the variables.

PLSR is closely related to Principal Component Regression (PCR). However, PLSR is performed in a slightly different manner. Take the case where we want to use spectral reflectance data to model and then estimate the value/concentration of a soil property: Instead of first decomposing the spectra into a set of eigenvectors and scores, and regressing them against the soil values as a separate step, PLSR actually uses the soil information during the decomposition process (the decomposition of both the spectral and the soil data into their most common variations is performed simultaneously). PLSR takes advantage of the correlation that exists between the spectra and the soil values. So, the resulting spectral vectors are directly related to the soil values/concentrations.

<u>Advantages of PLSR</u>:

- handles multicollinearity
- robust in terms of data noise and missing values
- balances the two objectives of explaining response and predictor variation thus predictions are more robust
- calibrations generally more robust

- single step decomposition and regression
- can give good insight into underlying relationship between variables

Disadvantages of PLSR:

- calculations are slightly slower than more classical methods, although this is no longer much of an issue

Contact Information: Raphael Viscarra-Rossel, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney NSW 2006, Phone: +61 2 9351 5813, Fax: +61 2 9351 3706, Email: r.rossel@agec.usyd.edu.au,
Web: www.usyd.edu.au/su/agric/acpa/people/rvrossel/

Visible, Near-Infrared, Mid-Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties

R. A. Viscarra-Rossel[1], *D. J. J. Walvoort*[2], *A. B. McBratney*[1], *L. J. Janik*[3] and *J. O. Skjemstad*[3]
[1]Australian Centre for Precision Agriculture, The University of Sydney, NSW 2006, Australia
[2]Wageningen University & Research Centre, Wageningen, The Netherlands
[3]CSIRO Land and Water, Glen Osmond SA 5064, Australia

Historically, our understanding of the soil and assessment of its quality and function has been gained through routine soil chemical and physical laboratory analysis. There is a global thrust towards the development of more time- and cost-efficient methodologies for soil analysis as there is a great demand for larger amounts of good quality, inexpensive soil data to be used in environmental monitoring, modeling and precision agriculture. Diffuse reflectance spectroscopy provides a good alternative that may be used to enhance or replace conventional methods of soil analysis, as it overcomes some of their limitations. Spectroscopy is rapid, timely, less expensive, non-destructive, straightforward and sometimes more accurate than conventional analysis. Furthermore, a single spectrum allows for simultaneous characterisation of various soil properties and the techniques are adaptable for 'on-the-go' field use.

The aims of this paper are threefold: (i.) determine the value of qualitative analysis in the visible (VIS) (400 – 700 nm), near infrared (NIR) (700 – 2500 nm) and mid infrared (MIR) (2500 – 25000 nm), (ii.) compare the simultaneous predictions of a number of different soil properties in each of these regions and the combined VIS-NIR-MIR to determine whether the combined information produces better predictions of soil properties than each of the individual regions and (iii.) deduce which of these regions may be best suited for simultaneous analysis of various soil properties. In this instance we implemented partial least-squares regression (PLSR) to construct calibration models, which were independently validated for the prediction of various soil properties from the soil spectra. The soil properties examined were soil pH_{Ca}, pH_w, lime requirement (LR), organic carbon (OC), clay, silt, sand, cation exchange capacity (CEC), exchangeable calcium (Ca), exchangeable aluminium (Al), nitrate-nitrogen (NO_3-N), available phosphorus (P_{Col}), exchangeable potassium (K) and electrical conductivity (EC).

Our results demonstrated the value of qualitative soil interpretations using the loading weight vectors from the PLSR decomposition. The MIR was more suitable than the VIS or NIR for this type of analysis due to the higher incidence spectral bands in this region as well as the higher intensity and specificity of the signal. Quantitatively, the accuracy of PLSR predictions in each of the VIS, NIR, MIR and VIS-NIR-MIR spectral regions varied considerably amongst properties. However, more accurate predictions were obtained using the MIR for pH, LR, OC, CEC, clay, silt and sand contents, P and EC. The NIR produced more accurate predictions for exchangeable Al and K than any of the ranges. There were only minor improvements in predictions of clay, silt and sand content using the combined VIS-NIR-MIR range. This work demonstrates the potential of diffuse reflectance spectroscopy using the VIS, NIR and MIR for more efficient soil analysis and the acquisition of soil information.

Contact Information: Raphael Viscarra-Rossel, Australian Centre for Precision Agriculture, Faculty of Agriculture, Food & Natural Resources, McMillan Building A05, The University of Sydney NSW 2006, Phone: +61 2 9351 5813, Fax: +61 2 9351 3706, Email: r.rossel@agec.usyd.edu.au,
Web: www.usyd.edu.au/su/agric/acpa/people/rvrossel/

A Photogrammetric Method for Collecting Three-Dimensional Soil Surface Data

Neffra Matthews, Tom Noble and ***William Ypsilantis***
USDI, Bureau of Land Management, National Science & Technology Center, Denver, CO, USA

Extreme close-range photogrammetry is an excellent method for capturing detailed information about surface features of soils. The term extreme close-range refers to object to camera distances of 50 meters or less. The technique developed enables precise measurements to be made from photographs taken with affordable six megapixel or higher single lens reflex or eight megapixel point and shoot digital cameras.

The two major requirements to create these measurements are complete coverage by overlapping photographs for stereoscopic viewing, and adequate x, y, z control for defined points within the overlapping area of the photographs. Three-dimensional measuring and modeling software is used to calibrate the camera for focal length, format size, principal point, and distortion coefficients. The calibrated camera is then used to capture ground control x, y, z coordinate data. This data can be captured in a series of oblique, orientation photographs which are taken of the subject area with circular reference targets, circular coded targets, and an object of known dimension placed within the target layout. A series of stereoscopic photographs are taken of the subject concurrently with the oblique-orientation photographs to facilitate stereoscopic viewing of the subject area.

Digital, or softcopy, workstations consist of a specialized display that allows the operator to view and collect data in a three-dimensional environment. The softcopy photogrammetric workstation recreates, in three dimensions, the geometry of the soil surface for the purpose of compiling data on it. Microtopographic data results from the automated digital terrain extraction process. The software creates a digital terrain model that consists of a closely spaced grid of thousands of x, y, z data points. Ditigal terrain model grid spacings of one to two millimeters and positional accuracies of one quarter of a millimeter can be determined for areas up to five square meters.

It is possible to make repeat visits to the same location for the purpose of monitoring or change detection. It is necessary to establish monumented points to achieve high levels of accuracy of change detection. The data collected over a time period can be used to precisely document erosion rates or other changes in soil surface features, such as microtopography of biological soil crusts.

The three dimensional monitoring and modeling software also has the capability to use historical photographs to detect change from previous periods of time. The accuracy for these type projects will be less due to the lack of camera calibration and the lack of oblique photographs needed to establish camera orientation.

Contact Information: William Ypsilantis, USDI, Bureau of Land Management, National Science & Technology Center, Denver Federal Center, Bldg. 50, P.O. Box 25047, Denver, CO, 80225, USA, Phone: 303-236-3404, Fax: 303-236-3508, e-mail: bill_ypsilantis@blm.gov

Purposive Sampling for Soil Mapping: Successes and Challenges

A-Xing Zhu[1,2], *Baolin Li*[1], *Edward English*[2], *Lin Yang*[3], *Chengzhi Qin*[1], *James E. Burt*[2] and *Chenghu Zhou*[1]

[1]State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.

[2]Department of Geography, University of Wisconsin, Madison, Wisconsin, U.S.A. Email:

[3]Department of Geography, Beijing Normal University, Beijing, China

Knowledge of relationships between soil and its observable environmental conditions (referred as soil landscape model), is needed for mapping spatial distribution of soil and its properties. This knowledge, Soil-landscape model, is often developed through extensive fieldwork which is not only very labor intensive and slow, but also very costly. With the increasing availability of geospatial data in digital form (GIS) and the sophistication of modern spatial information processing techniques, it is possible to develop methods to enhance the effectiveness of field sampling and to improve the efficiency of developing soil-landscape model.

This paper presents a purposive sampling strategy based on prototype theory and fuzzy logic theory. The basic idea in this approach is that we only need to sample the locations of soils which are the prototypes of soil classes (such as soil series) and the relationships between the soils and the environmental conditions at these locations can be used to approximate the soil-landscape model of the area. Locations whose soils are between these prototypes do not need to be sampled and can be handled by a fuzzy representation.

The key problem in this pursue is to distinguish locations where the soils are the prototypes from other locations where soils are somewhere between soil types without extensive sampling. To address this question, the classic soil-landscape model concept is employed. We assume that similar environmental conditions will produce/reflect similar soils (soil conditions) and that the prototypes of soil classes correspond to the prototypes of the environmental configurations. Thus, the problem now becomes that of finding prototypes of environmental configurations. GIS/RS techniques are used to characterize the soil environmental conditions and classification techniques are used to identify the classes that exist in the environmental data set. These classes are considered to be directly associated with the different soil classes. Fuzzy maps of these derived classes are then used to determine locations where the soils are of prototypes. Thus, discovering the relationships between soils and their environmental conditions is a matter of associating the unique combination of environmental conditions with the soil classes at these locations.

In this illustration, we employed a fuzzy *c*-means classifier to identify the natural clusters in the environmental space and use the centroids of these fuzzy clusters as the prototypes for developing the soil landscape models in two study areas: one in Wisconsin and one in Northeast China. We found the approach is effective in helping local soil scientists to develop their understanding (knowledge) of soil-environmental relationships in the Wisconsin watershed where the developed model can reach an overall accuracy of 76%, with some area as high as 90%, in mapping the soils of the area. But we did not experience the same success with the approach in the Northeast China watershed and where the accuracy is only 56%.

We attribute the mild success in the China watershed to the following aspects: the unique nature of the landscape and the limitation of fuzzy *c*-means classification. The area is characterized by rounded hills and with low relief. The micro-topography (small drops and minor terraces along the slope) plays a key role in soil developing. It is very difficult to pick up these micro-topographic features with current GIS techniques. The fuzzy *c*-means classification technique determines class centroids based on frequencies of environmental conditions. Class centers are more likely to be allocated to environmental conditions with high frequencies than to those with low frequencies. The latter are often associated with distinct geomorphic features (such as ridges, noses) and are more likely to be the prototypes of soils. The rounded nature of the watershed in Northeast China really challenges the classifier in determining the natural clusters. The stronger relief in the Wisconsin watershed made this limitation of fuzzy *c*-means classifier less a problem.

Contact Information: A-Xing Zhu, Department of Geography, University of Wisconsin, 550 North Park Street, Madison, Wisconsin 53706, U.S.A. Email: axing@geography.wisc.edu

Author Index

Bold numbers indicate presenting authors.

Notes

Notes