Book and Software Review of "Yvan Pannatier: Variowin — Software for Spatial Data Analysis in 2D"

Jürgen Symanzik and Noel Cressie

Department of Statistics, Iowa State University, Ames, IA 50011, USA

symanzik@iastate.edu, ncressie@iastate.edu

The book discussed in this review is one of the most recent additions to Springer's series on"Statistics and Computing". In contrast to many other books in that series, where multi–purpose statistical packages such as S – PLUS® and XploRe are the matter of interest, Pannatier's book covers a very specialized collection of four Windows™ programs that can be used for spatial data analysis and variogram modeling of irregularly located data in two dimensions.

The 91 pages of the book are split into six chapters, appendix, bibliography, and index. It includes a 3.5" diskette with programs containing the latest version of *Variowin*, i. e., version 2.2.

The introduction in Chapter 1 gives an overview of the main features of the program and summarizes the history of the software, originally developed as part of Pannatier's Ph.D. research.

Chapter 2 is probably the most important part of the book for readers with little time. It gives a quick start into the four programs included in *Variowin 2.2*: *Prevar2D* (construction of a pair comparison file), *Vario2D with PCF* (exploratory variography), *Model* (interactive variogram modeling), and *Grid Display* (displaying grid files as pixel maps). Each page of this

S–PLUS is a registered trademark of MathSoft, Inc.
Windows is a registered trademark of Microsoft, Inc.

chapter contains screendumps from one of the four previously listed programs. The text indicates which steps are required to obtain the window contents visible in the screendumps. The required demonstration data used here is provided on the accompanying diskette.

Chapter 3 is a very short introduction of pair comparison files (PCFs) and describes how these are created using *Prevar2D* and what kind of problems had to be overcome.

In Chapter 4, the interactive exploratory variography features that are available in *Vario2D with PCF* are described. This program allows the calculation and display of (cross) h–scatterplots, (cross) variogram surfaces, directional (cross) variograms, and (cross) variogram clouds. Whenever one of the options `Calculate | Variogram Surface` or `Calculate | Directional Variogram` is selected, the following measures of spatial continuity can be displayed: (cross) variograms, standardized (cross) variograms, (cross) covariances, (cross) correlograms, and (cross) madograms. Definitions of these are also given in this chapter.

Pannatier uses (what we call)

$$\hat{\nu}_{ij}(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} (z_i(x_k) - z_i(x_k + h)) \cdot (z_j(x_k) - z_j(x_k + h)),$$

where h is a separation vector, to calculate the experimental cross variogram (p. 38). This is an estimator for

$$\nu_{ij}(h) = \frac{1}{2}\mathrm{Cov}((z_i(s + h) - z_i(s)), (z_j(s + h) - z_j(s))).$$

Even though $2\nu_{ij}(h)$ is traditionally recommended for the cross variogram (Journel and Huijbregts, 1978, p. 324), it can be used only under the assumption that cross covariances are symmetric in $(h, -h)$ (Journel and Huijbregts, 1978, p. 326). However, there exists another definition of the cross variogram, namely,

$$2\gamma_{ij}(h) = \mathrm{Var}(z_i(s + h) - z_j(s)),$$

which was first considered by Clark, Basinger, and Harper (1989). It has been shown in Ver Hoef and Cressie (1993, 1994) that $2\gamma_{ij}(h)$ does not depend on the "symmetric–cross–covariance condition" and, moreover, linear prediction based on $2\gamma_{ij}(h)$ is more efficient than linear prediction based on $2\nu_{ij}(h)$. We recommend that $\{z_i(x_k)\}$ and $\{z_j(x_k)\}$ be first standardized so that $\overline{z}_i = \overline{z}_j = 0$ and $s_i^2 = s_j^2 = 1$ and then, that $2\gamma_{ij}(h)$ be estimated by

$$2\hat{\gamma}_{ij}(h) = \frac{1}{N(h)} \sum_{k=1}^{N(h)} (z_i(x_k + h) - z_j(x_k))^2.$$

While all other programs of the *Variowin 2.2* package are almost self–explanatory, the program *Model* (used for interactive variogram modeling)

and its input window require an explicit description, given in Chapter 5. Five variogram models are supported by the software: pure nugget–effect models, spherical models, exponential models, Gaussian models, and power models.

Chapter 6 is intended for the advanced reader/user of the software since it discusses the format of data files used and generated by the four programs in the *Variowin 2.2* package.

An appendix on geostatistical concepts provides some of the necessary theoretical background for the work possible with the *Variowin 2.2* package. However, it might be very helpful for the reader to look additionally at some of the standard references on geostatistics, e. g., Isaaks and Srivastava (1989) and Cressie (1993). A bibliography and an index conclude the book.

The *Variowin 2.2* software runs on any PC with a 386 processor and above (running Windows™ 3.x, 95, or NT), a coprocessor, and approximately 2MB hard disk capacity. The four programs are mostly operated through the mouse while function keys serve as accelerators. Unfortunately, the split of the software into four (dependent) programs is a very user–unfriendly feature. In a typical application, the program *Prevar2D* has to be started first to create the PCF. Then, *Vario2D with PCF* is started. It would have been more convenient if a single option, e. g., in the "File" menu of *Vario2D with PCF*, provided the functionality of *Prevar2D*. Similarly, after the interactive variogram modeling has been done in *Model*, the user explicitly has to create external files and start the program *Grid Display* to display pixel maps. We would recommend that a future version of *Variowin* joins the four programs into one, thus removing the burden from the user to (re–)start different programs and to create and handle external files.

Two more things should be mentioned about the user interface: The way in which multiple subwindows within the *Variowin* main window can be customized and handled is very convenient and helpful. Otherwise, on–line help for the features and options of the programs is almost non–existent.

As far as the features on exploratory variography (provided by *Vario2D with PCF*) and variogram modeling (provided by *Model*) are concerned, *Variowin* is definitively one of the most sophisticated programs that currently provide this type of application. Therefore, the book should be considered as a users' guide to *Variowin* that directly addresses exploratory geostatistical data analysis. Without the software, the book has very little intrinsic merit.

One further critical comment about the scope of the software should be made. Recently, many links between software packages have been developed where the statistical application operates directly on data extracted from a geographic information system (GIS), e. g., S–PLUS for ARC/INFO by MathSoft, the link between SpaceStat and ArcView™ (Anselin & Bao 1996), and the link between XGobi and ArcView™ (Cook, Majure, Symanzik & Cressie 1996); however, the *Variowin 2.2* package can be used only as stand–

™ *ArcView* is a trademark of Environmental Systems Research Institute, Inc.

alone software. The new links between software packages referred to above allow one to modify displays in the statistical package while the sample map in the GIS, that also displays many other types of geographic information, is concomitantly modified (and vice versa). *Variowin* only allows simultaneous modifications of sample maps and plots in its window but, unfortunately, no additional geographic information can be displayed.

In the long run, *Variowin* also has to compete against S+SpatialStats from MathSoft. Many researchers are already familar with the basic S – PLUS® software and, therefore, may not be willing to learn how to use a new package. In general, S+SpatialStats offers a much broader functionality for the exploration and modeling of spatially correlated data, including some (but not all) of the features provided by *Variowin*. However, it should be left to the reader to decide if he or she is interested in a comprehensive software package for the analysis of spatial data (provided by S+SpatialStats) or in the very specialized and sophisticated exploratory variography and variogram modeling for geostatistical data (provided by *Variowin*).

References

Anselin, L. & Bao, S. (1996), Exploratory Spatial Data Analysis Linking SpaceStat and ArcView, Technical Report 9618, West Virginia University, Morgantown, WV.

Clark, I., Basinger, K. L. & Harper, W. V. (1989), MUCK — A Novel Approach to Co–Kriging, *in* B. E. Buxton, ed., 'Proceedings of the Conference on Geostatistical, Sensitivity, and Uncertainty Methods for Ground–Water Flow and Radionuclide Transport Modeling', Battelle Press, Columbus, OH, pp. 473–493.

Cook, D., Majure, J. J., Symanzik, J. & Cressie, N. (1996), 'Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software', *Computational Statistics* **11**(4), 467–480.

Cressie, N. A. C. (1993), *Statistics for Spatial Data (revised edition)*, Wiley, New York, NY.

Isaaks, E. H. & Srivastava, R. M. (1989), *Applied Geostatistics*, Oxford University Press, New York, NY.

Journel, A. G. & Huijbregts, C. J. (1978), *Mining Geostatistics*, Academic Press, London.

Ver Hoef, J. M. & Cressie, N. (1993), 'Multivariable Spatial Prediction', *Mathematical Geology* **25**(2), 219–240.

Ver Hoef, J. M. & Cressie, N. (1994), 'Errata: Multivariable Spatial Prediction', *Mathematical Geology* **26**(2), 273–275.