



Project Summary

Geostatistics for Waste Management: A User's Manual for the GEOPACK (Version 1.0) Geostatistical Software System

S.R. Yates and M.V. Yates

GEOPACK, a comprehensive user-friendly geostatistical software system, was developed to help in the analysis of spatially correlated data. The software system was developed to be used by scientists, engineers, regulators, etc., with little experience in geostatistical techniques and still satisfy the requirements of more advanced users. By using GEOPACK, and spending a little time becoming familiar with geostatistics, end-users will be able to include these geostatistical techniques in their work and research environments.

This Project Summary was developed by EPA's Robert S. Kerr Environmental Research Laboratory, Ada, OK, to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

Using the geostatistical techniques in the analysis of spatially correlated data generally requires the use of a computer to handle the large number of samples and carry out the lengthy calculations. Unless one knows someone who is willing to provide the necessary computer programs, one is faced with the difficult task of finding, purchasing or developing the required computer software. Although there are a number of practicing geostatisticians who undoubtedly have

access to the necessary programs, these programs are not generally available or they are proprietary codes. Often, the programs which are developed for research purposes are subject to limited availability and are difficult for others to use or modify for purposes other than those for which they were originally designed.

GEOPACK was developed with the philosophy that geostatistical software is needed that can be used by individuals with a minimum level of geostatistical expertise and yet can also satisfy the needs of more sophisticated users. The specific objectives in developing this program were:

- (1) develop geostatistical software for individuals without a great deal of geostatistical training and allow those individuals to learn these techniques and eventually use them in their work environment
- (2) develop a system which is adaptable in the sense that additional programs could be incorporated into the system at a later date without having to alter previous programs or recompile the entire system
- (3) develop programs which produce graphic output in a variety of forms and of publishable quality to meet the needs of research scientists and engineers
- (4) include on-line help facilities and extensive error checking in the programs

The on-line help facilities offer information concerning the operation of

the system, its capabilities and limitations, how to alter the system, as well as programming conventions and definitions. GEOPACK allows the incorporation of other programs, such as the GEO-EAS (EPA/600/4-88/033) system. Examples showing how this software can be used in the analysis of spatially correlated data can be found in the GEOPACK users manual.

Basic System Description

The GEOPACK system includes programs to do the more common statistical and geostatistical analyses. The system is estimation oriented in that if the ordering in the menu system is followed, a grid of estimates for the selected variable in the data set will result. A description of the various components of the system follows.

Basic Statistics

Basic statistics such as the mean, median, variance, standard deviation, skew, kurtosis and maximum and minimum values can be determined for the selected data set. Routines are also available for linear regression, polynomial regression, Kolmogorov-Smirnov test for distribution and calculating the percentiles of the data set. A user-supplied statistics package can be incorporated into GEOPACK to allow the user to access the comprehensive statistical analyses that are contained in many commercial statistics packages.

Variography

The sample semivariogram or cross-semivariogram for a two-dimensional spatially-dependent random function can be determined. The approach used in determining the sample semivariogram is similar to that outlined in Journel and Huijbregts (1978). A model can be fitted to the sample semivariogram using the nonlinear least-squares fitting procedure of Marquardt (1963). This provides a first estimate for the coefficients to be used in a cross-validation program and helps to automate the model-fitting procedure. If the least-squares technique fails, or other information is available which should be included in the model-fitting process, the traditional iterative method of selecting the model coefficients and viewing a graph comparing the sample values to the model can be utilized.

Linear Estimation

GEOPACK includes programs to calculate the ordinary kriging and cokriging estimators in two dimensions along with their associated estimation

variance. The programs include punctual and block kriging and geometric anisotropy. There is a cross-validation option which uses the kriging estimator in a jackknifing mode to cross validate the spatial correlation structure. It is possible to include indicator kriging in an analysis by first transforming the data.

Nonlinear Estimation

Nonlinear estimators such as the disjunctive kriging and cokriging estimators can be determined along with the estimation variance and the conditional probability that the value is greater than a specified cutoff level. Up to 10 cutoff levels are allowed. As with the linear estimation method, this type of an analysis can be done on punctual or block support and may include anisotropy.

Help Facilities

The program includes on-line help facilities to provide the user with information concerning the operation of the program, data requirements, conventions, definitions, run-time errors, missing files, etc. that are encountered during execution. At the main menu level, the help information is of a general nature. During execution of a program, the help is more specific, such as defining a term. Virtually all the information needed to operate or modify the system is available from the HELP facility.

Other Features of GEOPACK

The program also includes various graphics capabilities such as linear or logarithmic line plots, contour and pixel diagrams. The program can be interfaced with any user-supplied graphics package so that custom diagrams can be developed.

GEOPACK uses dynamic allocation of memory so that data sets with a wide range of variables and positions can be used without having to alter the program. A large storage array is partitioned based on the number of samples and variables so that there is little wasted space compared to defining the arrays to have a fixed number. One limitation is that GEOPACK allows a maximum of 10 variables plus their x and y positions and a sample or position number. If an array must be created by a program, the space is obtained from the large storage array. If attempts are made to use more memory than is available, an error message is printed out giving the memory status. From this information, a

decision can be made on how to reduce the memory needs to allowable limits.

GEOPACK uses data in a standard ASCII format for data input. Data can be entered with any program (data base or spreadsheet) or word processor that supports ASCII format. There is a seven line header associated with each data file. This header consists of three lines of title information, the number of random variables, total number of samples, the names to use for the random variables, and a format specifier which describes the way the data is to be read into GEOPACK. The format specifier follows the ANSI FORTRAN convention. The sample data file is on the following page.

Program Structure

The program has been structured to enable the addition of programs by end-users. This has led to the development of a menu system from which a particular program is executed. Part of the system is hard coded and cannot be changed; but by using what is termed a USER's menu, a program or another user-defined menu can be added to the system at any time. The user menu is accessed from the F5 key and reads the instructions contained in a data file. The data file can be modified to include a different set of instructions, and thus allow the system to be modified to suit the end-user's needs. Through this menu, the end-user can add any number of programs, menus and subdirectories to the system.

Program Utilities

Many utility programs are included enabling the user to access a variety of information and other computer functions while using GEOPACK. A sample of the utilities from within GEOPACK are those to 1) select, edit or modify an existing data set, 2) pack the contents of the temporary directory into a compressed format for later use or to extract all or some of the files contained in a compressed file, 3) display the program structure, 4) temporarily leave the program in a DOS shell, 5) execute a DOS command, 6) view a file. Also included are a number of utilities which help the user to tailor GEOPACK to specific needs by facilitating the passage of information to the new user-supplied applications.

Computer Requirements

The programs have been written in a combination of FORTRAN and C programming languages and run on IBM-compatible microcomputers such as the PC-AT, Compaq-286, -386, Zenith, etc.,

This is a typical data file. There are 4 random variables: MOIST, TEMP, SAND and OIL-%. There are 119 positions where data was collected (only 4 positions are shown).

4	119						
	1	2	3	5	4	6	7
	MOIST		TEMP		SAND		OIL-%
	(G5.0,12F10.3)						
1	6.0000		7.0000		46.8500	999.9990	56.5102 6.5362
2	6.0000		10.0000		46.2900	5.9250	55.6444 5.2454
.
.
.
118	24.0000		21.0000		46.3500	999.9990	54.4012 4.0463
119	22.0000		24.0000		47.1400	999.9990	52.5845 2.5345

using an MS-DOS operating system (ideally version 3.2 or greater) and 640 K memory. (PC-XT and compatibles are not recommended for GEOPACK.) A math coprocessor is recommended but not required. This is a mathematical intensive system and calculations will be significantly faster if a math coprocessor is installed. The system can support the use of a virtual disk (RAM disk) as the temporary storage device. GEOPACK also requires that the ANSI.SYS driver be installed if the screen output is to perform properly. Because of its integrated nature, GEOPACK requires a hard disk storage with about 4 Mbytes free disk space. A graphics monitor is required due to the graphical nature of the output.

GEOPACK supports either a CGA, EGA, VGA or HERCULES graphics adapter and the appropriate monochrome or color monitor. The system includes a graphics program for the printing of graphical images and supports the following devices: HPGL compatible plotters, HPCL compatible printers, and Epson compatible dot matrix printers.

Software Availability

The GEOPACK system in its executable form is in the public domain and can be obtained by sending four preformatted diskettes [either 5.25 inch high density (1.2Mb) or 3.5 inch high density (1.44Mb)] to:

GEOPACK Distribution
 US EPA
 Robert S. Kerr Environmental
 Research Laboratory
 P.O. Box 1198
 Ada, OK 74821-1198
 Telephone: (405) 332-8800

References

- Journel, A. G., and Ch. J. Huijbregts, Mining Geostatistics, Academic Press, New York, 1978.
- Marquardt, D. W., An Algorithm for Least-squares Estimation of Non-linear Parameters, J. Soc. Ind. Appl. Math., 11:431-441, 1963.

S. R. Yates is with the U.S. Salinity Laboratory, Riverside, CA 92501 and M. V. Yates is with the University of California, Riverside, CA 92521

David M. Walters is the EPA Project Officer (see below)..

The complete report, entitled "Geostatistics for Waste Management: A User's Manual for the GEOPACK (Version 1.0) Geostatistical Software System," Order No. PB 90-186 420/AS; Cost: \$17.00 subject to change) will be available only from:

*National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650*

The EPA Project Officer can be contacted at:

*Robert S. Kerr Environmental Research Laboratory
U.S. Environmental Protection Agency
Ada, OK 74820*

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Official Business
Penalty for Private Use \$300

EPA/600/S8-90/004

•

•