UNCERT: geostatistics, uncertainty analysis and visualization software applied to groundwater flow and contaminant transport modeling

William L. Wingle *, Eileen P. Poeter1, Sean A. McKenna2

Department of Geology and Geological Engineering, Colorado School of Mines, Golden, CO 80401, USA

Received 2 July 1998; received in revised form 27 August 1998; accepted 27 August 1998

Abstract

UNCERT is a 2D and 3D geostatistics, uncertainty analysis and visualization software package applied to groundwater flow and contaminant transport modeling. It is a collection of modules that provides tools for linear regression, univariate statistics, semivariogram analysis, inverse-distance gridding, trend-surface analysis, simple and ordinary kriging and discrete conditional indicator simulation. Graphical user interfaces for MODFLOW and MT3D, groundwater flow and contaminant transport models, are provided for streamlined data input and result analysis. Visualization tools are included for displaying data input and output. These include, but are not limited to, 2D and 3D scatter plots, histograms, box and whisker plots, 2D contour maps, surface renderings of 2D gridded data and 3D views of gridded data. By design, UNCERT’s graphical user interface and visualization tools facilitate model design and analysis. There are few built in restrictions on data set sizes and each module (with two exceptions) can be run in either graphical or batch mode. UNCERT is in the public domain and is available from the World Wide Web with complete on-line and printable (PDF) documentation. UNCERT is written in ANSI-C with a small amount of FORTRAN77, for UNIX workstations running X-Windows and Motif (or Lestif). This article discusses the features of each module and demonstrates how they can be used individually and in combination. The tools are applicable to a wide range of fields and are currently used by researchers in the ground water, mining, mathematics, chemistry and geophysics, to name a few disciplines.

Keywords: Geostatistics; Variogram analysis; Kriging; Trend-surface analysis; Simulation; Visualization; Ground water modeling; Contaminant transport modeling

1. Introduction

UNCERT was developed as a software package designed to aid hydrogeologists to simulate the distribution of materials and material properties in the subsurface, evaluate groundwater flow and contaminant transport and design and evaluate alternative contaminant remediation actions. The flexibility of UNCERT

* Corresponding author. Tel.: +1-303-273-3905; fax: +1-303-273-3859; e-mail: wwingle@mines.edu.
1 Fax: +1-303-273-3859; e-mail: epoeter@mines.edu
2 Present address: Sandia National Laboratories, Albuquerque, NW 87185, USA. Fax: +1-505-844-4426; e-mail: samcken@nwer.sandia.gov
has lead to its use by a wide range of scientists from fields as diverse as chemistry, geophysics, mining and meteorology. The package is written in ANSI-C and a small amount of FORTRAN77 and runs on UNIX systems with X-Windows and Motif (or Lessft). The modules use graphical user interfaces (GUI's) but may also be run from the command line or in batch mode without a graphical display. The ability to run in batch mode makes it practical to model large data sets, model and process series of Monte Carlo simulations or model alternative conceptual models of a site with minimal modeler interaction and easily reproduce model results.

This package was initially developed in the early 1990s to give small consulting firms and academics an inexpensive (free) set of tools to aid in data visualization, to organize many of the data handling requirements in geostatistical and ground water modeling analysis and to act as a test bed for implementing new research tools and algorithms. Few tools were available at the time and were limited in functionality. For example, GeoEAS was a 2D package (Englund and Sparks, 1988), other packages had limited visualization tools, such as GSLIB, which only supported Postscript output (Deutsch and Journel, 1992), and MODFLOW which had no graphical output (McDonald and Harbaugh, 1984); others were expensive. UNCERT merged many of the best or most used features of each into a single package. It was also understood that UNCERT would never solve everyone's problems and that it would be important to make it simple to import data from other programs (UNCERT can read most GSLIB data and grid files even though it does not support many of the algorithms) or export data to other programs (some data manipulation may be required, but simple programs can be written to make the conversions). This flexibility allows modelers to alternatively use UNCERT and incorporate other packages, but UNCERT performs all the basic steps required for geostatistical and ground water analysis (Fig. 1).

There are thirteen main software modules within the UNCERT package. These modules allow the modeler to (1) input raw field data or data from a pre-existing database, (2) analyze the data using classical statistics, (3) evaluate trends, (4) evaluate data using geostatistical techniques such as semivariogram analysis, various kriging techniques (simple and ordinary) and stochastic indicator simulation. When the data are analyzed or when data are prepared from other sources, graphical tools are available to view the results in two-, two-and-a-half- and three-dimensions. Once the spatial variations have been determined, tools allow this information to be (5) imported and modeled with ground water flow and contaminant transport models such as MODFLOW and MT3D and the results can be (6) evaluated and analyzed for individual runs as well as for the composite results of multiple model simulations.

2. Modules

UNCERT is a collection of modules that can be used independently or in combination (some modules call other modules, usually for data visualization). There are thirteen main modules. Most of these modules support some sort of graphical output and any output that is sent to a graphics monitor can also be printed in Postscript (the appearance may vary slightly). All the modules except for modmain and mt3dmain (pre- and post-processors or MODFLOW and MT3D, respectively) can be (1) run in batch mode, (2) have any or all of their parameters defined on the command line or (3) have any or all of their parameters default values defined in a preference file (there are minor exceptions to the last two items). This last option is particularly useful when one set of defaults can be used repeatedly for one or more data sets in a project.

2.1. Plotgraph: X–Y graphs and regression analysis

Plotgraph is an X–Y graph plotting tool. It supports an unlimited number of data points (limited only by computer memory) and up to ten columns of data in a single data file. The module also performs least-squares regression for up to tenth order polynomials. This is a basic utility, with great flexibility for displaying a large variety of data. In addition to being able to display data from a single file, data from multiple data files can be merged into a single graph file. An example graph identifying discriminant groups (a future module) is shown in Fig. 2.

2.2. Histo: univariate statistics

Histo is used primarily for calculating and displaying univariate statistical data and up to 20 populations can be simultaneously compared. This program can individually evaluate single or multiple columns from column formatted data sets and it can evaluate the statistics of 2D and 3D gridded data files. Statistical properties such as the mean, median, variance, standard deviation, skew and kurtosis are calculated and evaluated for each data set. This and other information can be displayed in tabular form or graphically with histograms, cumulative distribution plots and box and whisker plots. For evaluating data normality (or log-normality), a Gaussian distribution bell curve, with the same mean and variance, can be displayed over the sample histograms. Probability plots can also be generated to qualitatively evaluate the normality of the data.
Fig. 1. Detailed flow chart of uncertainty analysis software package.

1. Data Input
   - Trend Surface Analysis
     - Classic Statistical Data Analysis
       • Mean, variance, standard deviation
       • Goodness of fit
       • Histograms
       • Probability Plots
       • Log-Transforms
   - Spline Contouring
   - Continuity Analysis
     - Indicator Kriging
     - Stochastic Simulations
   - Shaded Surface Modeling
     - Finite-Difference Grid Display
     - 3-D Floating Bodies
     - 3-D Block Display
   - 3-D Block Editing
2. Manual Semivariogram Analysis
3. Automated Semivariogram Analysis
   - Jackknifing and Latin-Hypercube Sampling
   - Ordinary Kriging
   - Postscript Output
3. Finite-Difference Grid Editing and Development
   - MT3D
   - MODFLOW
   - Other Applications
set or identify portions of the data set that may suggest the sample population is influenced by multiple factors.

2.3. Distcomp: sample population comparison analysis

Distcomp is a first step toward evaluating multivariate data; this package allows different data populations to be displayed together using histogram and cumulative histogram plots or to be directly compared using P–P (compare cumulative probabilities) and Q–Q (compare quantiles) plots.

2.4. Vario and variofit: semivariogram analysis

Vario and variofit allow the modeler to calculate experimental and model semivariograms, respectively, for 2D and 3D data.

Vario is used to calculate experimental semivariograms for an unlimited number of samples (the limit is determined by available computer memory) for both scattered and gridded data. For irregularly scattered data, traditional 1D semivariograms (Fig. 3) for any lag, search direction, plunge, half-angle and bandwidth

Fig. 2. Distribution of sample data and group means based on two-primary discriminant factors.

Fig. 3. 1D experimental semivariogram.
can be calculated. In addition to semivariograms, other spatial relation equations can be used: the cross-
semivariogram, covariance, correlogram, general relative semivariogram, pairwise relative semivariogram,
log semivariogram, semirodogram, semimadogram, indicator semivariogram and soft indicator covariance
(Deutsch and Journel, 1992). If the user is willing to
limit flexibility, 2D spatial equations can be calculated.
Here, the data are evaluated along 18 vectors separ-
at ed by $10^\circ$ on a user specified plane. Because of this
design, bandwidths are not applicable; half-angle's are
restricted to $5^\circ, 15^\circ, 25^\circ, \ldots, 75^\circ, 85^\circ$ and $90^\circ$ and no
covariance type equations are solved. The results can
be useful for quickly identifying principle directions of
anisotropy (Fig. 4). If gridded data are evaluated, sig-
nificant computational efficiency can be obtained, but
the parameters that can be varied are severely
restricted: (1) the band-widths and half-angles are 0.0,
(2) search direction must align exactly with grid cells
(this does not have to be axis aligned) and (3) lags
must equal grid spacing (or multiples thereof).

Variofit allows the modeler to fit model semivari-
ograms to the experimental semivariogram regardless of
how it was calculated. Several least-squares regression
techniques are available for automatically fitting the
semivariogram models, but these should only be used
to obtain approximate fits. Semivariogram model fit-
ting requires some skill and practice to properly fit and
model appropriate structures (Isaaks and Srivastava,
1989) and automatic fitting routines (these included)
often fall short of a perfect solution (Greton, 1998) for
model fitting. This module allows modelers to fit ex-
perimental semivariograms with spherical, exponential
and Gaussian models with up to four nested structures.

Fig. 4. (A) 2D experimental semivariogram and (B) 1D experimental semivariogram profile line (A–A').

Fig. 5. Example model semivariogram with experimental semivariogram lag estimates and number of sample pairs posted.
Variofit also supports two unusual features: (1) calculating jackknifed semivariograms (Wingle and Poeter, 1992) and (2) displaying ergodic fluctuations of semivariogram models from multiple simulations (Fig. 6). Calculating the experimental semivariogram and displaying the experimental semivariograms for a complete series of simulations can be useful for evaluating whether model results reflect the input parameters and the degree of variation between the different simulations in a qualitative manner.

2.5. Grid: gridding algorithms

Grid allows a modeler to interpolate between randomly scattered 2D and 3D data. Three methods are available: (1) inverse-distance, (2) trend-surface analysis (up to fifth order) and (3) simple, ordinary and universal kriging. With trend-surface analysis, the trend and residual are also evaluated at each sample location. In the kriging packages, the sample locations can be cross validated. The inverse-distance technique was included because it is fast, simple to use and in some situations it produces reasonable results. Kriging was included, because mathematically it is the best-linear-unbiased-estimator (BLUE, Clark, 1979). Trend-surface analysis was included because it is often appropriate to remove the regional trend from the data; the residuals are then kriged and the kriged results are added back to the trend surface (Fig. 7). If the trend is not removed from the data, the assumption of secondary stationarity in kriging may be violated.

Within the kriging portion of this module, there are two experimental features which allow the modeler to (1) use unique semivariogram models in different orthogonal directions (Wingle and Poeter, 1998a; the model types, ranges and sill components can all vary; most algorithms require one model and the orthogonal differences (ranges) are defined using anisotropy factors) and (2) a zonal package (Wingle and Poeter, 1996) which allows different semivariogram models to be applied and merged across the model area (Fig. 8).

2.6. Sisim and sisim3d: conditional simulation

The Sisim (the GUI) and sisim3d (the simulator) modules combine to make a discrete conditional indicator simulator. This is the only simulator included in UNCERT, but it is useful for modeling the possible distribution of material units which are important for modeling hydraulic conductivity for ground water flow and contaminant transport modeling. This package is a substantially modified version of *isim3d* by Gómez-Hernández and Srivastava (1990). This module can incorporate various types of imprecise (soft) data, use directional semivariograms and zonal kriging techniques (Fig. 8) mentioned in Section 2.5 and calculate simulations based on semivariograms of the indicators themselves, rather than on semivariograms of the thresholds between indicators. For numerical reasons, this last item better accounts for order relation violations and simplifies sensitivity analysis (Wingle and Poeter, 1998b).

Simulation is important, because instead of generating a single interpretation of site conditions, it generates multiple, unique, interpretations of site conditions that honor the sample data and mimic the random variability of the parameter of concern. This differs from kriging methods which determine the most prob-
able value at a particular location. A kriged surface represents an average, whereas simulations generate multiple possible configurations of reality. By generating multiple interpretations of the parameter of concern, particularly when combined with other modeling processes (e.g. groundwater flow), a range of probable results and statistics can be evaluated (Fig. 9), rather than a single result based on a single best estimate (examples: McKenna and Poeter, 1995; Poeter and McKenna, 1995).

2.7. Contour: contouring

Contour provides a basic 2D contouring utility for regular grids (\( \Delta x \), \( \Delta y \) and \( \Delta z \) are constants, but they may be different from each other). Although only \( X-Y \), \( X-Z \) or \( Y-Z \) planes of a grid can be viewed, it is simple to switch and step through different planes. In addition to contouring the data, line work (e.g. representing roads, buildings, formation contacts) can be drawn on top of the contours, portions of grids can be hidden, data locations and values/labels can be posted, gradients between neighboring cells can be defined by color or direction arrows and profile (transect) lines can be plotted (Fig. 4).

2.8. Surface: surface rendering

Surface, like contour is a 2D visualization utility. Instead of contouring the grid values, it treats the grid values as elevations to create an undulating surface (quasi-3D). The surface can be rendered as a mesh, a color contour map, a map displaying the gradient between neighboring cells (Fig. 10A) or a shaded relief map which mimics how a light source from a specified position would illuminate the surface (Fig. 10B). The last feature can be useful for identifying structural features such as lineaments or joints (Lillesand and Kiefer, 1987) or for identifying problem areas due to the gridding technique (Wingle, 1992).

Fig. 7. Estimates of bedrock surface: (A) was generated using second-order regional trend (B) plus the simple kriged results of the local residuals; (C) was generated using only simple kriging. Differences (D) between (A) and (C) in the central portion of the site are quite small (\( \pm 1 \) ft (0.3 m)); the differences can be extreme near the edges (180 ft. (55 m) in the Northwest corner). Here ordinary kriging generates a valley, where site conditions actually show a continued gentle downward slope (to the Northwest).
2.9. Block: scatter point and three-dimensional voxel visualization

Block is a utility for viewing 3D scatter data and 3D gridded data. Block can also be used to transform and subset 3D scatter data. When used with scatter data, line work can be displayed along with the sample data (data from multiple data sets can be merged into a single figure). The color mapping of the data can be based on the range of the sample data, an arbitrary range or can be specifically defined for each point. Instead of plotting the samples as points, bars can be

Fig. 8. Single-zone and two-zone realization pair. These models were calculated using the same ‘random’ path, but the second model used two zones where (1) the semivariograms models differed, and (2) the prior distribution of the indicators varied. Note the difference in textures and the increase in indicator 1 in the East half of the two-zone realization.
Fig. 9. Based on 200 hydraulic conductivity simulations, and ground water flow was models each realization; (A) the resulting mean hydraulic head at each location, (B) the standard deviation of hydraulic head (the East and West boundaries were constant heads), (C) 97.5% of realizations hydraulic heads were greater than shown for each cell and (D) 97.5% of realizations hydraulic heads were less than shown for each cell.

Fig. 10. Two maps or surfaces displaying the same information in unique ways: (A) a gradient map of surface and (B) 3D shaded-relief surface. Each map has a different purpose and emphasizes different attributes of the same data.
attached reflecting the magnitude of each sample (Fig. 11). Using transformation, planes and clipping options, it is also possible to divide the data set and save sample subsets to different files. Block also is used to display 3D grid data (Fig. 8). Each cell is assigned a color based on its value. Tools allow the modeler to hide all cells with a given value, make a bench cut into the model or mask certain cells regardless of their value.

2.10. Modmain and mt3dmain: ground-water flow and contaminant transport modeling

Modmain offers a complete GUI interface for MODFLOW (McDonald and Harbaugh, 1984). The data input files can be created, or input from other sources, MODFLOW can be executed and the results can be extracted from the output file and visualized (because MODFLOW does not necessarily track the Z dimension, the Z aspect is arbitrary). Much of the data are input though text fields, but any gridded information can be imported from grid files (the grid dimensions must be equal) and some point data can be interpolated from grid files estimated elsewhere. A point and click editor facilitates editing grid layer values as well as adding and deleting features such as wells, river nodes and boundary cells. The interface assumes the user is familiar with MODFLOW.

Mt3dmain is a complete GUI for MT3D (Zheng, 1990). The interface and the features are nearly identical to modmain.

2.11. Array: grid mathematics

Array is a utility for performing mathematical calculations on gridded data. These functions are comparable to many grid functions found in GIS packages (Burrough, 1986). Constants can be added, subtracted, multiplied or divided against every cell within a grid. Each cell in two grids can be added, subtracted, multiplied or divided to, from or by each other, respectively. Cells within a grid can be reclassified based on the value (e.g. all calculated concentrations less than 0.0 are assigned 0.0; negative concentrations make no physical sense). A series of grids (from a simulation ensemble for example) can be averaged, the variance can be estimated for each cell, the minimum or maximum values can be calculated, the frequency with which a cell has a value within a specified range can be

Fig. 11. MADE, the MAcroDispersion Experiment, at Columbus Air Force Base, Mississippi. Natural gamma borehole logs with well ID’s are shown with cultural (road, thick black), geomorphic (paleo-river channel, thick blue), and site features (plume well field, thick red).
determined, or a list of values for a particular cell (one value for each realization) can be extracted.

3. Compatibility with other programs

UNCERT was designed to be relatively easy to use with other programs. UNCERT will not solve all the requirements of every user; even the authors use GSLIB extensively. When applicable, UNCERT is powerful and can enhance modeling regardless of the origin of the data, or the ultimate use of the results.

Although UNCERT is easy to use with other applications all files are not directly transferable, however, the input and output file formats are well documented, simply laid out and with a little programming knowledge can be converted into an appropriate format. Some data file conversion tools are built into UNCERT already. For example, most GSLIB files can be read directly into UNCERT, GeoEAS (and GSLIB) column data files can be read without modification and so can ASCII Surfer grid files. Using array, the X, Y and Z (column, row and layer) ordering also can be independently redirected.

4. UNCERT’s future

UNCERT is a work in progress. New features continue to be added and new and old bugs are fixed. We are currently working on other projects using Java and will be converting many of the UNCERT utilities to this language. The advantage of Java is that it is platform independent (Macintosh and MS-Windows users will be able to use the packages); the disadvantage is that it is slower than conventional compiled languages such as C, C++ or FORTRAN. A great deal of research is underway with Java and the performance differences are becoming less significant, due to virtual Java computers and Just In Time (JIT) compilers. It is also possible to leave computationally expensive tasks in C and if carefully handled, cross-platform compatibility may not be compromised.

We continue to conduct research in geostatistics, multivariate analysis and inversion. As tools from these activities become available and are tested, they will be merged into UNCERT.

5. Conclusions

UNCERT is a collection of tools designed to allow a modeler to perform all the steps required to evaluate site data, perform statistical and geostatistical analysis and execute ground-water flow and contaminant transport models for risk assessment and remediation design; with an easy to use GUI, with powerful visualization tools. Although UNCERT has all the necessary components, modelers will often want to substitute or add other tools for specific tasks, therefore the data structures are basic, facilitating data transfer between applications.

The utility of UNCERT comes from several basic features: (1) the modules were written so that the only restrictions (with a few exceptions) on file sizes are the memory of the computer being used; (2) the consistent graphical user interface simplifies data handling and model setup for users; (3) scripting options and preference files make it easy to evaluate large or complex models and reproduce work; (4) computer graphics facilitate result visualization; (5) Postscript output makes it simple to present work and (6) because there is no charge for the software, it is readily available to students, consultants and corporations of any size. The packages are written in ANSI-C and FORTRAN77 for UNIX systems and have proven to be portable between UNIX platforms. Work is in progress to transport these tools to the Microsoft Windows and Macintosh environments.

Acknowledgements

This work was supported by the US Bureau of Reclamation, the US Army Corp of Engineers Waterways Experiment Station and Sandia National Laboratories.

Appendix A. Acquiring software, demo data files and documentation

The UNCERT software package and users manual in HTML and PDF formats are available from http://uncert.mines.edu/ or ftp://uncert.mines.edu/pub/uncert/. If you are using a World Wide Web browser, type in the http addresses above as shown. If you are using a ftp program, connect to uncert.mines.edu. Use the login name anonymous and use your e-mail address for the password. The files are kept in the directory/pub/uncert/.

References


