United States
Environmental Protection
Agency

Environmental Monitoring
Systems Laboratory
Las Vegas NV 89193-3478

Research and Development

EPA/600/S4-88/033 Feb. 1989

EPA

Project Summary

Geo-EAS (Geostatistical Environmental Assessment Software) User's Guide

Evan J. Englund and Allen R. Sparks

This report describes how to install and use the Geo-EAS (Geostatistical Environmental Assessment Software) software package on an IBM-PC compatible computer system. A detailed example is provided showing how to use the software to conduct a geostatistical analysis of a data set.

Thirteen Geo-EAS programs are documented. The principal functions of the package are the production of 2-dimensional grids and contour maps of interpolated (kriged) estimates from sample data. Other functions include data preparation, data maps, univariate statistics, scatter plots/linear regression, and vario-gram computation and model fitting. Extensive use of screen graphics such as maps, histograms, scatter plots and variograms help the user search for patterns, correlations, and problems in a data set. Data maps, contour maps, and scatter plots can be plotted on an HP compatible pen plotter. Individual programs can be run independently; the statistics and graphics routines may prove useful even when a full geostatistical study is not appropriate. For ease of use, the programs are controlled inter-actively through screen menus, and use simple ASCII data files.

This Project Summary was developed by EPA's Environmental Monitoring Systems Laboratory, Las Vegas, NV, 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

Geostatistical methods are useful for site assessment and monitoring situations where data are collected on a spatial network of sampling locations, and are particularly suited to cases where contour maps of pollutant concentration (or other variables) are desired. Examples of environmental applications include lead and cadmium concentrations in soils surrounding smelter sites, outdoor atmospheric NO_2 concentrations in metropolitan areas, and regional sulfate deposition in rainfall. Kriging is a weighted moving average method used to interpolate values from a sample data set onto a grid of points for contouring. The kriging weights are computed from a variogram, which measures the degree of correlation among sample values in the area as a function of the distance and direction between samples.

Kriging has a number of advantages over most other interpolation methods:

Smoothing – Kriging smooths, or regresses, estimates based on the proportion of total sample variance accounted for by random "noise". The noisier the data set, the less individual samples represent their immediate vicinity, and the more they are smoothed.

Declustering – The kriging weight assigned to a sample is lowered to the degree that its information is duplicated by nearby, highly correlated samples. This helps mitigate the impact of oversampling "hot spots".

Anisotropy – When samples are more highly correlated in a particular direction, kriging weights will be greater for samples in that direction.

Precision – Given a variogram representative of the area to be estimated, kriging will compute the most precise estimates possible from the available data. In practice, this is only approximated, as the variogram must itself be estimated from the available data.

Estimation of the variogram from sample data is a critical part of a geostatistical study. The procedure involves interpretation and judgment, and often requires a large number of "trial and error" computer runs. The lack of inexpensive, easy-to-use software has prevented many people from acquiring the experience necessary to use geostatistical methods effectively. This software is designed to make it easy for the novice to begin using geostatistical methods and to learn by doing, as well as to provide sufficient power and flexibility for the experienced user to solve real-world problems.

The basic tools of applied geostatistics, variogram analysis and kriging, were developed in the early 1960's for use in the mining industry, and applied to the problems of ore reserve estimation and grade control. In recent years, use of geostatistical methods has become widespread among researchers in various fields such as meteorology, hydrology, and soils science. The ongoing USEPA research program in environmental geostatistics at the Environmental Monitoring Systems Laboratory - Las Vegas (EMSL-LV) was begun about six years ago. As an incidental result of various research projects conducted under Cooperative Agreements with Stanford University, the University of Wyoming, and the University of Arizona, EMSL-LV acquired a miscellaneous collection of public domain geostatistical and statistical software. Development of the Geo-EAS package began in the summer of 1986, as a means of making the software widely available, adequately documented, and easy to use. The system was designed in cooperation with the Applied Earth Sciences Department of Stanford University, and programming was done by the Computer Sciences Corporation, Las Vegas, NV. Design features such as the simple ASCII file formats, standardized menu screens, and integrated graphics, give Geo-EAS flexibility for future expansion; it is anticipated that Geo-EAS will become a significant technology transfer mechanism for more advanced methods resulting from the EMSL-LV research and development program.

The Geo-EAS System

Geo-EAS (Geostatistical Environmental Assessment Software) is a collection of interactive software tools for performing two-dimensional geostatistical analyses of spatially distributed data. Programs are provided for data file management, data transformations, univariate statistics, variogram analysis, cross validation, kriging, contour mapping, post plots, and line or scatter graphs. Features such as hierarchical menus, informative messages, full-screen data entry, parameter files, and graphical displays are used to provide a high degree of interactivity, allowing users to easily alter parameters and recalculate results.

Data Files

The Geo-EAS programs use a simple ASCII file structure for data input. The files contain a header record, the number of variables, a list of variable names and units, and a numeric data table. The first few lines of an example file are listed below.

```
Example.dat - Geostatistical
Environmental Assessment Software
 5
Easting     feet
Northing    feet
Arsenic     ppm
Cadmium     ppm
Lead        ppm
    288.0   311.0   .850   11.5  18.25
    285.6   288.0   .630   8.50  30.25
    273.6   269.0   1.02   7.00  20.00
    280.8   249.0   1.02   10.7  19.25
    -----   -----   -----  ---   -----
```

Menu Screens

All Geo-EAS programs are controlled interactively through menu screens (Figure 1) which permit the user to select options and enter control parameters. While default choices for many parameters and options make the programs easier to use, the programs are structured to avoid a "black box" approach to data analysis. Several of the more complex programs permit the user to save and read "parameter files", making it easy to quickly rerun a program at a later date.

File Utilities

The programs DATAPREP and TR/ provide capability for manipulating G EAS data files. Files can be appende merged, and variables can be crea transformed, or deleted. Transforma operations include natural log, squ root, rank order, indicator, and arithm operations.

Data Maps

POSTPLOT creates a map (Figure of a data variable in a Geo-EAS c file. Symbols representing the quartile the data values and/or the data val themselves are plotted at the sam locations.

Univariate Statistics

STAT1 computes univariate statis such as mean, standard deviation,etc., variables in a Geo-EAS data file, plots histograms (Figure 3) and pr ability plots.

X-Y Plots

SCATTER and XYGRAPH both cre x-y plots with optional linear regress for any two variables in a Geo-EAS d file. SCATTER (Figure 4) is most us for quick exploratory data analysis, w XYGRAPH provides additional capa ities such as multiple 'y' variables, scaling and labelling options.

Variogram Analysis

PREVAR creates an intermedi binary file of data pairs for use in VAR which computes and displays plots variograms for specified distance directional limits. Variogram models be interactively fitted to the experime points (Figure 5). The fitted model n be the sum of up to five independ components, which can be a combination of nugget, linear, spheri exponential, or gaussian models. XVA is a cross-validation program which test a variogram model by estimat values at sampled locations fr surrounding data and comparing estimates with the known sample value

Interpolation

KRIGE provides kriged estimates for a two-dimensional grid of points blocks. The user may specify the g parameters; an elliptical sear neighborhood; and up to five "neste independently anisotropic, variogr model components. During execution shaded map of estimated values

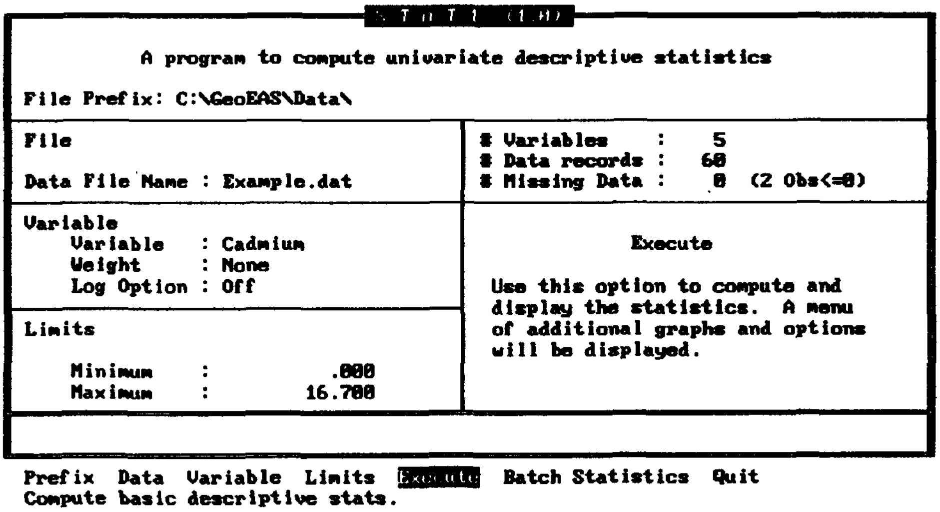

Figure 1. *Initial menu for program STAT 1.*

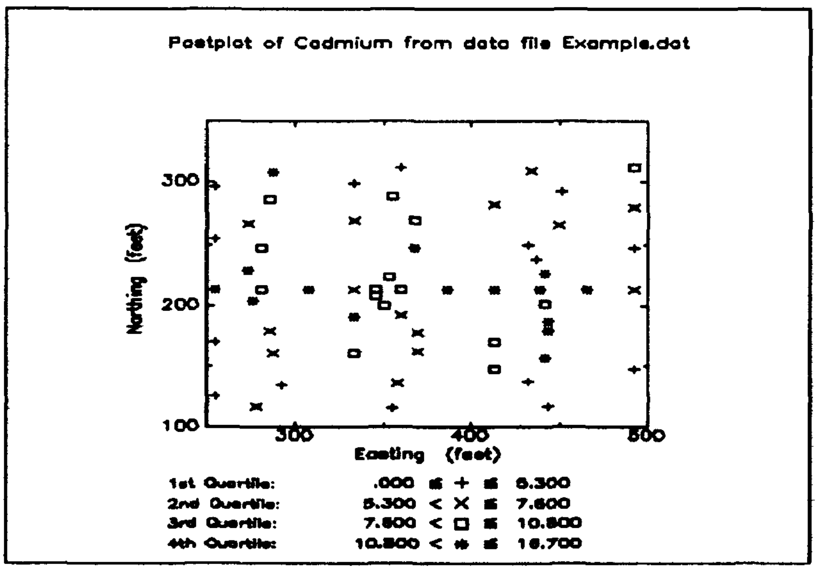

Figure 2. *Sample location map from program POSTPLAT.*

displayed (Figure 6), and a Geo-EAS file of kriged grid results is produced.

Contour Maps

CONREC generates contour maps (Figure 7) from a gridded Geo-EAS data file, usually the output from KRIGE. Options are provided for contour intervals and labels, degree of contour line smoothing, high-low symbols, and for dashed contour lines below a specified level.

Pen Plotting

The programs POSTPLOT, XYGRAPH, and CONREC are based on subroutines originally developed by the National Center for Atmospheric Research (NCAR), and produce graphics "metafiles" which can be saved and replotted later. HPPLOT reads a metafile and produces a file of HPGL commands which can be plotted on Hewlett-Packard compatible plotters. VIEW reads a metafile and displays the plot on the monitor for review.

User Profile

The Geo-EAS user should have some familiarity with personal computers and DOS (Disk Operating System) commands such as DIR (Directory), CD (Change Directory), COPY, etc., and how to insert and use diskettes. EDLIN or some other ASCII file text editor will often be needed for the initial formatting of Geo-EAS data files (adding header records, missing value codes, etc.). It is assumed that the Geo-EAS user has a basic understanding of geostatistical

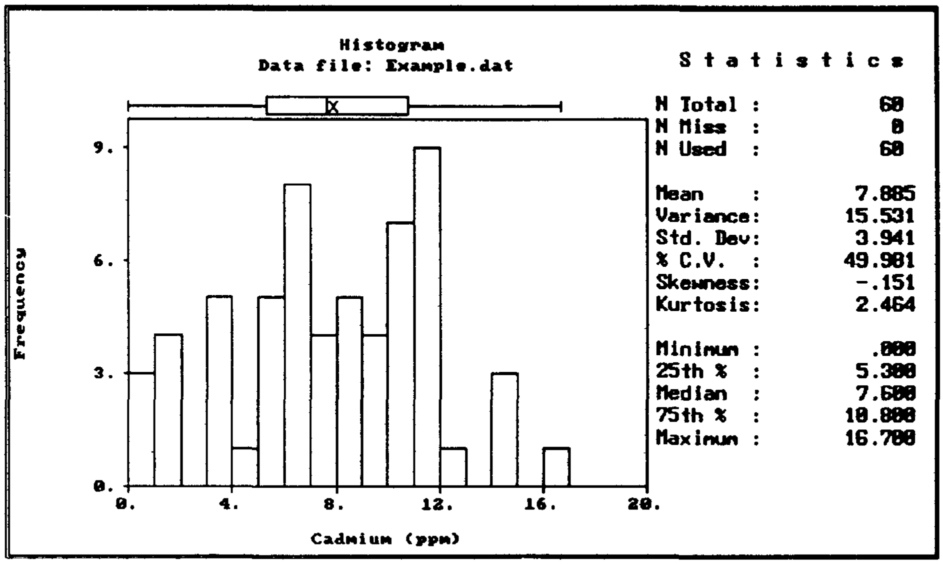

Figure 3. *Histogram from program STAT 1.*

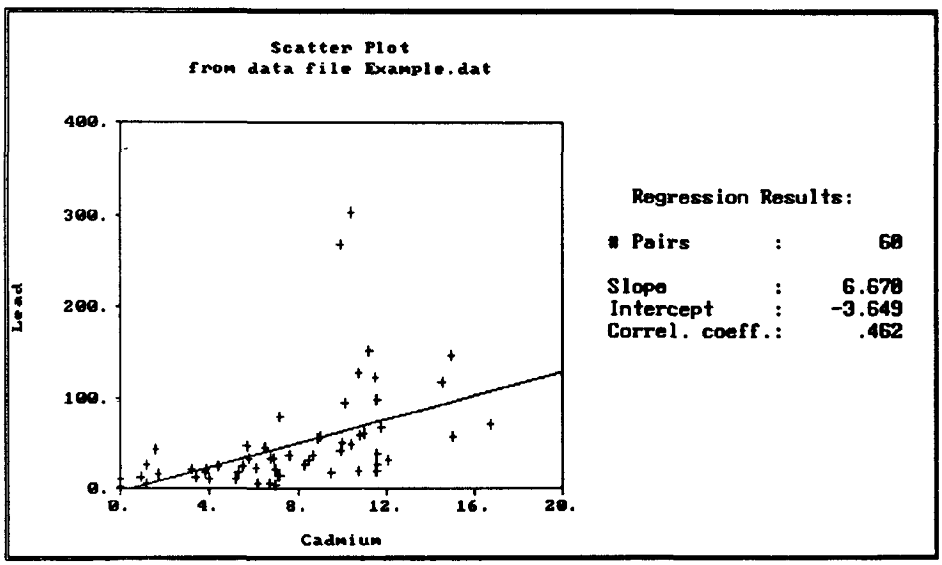

Figure 4. *XY plot with linear regression from program SCATTER.*

concepts, and preferably some prior experience in variogram analysis and kriging; however, the Geo-EAS User's Guide includes a fully worked example which can provide a starting point for the novice user.

Hardware Requirements

This system was designed to run under DOS (Disk Operating System) on an IBM PC, XT, AT, PS2, or compatible computer. Graphics capability is not required, but is highly recommended, as most programs will produce graphics output. Graphics support is provided for the Hercules monochrome graphics card, the Color Graphics Adapter (CGA), and he Enhanced Graphics Adapter (EGA). 640 kilobytes (Kbytes) of random access memory (RAM) is required. An arithmetic co-processor chip is strongly recommended due to the computationally intensive nature of the programs, but is not required for use. Programs may be run from floppy diskette, however, a fixed disk is required to use the programs from the system menu. The system storage requirement is approximately thı megabytes. For hardcopy of results graphics printer (IBM graphics cc patible) is required. Support is provic for plotters which accept HPGL plott commands.

Geo-EAS Software Availabilit

The Geo-EAS software in its exe table form is entirely in the pul domain, and can be obtained by send the appropriate number of diske (PRE-FORMATTED, PLEASE!) to following address:

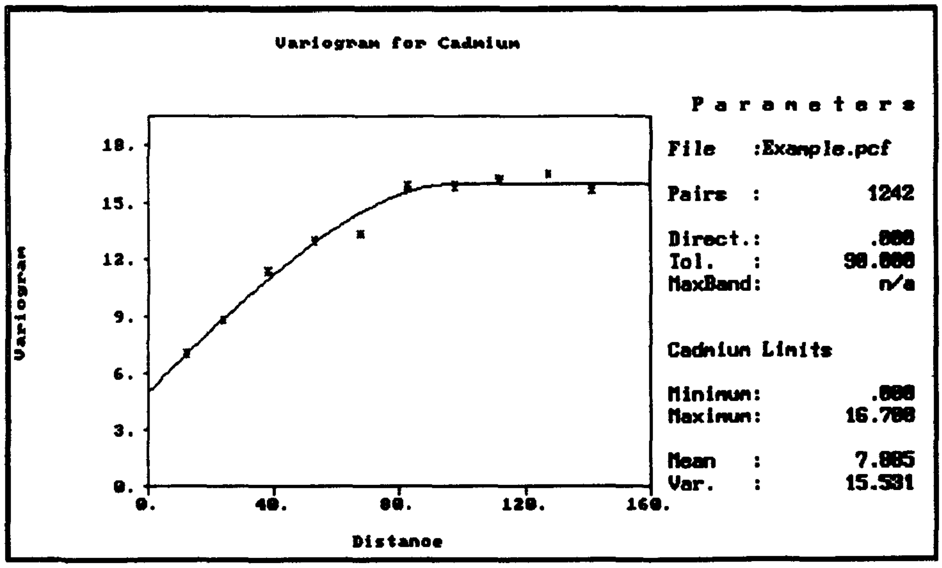

Figure 5. Variogram with fitted model from program VARIO.

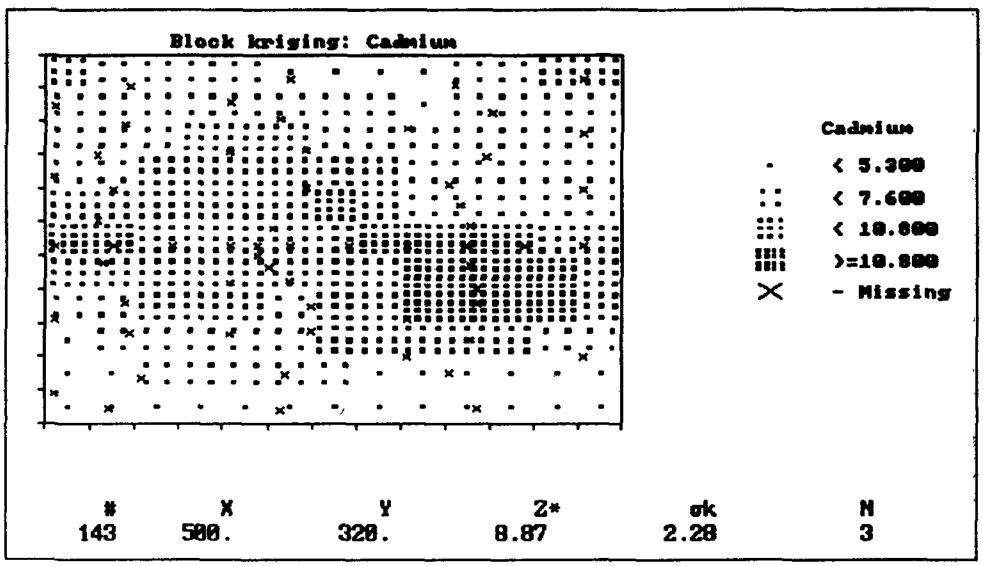

Figure 6. Screen display of interpolated results from program KRIGE.

Evan J. Englund (Geo-EAS)
USEPA EMSL-LV, EAD
P.O. Box 93478
Las Vegas, NV 89193-3478

The executable files and example data sets take approximately 3 megabytes of storage, and require several diskettes, depending on the diskette type.

Type		Number
5 1/4"	1.2MB	3
5 1/4"	360KB	9
3 1/2"	1.44MB	3
3 1/2"	722KB	6

The source code is written in FORTRAN using the Microsoft 4.01 (Microsoft Corporation, Redmond, WA) FORTRAN compiler. The graphics in Geo-EAS are accomplished through the use of a slightly modified version of the proprietary GRAFLIB graphics utilities (Sutra Software, Sugarland TX), and a PC version of the NCAR graphics utilities developed at the National Center for Atmospheric Research in Boulder, Colorado. The NCAR and Geo-EAS source code are in the public domain. For further information on the Geo-EAS source code and programmer documentation, contact:

Geo-EAS
Computer Sciences Corporation
P.O. Box 93478
Las Vegas, NV 89193-3478

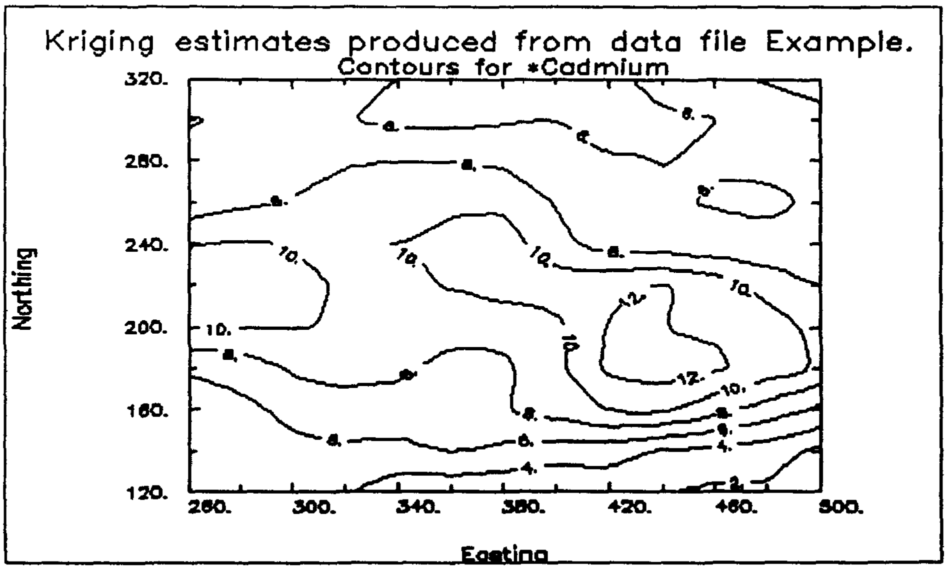

Figure 7. *Contour map of kriged estimates from program CONREC.*

*The EPA author, **Evan J. Englund** (also the EPA Project Officer see below), is with the Environmental Monitoring Systems Laboratory, Las Vegas, NV, and Allen R. Sparks is with Computer Sciences Corporation, Las Vegas, NV 89119.*

The complete report consists of paper copy and software, entitled "Geo-EAS (Geostatistical Environmental Assessment Software) User's Guide,"

Paper Copy, (Order No. PB 89-151 252/AS/AS; Cost: $21.95)

Software, (Order No. PB 89-151 245/AS/AS; Cost: $60.00; cost of software includes paper copy)

The above items will be available only from: (costs subject to change)

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

The EPA Project Officer can be contacted at:

Environmental Monitoring Systems Laboratory
U.S. Environmental Protection Agency
Las Vegas, NV 89193-3478

United States Environmental Protection Agency
Center for Environmental Research Information
Cincinnati OH 45268

EPA/600/S4-88/033