ELSEVIER

Computers & Geosciences 30 (2004) 221–238

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

GEOSSAV: a simulation tool for subsurface applications ☆

Christian Regli[a,*], Lukas Rosenthaler[b], Peter Huggenberger[a]

[a] *Department of Geosciences, Applied and Environmental Geology, University of Basel, Bernoullistr. 16, 4056 Basel, Switzerland*
[b] *Imaging & Media Lab, University of Basel, Bernoullistr. 32, 4056 Basel, Switzerland*

Received 3 December 2002; received in revised form 23 October 2003; accepted 30 October 2003

Abstract

Geostatistical Environment fOr Subsurface Simulation And Visualization (GEOSSAV) is a tool for the integration of hard and soft data into stochastic simulation and visualization of distributions of geological structures and hydrogeological properties in the subsurface. GEOSSAV, as an interface to selected geostatistical modules (bicalib, gamv, vargplt, and sisim) from the Geostatistical Software LIBrary, GSLIB (GSLIB: Geostatistical Software Library and User's Guide, 2nd Edition, Oxford University Press, Oxford, 1998, 369pp), can be used for data analysis, variogram computation of regularly or irregularly spaced data, and sequential indicator simulation of subsurface heterogeneities. Sequential indicator simulation, based on various kriging techniques (simple, ordinary, and Bayesian), is suitable for the simulation of continuous variables such as hydraulic conductivity of an aquifer or chemical concentrations at a contaminated site, and categorical variables which indicate the presence or absence of a particular lithofacies. The software integration platform and development environment of GEOSSAV is Tool command language (Tcl) with its graphical user interface, Toolkit (Tk), and a number of Tcl/Tk extensions. The standard Open Graphics Library application programming interface is used for rendering three-dimensional (3D) data distributions and for slicing perpendicular to the main coordinate axis. Export options for finite-difference groundwater models allow either files that characterize single model layers (which are saved in ASCII matrix format) or files that characterize the complete 3D flow model setup for MODFLOW-based groundwater simulation systems (which are saved in block-centered flow package files (User's documentation for MODFLOW-96, an update to the US Geological Survey modular finite-difference ground-water flow model, Geological Survey Open-File Report 96-485, Reston, VA, 1996, 56pp)). GEOSSAV can be used whenever stochastic solutions are preferred to solve site-specific heterogeneity problems, e.g., in the field of hydrology, groundwater, groundwater and/or soil contamination, site remediation, air pollution, and ecology. An example from the Rhine/Wiese aquifer near Basel demonstrates the application of GEOSSAV on geostatistical data analysis and subsurface visualization. GEOSSAV has been successfully tested on Microsoft Windows NT 4.0/2000/XP and on SuSE Linux 7.3. The current version is available at http://www.unibas.ch/earth/pract.

Keywords: Software; Geostatistics; Soft kriging; Variogram modeling; Sequential indicator simulation; Heterogeneity; Aquifer stratigraphy

1. Introduction

Numerical models of physical systems play an important role in decision-making processes, especially in the context of better characterization of parameter distributions and prediction of dynamic behavior of a given system. In the earth sciences and in many other research disciplines, great efforts have been made within

☆ Code available from server at http://www.iamg.org/CGEditor/index.htm

*Corresponding author. Tel.: +41-61-267-3447; fax: +41-61-267-2998.

E-mail address: christian.regli@unibas.ch (C. Regli).

0098-3004/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2003.10.007

the last 20 years on evaluation and integration of data in the characterization of the subsurface (among others Wingle et al., 1997; Deutsch and Journel, 1998). One of the objectives of all these projects is to model reasonable variations in the subsurface while constraining results as much as possible with available data. Although one particular data set may suggest a wide range of alternatives, if all the available data are combined, the possible solution population should be greatly reduced

Data used in subsurface models may be divided into two basic types: 'hard data' and 'soft data' (Poeter and McKenna, 1995). Hard data can be directly obtained and relate directly to the phenomenon being modeled. Examples of hard data include outcrop data and, in some cases, drill-core data because these explicitly define sedimentary structure types (in the following called structure types or lithofacies types) and properties. There is measurement or interpretation uncertainty in hard data, but it is considered small enough to be ignored. Soft data are less precise and generally indirect, thus greater uncertainties are associated with the soft data values. Soft data may not be honored exactly in conditional simulations (e.g., Wingle et al., 1997).

Ground penetrating radar (GPR or georadar) data are an example of soft data. Georadar systems measure the dielectric constant of the subsurface (Sensors & Software Inc., 1993), which is dependent on the water content of the sedimentary structure types. The formation of reflection patterns is dependent on the reflection coefficient at contrast surfaces between different structure types (Huggenberger, 1993). Generally, the interpretation of georadar data is based on reflection pattern analysis and the applied georadar system configuration (Huggenberger, 1993; Beres et al., 1999). Therefore, only imprecise estimates can be made about structure type, structure properties, and location (Regli et al., 2002). As shown in another example, often hydraulic conductivity is not measured directly but is deduced from structure characteristics such as grain size distribution data (e.g., Sudicky, 1986; Hess et al., 1992) or from pumping tests (Furger, 1990).

Subsurface heterogeneity is one of several important factors (e.g., boundary conditions and groundwater recharge) for modeling groundwater flow. In practice, the subsurface heterogeneity is often underestimated. In the following discussion, only the subsurface heterogeneity is considered. All other factors influencing groundwater flow are not taken into account.

Accurate modeling of subsurface parameter distributions becomes more difficult with increasing heterogeneity and uncertainty with respect to spatial variability of available data. The uncertainty depends both on the quantity and the quality of available data. The technique used to model subsurface structures for a site-specific problem should be chosen specifically for the properties under consideration (e.g., lithofacies type, hydraulic conductivity, and porosity), the knowledge of the subsurface, and the causes of uncertainty (Ayyub and Gupta, 1997; Weissmann et al., 1999). For example, in a study area where only two structure types are present, relatively few alternative interpretations may be likely based on the raw data. However, although each model realization of parameter distributions may honor the data exactly, subsequent process modeling with each alternative realization (e.g., groundwater flow and transport modeling) can generate significantly different outcomes (Poeter and McKenna, 1995). In more heterogeneous media, alternative interpretations are possible and this increases the number of spatial parameter distributions that may be generated to honor the data.

In order to evaluate this inherent uncertainty, computers can be used to create multiple alternative realizations of the subsurface. Once multiple realizations are created, output from subsequent modeling (e.g., groundwater flow and transport modeling) can be used to compare modeled and field conditions. If a groundwater model response clearly does not match field observations, the possible subsurface configuration can be rejected even though it may satisfy statistical requirements of data distributions (Poeter and McKenna, 1995). Of the remaining realizations that appear reasonable, the modeled distributions may be further evaluated for the conditions already modeled or for future conditions that may be added. With computerized routines, a sufficient number of realizations may be evaluated so that a statistically robust assessment of the reasonable alternatives may be achieved (Wingle et al., 1997).

Stochastic simulation implies sampling from conditional distributions and consequently the spatial models are samples from a multivariate distribution characterizing the spatial phenomenon. Until now it was relatively inconvenient to perform different steps of generating and visualizing realizations of subsurface heterogeneities that can be easily exported into existing groundwater simulation systems. Among others this is caused: (1) by user-unfriendly aspects of codes (e.g., GSLIB codes), (2) due to the fact that the different steps of subsurface modeling have to be completed with several independent tools (e.g., SAGE2001 for variogram calculation and modeling [1], GSLIB codes for subsurface simulation, VTK for visualization (Schroeder et al., 2001)), and (3) by the different export routines which are needed to transform modeled parameter distributions in data formats usable in subsequent modeling systems. In addition, many of these tools are platform dependent (e.g., UNCERT, Wingle et al., 1997).

Therefore, a tool was developed that combines geostatistical analysis, simulation, visualization, and data export. In this paper, a user-friendly tool for

subsurface simulation and visualization, Geostatistical Environment fOr Subsurface Simulation And Visualization (GEOSSAV), is described. GEOSSAV is a software package developed to aid hydrogeologists using geostatistics to simulate and visualize the distribution of structure types and properties in the subsurface. This package is developed in a way that spatial data from other research disciplines may be analyzed (e.g., groundwater, groundwater and/or soil contamination, site remediation, air pollution, and ecology). It allows the modeler to: (1) import hard and soft field data or data from a pre-existing database, (2) update soft data using a bivariate calibration method (Deutsch and Journel, 1998), (3) compute and visualize spatial variability of data by variogram analysis, (4) generate distributions of structure types and structure properties of the subsurface using sequential indicator simulation based on a choice of kriging techniques (simple, ordinary, and Bayesian), and (5) visualize the data and the modeled spatial distributions in two and three dimensions by three-dimensional (3D) rendering and slicing perpendicular to the main coordinate axis. Furthermore, once spatial distributions of structure types and structure properties have been generated, data files for 2D and 3D finite-difference groundwater flow and transport models can be created.

This paper starts by describing the integration platform and software resources. In the following sections the central features of GEOSSAV are described, including an account of the geostatistical techniques, the visualization methods, and the data export options. Subsequently, an example is given to illustrate site-specific considerations of heterogeneity in subsurface modeling. This paper concludes with a description of hardware and software requirements for running GEOSSAV, planned new developments, and information for acquiring GEOSSAV.

2. Integration platform and software resources

GEOSSAV consists of an integration platform and individual simulation, visualization, and export modules, as shown in Fig. 1. The software integration platform and development environment for GEOSSAV is Tool command language (Tcl) and its graphical user interface Toolkit (Tk, Ousterhout, 1994). In addition, some of the Tcl/Tk extensions (Harrison, 1997) such as [incr Tcl], [incr Tk], [incr Widgets], and TkTable are integrated into GEOSSAV. Tcl/Tk was chosen because of its speed of use, breadth of functionality, flexibility for cross-platform deployment, and ease of integrating new extensions such as rendering 3D graphics through the Open Graphics Library (OpenGL) application programming interface (API). In addition, Tcl/Tk and its extensions allow the integration of diverse software resources irrespective of the programming language (e.g., Fortran, C/C++) in which they are written.

Compiled GSLIB modules without any enhancements are integrated into GEOSSAV: The sequential indicator simulation module 'sisim' (Deutsch and Journel, 1998) requires information about the spatial variability of the regularly or irregularly spaced data, which can be computed with the variogram module 'gamv' and then visualized with the module 'vargplt' and a PostScript display device (Deutsch and Journel, 1998). The indicator kriging approach allows the modeling of single property classes, represented by indicators, using single indicator variograms. In addition, the sisim algorithm is able to account for soft data. The integration of soft data in the indicator formalism is made possible by the Markov–Bayes option for cokriging using the module 'bicalib' (Deutsch and Journel, 1998).

The open, standardized API OpenGL (Fosner, 1997; Wright and Sweet, 2000) has been integrated in the Tcl/Tk environment and is used for rendering of the 3D data distributions and for slicing perpendicular to the main coordinate axis. Export formats are compatible with widely used finite-difference groundwater modeling environments (e.g., ASMWIN, Chiang et al., 1998; GMS, [2]; PMWIN, Chiang and Kinzelbach, 2001). Also data may be exported as individual files for single model layers (saved in ASCII matrix files) or as data files characterizing the complete 3D model setup for MODFLOW-based groundwater simulation systems and saved in block-centered flow (bcf) package files (Harbaugh and McDonald, 1996). Groundwater flow and transport simulation is external to GEOSSAV.

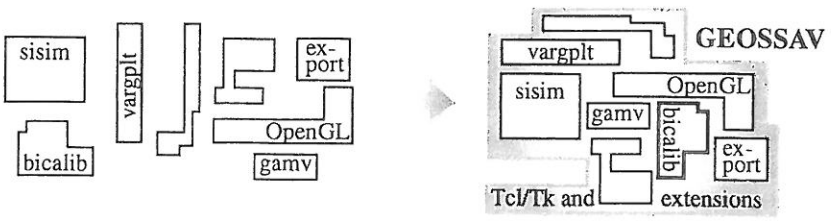

Fig. 1. Various modules integrated within GEOSSAV. Gamv computes spatial variability of regularly or irregularly spaced data. Sisim is used for sequential indicator simulation of categorical and continuous variables, and bicalib is used for considering soft data in indicator formalism. OpenGL supports visualization of data and results in two and three dimensions. Export options allow files to be generated in data formats for use in other applications.

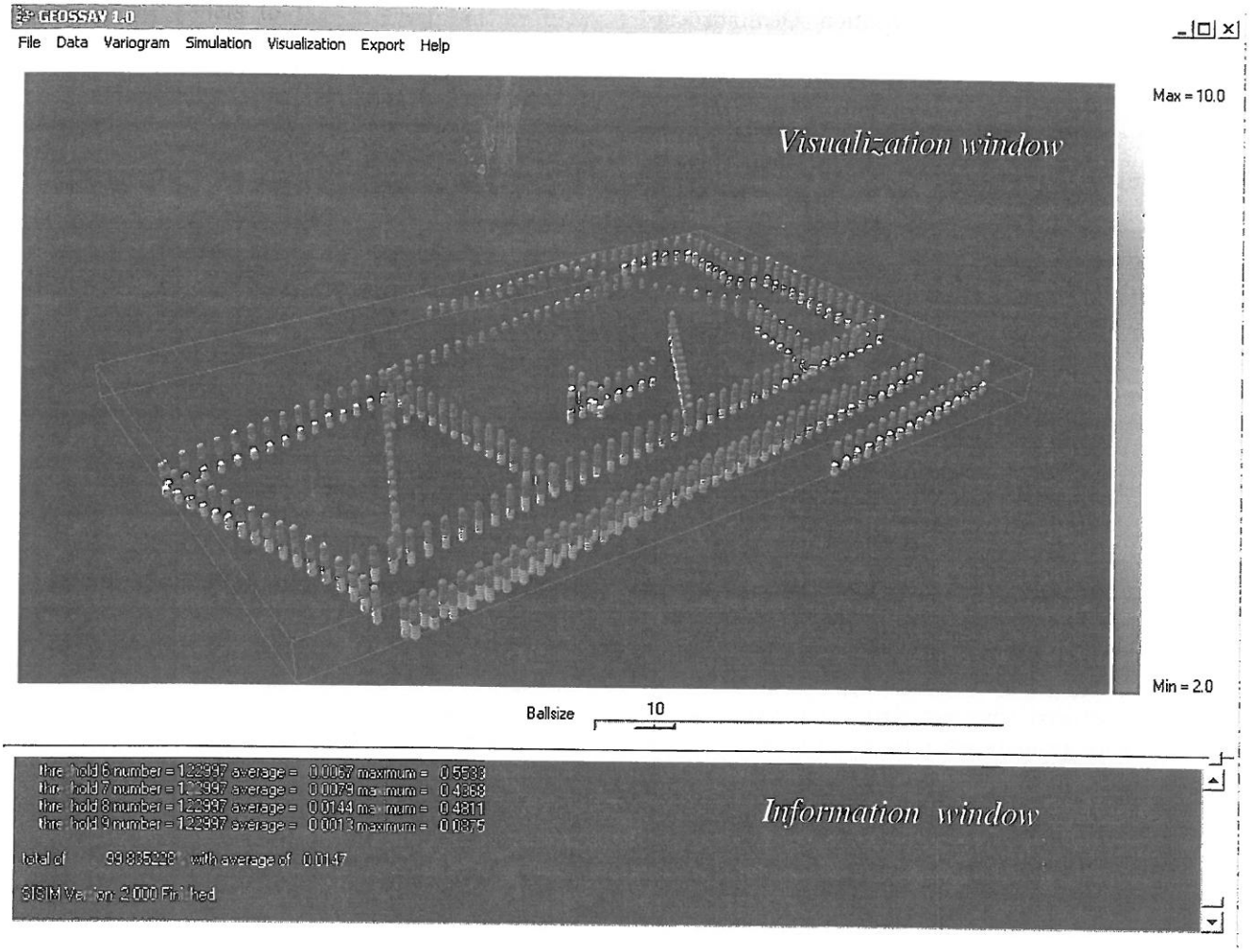

Fig. 2. Main GEOSSAV window with menu bar for pull-down menus, window for visualization of input and output data, and window for error checking.

Fig. 2 shows the main GEOSSAV window. The upper area is the visualization window where the input and output data are shown. Here all the hard and soft input data are displayed for error checking and the output data are shown for plausibility checking. The lower black area is an information window where the standard output of all modules is shown. That is, all of the output that would be seen at the prompt is shown here for error checking. The single modules (e.g., for data processing, simulation, visualization, final control, and data export) are placed in the corresponding pull-down menus. The 'File' menu controls the project settings. The 'Data' menu is for processing the input data (e.g., bivariate calibration of soft data). The 'Variogram', 'Simulation', 'Visualization', 'Export', and 'Help' menus are self-explanatory. When selecting a module from the pull-down menus, a corresponding dialog box for parameter input is opened (e.g., Figs. 4–7 and 9). The dialog box is designed to require the input of all options, which are necessary for running the module. The parameters can be written to a parameter file and the module can be executed.

3. Geostatistical techniques

3.1. Stochastic simulation

Stochastic simulation is the process of drawing multiple, equally probable realizations of random variables (RVs) from a random function (RF) model. The realizations represent possible images of spatial distributions of the data values $z(\boldsymbol{u})$ over the field A. Each realization reflects the spatial properties that have been imposed on the RF model $z(\boldsymbol{u})$. The more properties are inferred from the sample data and integrated in the RF model through the simulation algorithm, the higher the accuracy will be of the RF model and the resultant simulated realizations in representing the heterogeneity of the spatial phenomenon (Deutsch and

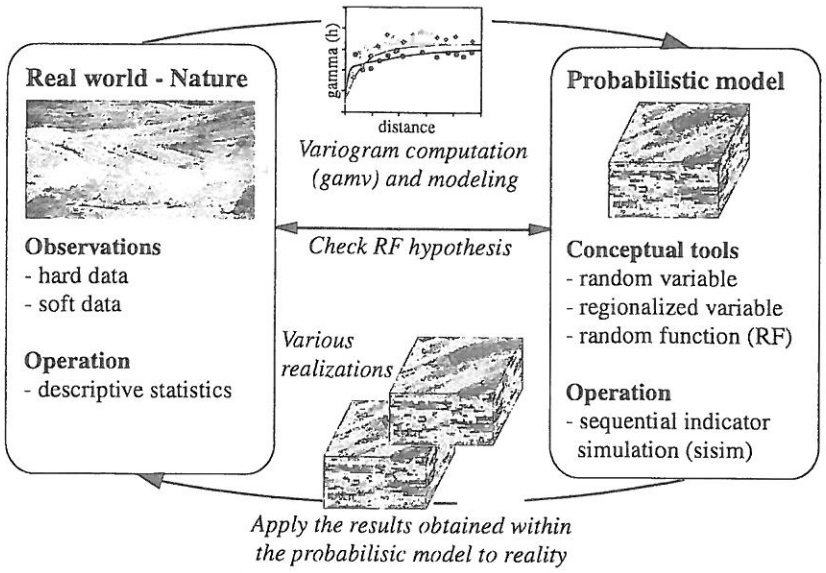

Fig. 3. Conceptual framework of stochastic simulation approach, modified from Pannatier (1996). Based on variogram modeling of hard input data and following stochastic simulation (e.g., sequential indicator simulation), various equiprobable realizations are generated and checked for consistency with RF hypothesis. Results obtained within probabilistic model should be checked by comparison of measured and calculated data.

Journel, 1998). The conceptual framework of the stochastic simulation approach is shown in Fig. 3. Based on the variogram computation of the hard input data and the following stochastic simulation, e.g., sequential indicator simulation, various equally probable realizations are generated, which represent images of the reality. The simulation is conditional if the resulting realizations honor data values at their locations.

The various realizations have to reproduce the input proportions within a defined accuracy: If the cumulative distribution functions (cdf) or the probability density functions (pdf) of the input data are significantly different from those of the resulting probabilistic model, then this possible subsurface configuration has to be rejected. Moreover, the accepted realizations should be applied to reality to compare measured and calculated data. If a model response clearly does not match field observations, the possible subsurface configuration has to be rejected. On the other hand, if a model response matches field observations, the possible subsurface configuration represents a probable realization of the reality. The different steps of stochastic simulation are described in Sections 3.2 and 3.3.

3.2. Variogram computation of irregularly spaced data

Modeling spatial variability of data is the key to any subsurface simulation. The variogram describes the spatial correlation of data as a function of the separation vector $\boldsymbol{h}$ (lag) between two data points. The indicator variogram for continuous variables such as hydraulic conductivity of an aquifer or concentrations over a contaminated site, or categorical variables, indicating the presence or absence of a particular lithofacies, is computed on a constructed indicator variable. This requires either the specification of a continuous variable and cutoff or a categorical variable and category to create the indicator transform. For the cutoff z_k (threshold) and data value $z(\boldsymbol{u}_\alpha)$ or the category s_k and data value $s(\boldsymbol{u}_\alpha)$, the indicator transform $i(\boldsymbol{u}_\alpha; z_k \text{ or } s_k)$ is defined as (Deutsch and Journel, 1998)

$$i(\boldsymbol{u}_\alpha; z_k \text{ or } s_k) = \begin{cases} 1 & \text{if } z(\boldsymbol{u}_\alpha) \leqslant z_k \text{ (continuous variable), or 1,} \\ & \text{if } s(\boldsymbol{u}_\alpha) = \leqslant s_k \text{ (categorical variable),} \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $\boldsymbol{u}_\alpha$ refers to a particular data location.

The indicator variogram, written for a specific category, is defined as half of the average squared increment between two indicators separated by $\boldsymbol{h}$ (Deutsch and Journel, 1998):

$$\gamma_I(\boldsymbol{h}; s_k) = \frac{1}{2N(\boldsymbol{h})} \sum_{1}^{N(h)} [i(\boldsymbol{u}; s_k) - i(\boldsymbol{u} + \boldsymbol{h}; s_k)]^2, \quad (2)$$

where $N(\boldsymbol{h})$ is the number of pairs, $i(\boldsymbol{u}; s_k)$ is the indicator at the start or tail of the pair, and $i(\boldsymbol{u} + \boldsymbol{h}; s_k)$ is the corresponding end or head indicator.

Experimental variograms derived on the basis of data have to be fitted using licit variogram models in order to satisfy the requirements. The variogram models are demanded in the sequential indicator simulation module for determining kriging weights. Several terms are used

to describe variogram models: the sill refers to the maximum value of the variogram model, the range corresponds to the maximum distance of spatial correlation, and the nugget refers to the behavior of the variogram model at lags $\boldsymbol{h} \rightarrow 0$. A variogram model consists of an isotropic or anisotropic nugget effect, a positive definite variogram structure (spherical model, exponential model, Gaussian model, power model, and hole effect model), and parameters defining the geometric anisotropy.

The module gamv from Deutsch and Journel (1998) can be used for spatial data analysis and variogram computation of irregularly spaced data in three dimensions. In fact, 10 experimental measures of spatial variability can be computed (variogram, cross variogram, covariance, correlogram, general relative variogram, pairwise relative variogram, variogram of logarithms, madogram, indicator variogram for continuous variables, and indicator variogram for categorical variables).

3.3. Variogram implementation

Fig. 4 shows the dialog box for the module gamv 2.000 in GEOSSAV. The first register (top panel of Fig. 4) is used for data input {*Input*}, data output {*Output*}, and lag specifications {*Lags*}. In general, the separation vector $\boldsymbol{h}$ is specified with some direction and distance tolerance. Experience shows that many directional sample variograms are required to obtain a reasonably representative model of 3D spatial variability.

The second register (bottom panel of Fig. 4) is used for variogram specifications {*Variograms*}. The module can handle many different directions, cutoffs or categories, and variogram types in a single pass. The azimuth is measured clockwise from North, and the dip is measured in negative degrees down from horizontal. Angular half-window tolerances are required for both the azimuth and the dip (Deutsch and Journel, 1998, p. 49, Fig. III.2). These tolerances may overlap, causing pairs to report to more than one direction and lag vector. The angle tolerances are restricted once the deviations from the direction vector exceed the horizontal or the vertical bandwidth. The results of the variogram computation are written into an output file.

The module vargplt from Deutsch and Journel (1998) takes the output format used by gamv and creates graphical displays for PostScript display devices. This module provides no facility for interactive model variogram fitting. Based on the experimental variogram, parameters of a model variogram have to be determined (e.g., with VARIOWIN from Pannatier, 1996; SAGE2001 from [1]).

Fig. 5 shows the dialog box for the module vargplt 2.000 in GEOSSAV, calling for a plot title, distance and variogram limits, as well as files containing the

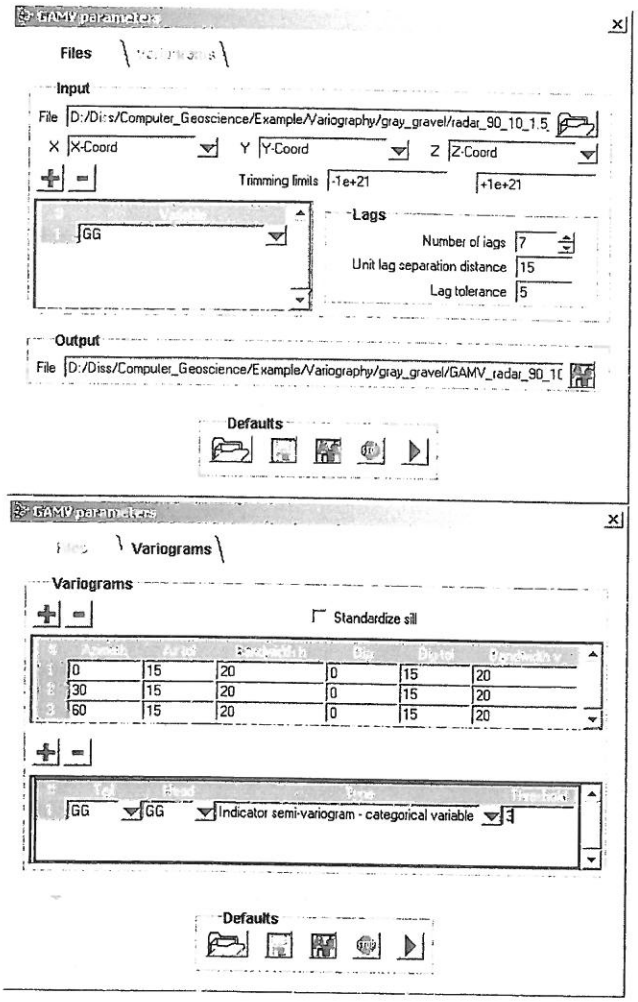

Fig. 4. Dialog box for irregularly spaced data variogram computation, used for data input and output as well as for lag and variogram specification (directions, variables, and variogram types).

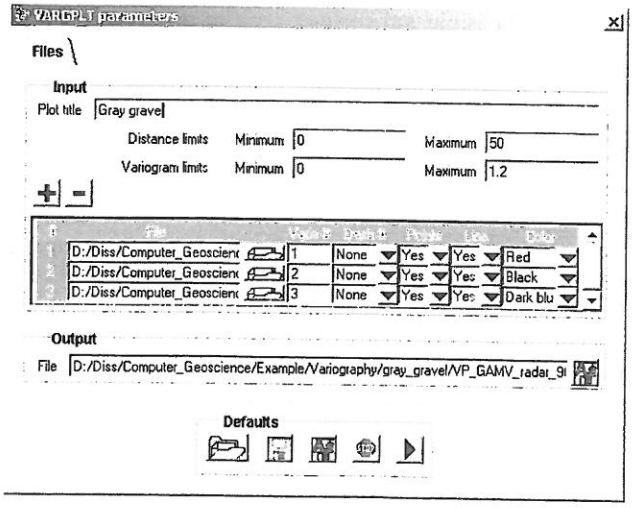

Fig. 5. Dialog box for variogram plotting, used for specifying and creating graphical displays for PostScript display devices.

calculated experimental variograms {*Input*}. For every single variogram, display specifications can be indicated. The graphical displays are written into an output file {*Output*} that can be visualized in public PostScript viewing program (e.g., GSview).

3.4. Sequential indicator simulation

Sequential indicator simulation is suitable for the simulation of continuous and integer-coded categorical variables. Simple indicator kriging (SIK) or ordinary indicator kriging (OIK) can be selected for simulation. The indicator kriging process is repeated for a series of K cutoffs z_k, $k = 1, \ldots, K$, which discretize intervals of variability of the continuous variable, or for a series of K categories s_k, $k = 1, \ldots, K$, which represent different lithofacies types. The cdf (working with continuous variables), built by assembling the K indicator kriging estimates, represents a probabilistic model for the uncertainty about the non-sampled value $z(\boldsymbol{u})$. In the case of categorical variables, the pdf corresponding to each category represents the uncertainty about the non-sampled category $s(\boldsymbol{u})$. For the categorical variable $s(\boldsymbol{u})$, e.g., set to 1 if a specific lithofacies type prevails at location $\boldsymbol{u}$, set to 0 if not, the indicator kriging of $s(\boldsymbol{u})$ provides a model for the probability that $s(\boldsymbol{u})$ be one. For the continuous variable $z(\boldsymbol{u})$ the correct selection of the cutoffs z_k, at which indicator kriging takes place, is essential. By defining too many cutoffs or categories, the computation becomes needlessly tedious, by marking too few, the details of the distribution are lost. The indicator methodology is a non-parametric approach to creating a useful distribution for modeling distributions, which can only be described inadequately by the mean and the variance of the data due to their polymodal form (Schafmeister, 1999).

For a specific category, the SIK estimate, i.e. the probability that s_k prevails at location $\boldsymbol{u}$, is written as a linear combination of the n nearby indicator-coded data (Deutsch and Journel, 1998):

$$[i(\boldsymbol{u}; s_k)]^*_{SK} = [Prob\{S(\boldsymbol{u}) = s_k | (n)\}]^*_{SK} = \sum_{\alpha=1}^{n} \lambda_\alpha(\boldsymbol{u}; s_k) i(\boldsymbol{u}_\alpha; s_k) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha(\boldsymbol{u}; s_k)\right] F(s_k), \quad (3)$$

where $F(s_k)$ is the stationary prior probability of category s_k, and the $\lambda_\alpha(\boldsymbol{u}; s_k)$ are the SIK weights corresponding to category s_k, which depend on the closeness of the data considered for the estimation. These weights are given by the SIK system (Deutsch and Journel, 1998):

$$\sum_{\beta=1}^{n} \lambda_\beta(\boldsymbol{u}; s_k) C_I(\boldsymbol{u}_\beta - \boldsymbol{u}_\alpha; s_k) = C_I(\boldsymbol{u} - \boldsymbol{u}_\alpha; s_k), \quad \alpha = 1, \ldots, n, \quad (4)$$

$C_I(\boldsymbol{h}; s_k) = Cov\{I(\boldsymbol{u}; s_k), I(\boldsymbol{u} + \boldsymbol{h}; s_k)\}$ are the requisite indicator covariances. The variogram models are converted into equivalent covariance models, because the kriging system is more easily solved with covariance matrices. For K categories s_k, SIK requires K indicator covariances $C_I(\boldsymbol{h}; s_k)$ in addition to the K pdf values $F(s_k)$.

OIK is the most commonly used variant of the SIK algorithm, whereby the sum of the kriging weights is constrained equal to 1 (OIK constraint). If data are abundant, OIK within moving data neighborhoods may be considered; this amounts to re-estimating locally the prior pdf values $F(s_k)$ (Deutsch and Journel, 1998).

Median indicator kriging calls for a single indicator variogram that is used for all K categories. It is used if the sample indicator variograms appear proportional to each other. Only one single indicator kriging system needs to be solved with the resulting weights being used for all categories. It is, therefore, a particularly simple and fast procedure.

The major advantage of the indicator kriging approach to generate posterior conditional distributions is its ability to account for soft data. As long as the soft data can be coded into prior local probability values, indicator kriging can be used to integrate that information into a posterior probability value. The prior information can take one of the following forms (Deutsch and Journel, 1998):

(1) Local hard indicator data $i(\boldsymbol{u}_\alpha; z_k)$ or $i(\boldsymbol{u}_\alpha; s_k)$ originating from local hard data $z(\boldsymbol{u}_\alpha)$ or $s(\boldsymbol{u}_\alpha)$, respectively:

$i(\boldsymbol{u}_\alpha; z_k)$
$= 1$, if $z(\boldsymbol{u}_\alpha) \leqslant z_k$, $= 0$ if not (continuous variables) or
$i(\boldsymbol{u}_\alpha; s_k)$
$= 1$, if $s(\boldsymbol{u}_\alpha) = \leqslant s_k$, $= 0$ if not (categorial variables). (5)

(2) Local hard indicator data $j(\boldsymbol{u}_\alpha; z_k)$ originating from ancillary information that provides hard inequality constraints on the local value $z(\boldsymbol{u}_\alpha)$. This type of prior information is valid only for continuous variables. If $z(\boldsymbol{u}_\alpha) \in (a_a, b_b)$, then

$$j(\boldsymbol{u}_\alpha; z_k) = \begin{cases} 0 & \text{if } z(\boldsymbol{u}_\alpha) \leqslant a_\alpha, \\ \text{undefined (missing indicator data)} & \text{if } z(\boldsymbol{u}_\alpha) \in (a_a, b_b), \\ 1 & \text{if } z(\boldsymbol{u}_\alpha) > b_\alpha. \end{cases} \quad (6)$$

(3) Local soft indicator data $y(\boldsymbol{u}_\alpha; z_k)$ or $y(\boldsymbol{u}_\alpha; s_k)$ originating from ancillary information providing prior probabilities of the value $z(\boldsymbol{u}_\alpha)$ or the category $s(\boldsymbol{u}_\alpha)$, respectively:

$$y(\boldsymbol{u}_\alpha; z_k) = Prob\{Z(\boldsymbol{u}_\alpha) \leqslant z_k | \textit{local ancillary information}\}, \in [0, 1] \text{ (continuous variables) or}$$

$$y(\boldsymbol{u}_\alpha; s_k) = Prob\{S(\boldsymbol{u}_\alpha) = s_k | \textit{local ancillary information}\}, \in [0, 1] \text{ (categorial variables).} \quad (7)$$

(4) Global prior information common to all locations $\boldsymbol{u}$ within the area A:

$$F(z_k) = Prob\{Z(\boldsymbol{u}) \leqslant z_k\}, \quad \forall \boldsymbol{u} \in A \text{ (continuous variables) or}$$

$$F(s_k) = Prob\{S(\boldsymbol{u}) = s_k\}, \quad \forall \boldsymbol{u} \in A \text{ (categorial variables).} \quad (8)$$

At any location $\boldsymbol{u} \in A$, prior information about the value $z(\boldsymbol{u})$ or the category $s(\boldsymbol{u})$ is characterized by any one of the four previous types of prior information. The indicator kriging process consists of a Bayesian updating of the local prior cdf or pdf into a posterior cdf or pdf using information supplied by neighboring local prior cdfs or pdfs, written for a specific category (Deutsch and Journel, 1998):

$$[Prob\{S(\boldsymbol{u}) = s_k | (n + n')\}]^*_{SK} = \lambda_0(\boldsymbol{u}) F(s_k) + \sum_{\alpha=1}^{n} \lambda_\alpha(\boldsymbol{u}; s_k) i(\boldsymbol{u}_\alpha; s_k) + \sum_{\alpha'=1}^{n'} \upsilon_{\alpha'}(\boldsymbol{u}; s_k) y(\boldsymbol{u}'_a; s_k). \quad (9)$$

The $\lambda_\alpha(\boldsymbol{u}; s_k)$ are the weights attached to the n neighboring hard indicator data of Eq. (5), the $\upsilon_{\alpha'}(\boldsymbol{u}; s_k)$ are the weights attached to the n' neighboring soft indicator data of Eq. (7), and λ_0 is the weight attributed to the global prior pdf. To ensure unbiasedness, λ_0 is usually set to (Deutsch and Journel, 1998):

$$\lambda_0(\boldsymbol{u}) = 1 - \sum_{\alpha=1}^{n} \lambda_\alpha(\boldsymbol{u}; s_k) - \sum_{\alpha'=1}^{n'} \upsilon_{\alpha'}(\boldsymbol{u}; s_k). \quad (10)$$

The conditional pdf model of Eq. (9) can be seen as an indicator cokriging that pools information of different types: the hard i indicator data and the soft y prior probabilities. If the soft information is not present or is ignored ($n' = 0$), Eq. (9) reverts to the SIK of Eq. (3).

3.5. Bivariate calibration of soft data

Soft information can be considered. If the spatial variability of the soft variable y is represented using the same covariance model $C_I(\boldsymbol{h}; s_k)$ as the indicator hard variable i, no posterior updating of prior probability values $y(\boldsymbol{u}'_a; s_k)$ at soft data locations $\boldsymbol{u}'_a$ is possible, i.e. the soft data are also treated as hard data.

Most often, the soft data originate from information related to, but different from the hard data. Thus, the soft y indicator spatial distribution is likely different from that of the hard i indicator data. The Markov–Bayes algorithm (see Eq. (9)) should be considered in order to perform full updating of all prior pdfs that are not already hard. The soft indicator data covariance and cross-covariance for a specific category are calibrated from the hard indicator covariance model, whereby (Deutsch and Journel, 1998):

$$C_{IY}(\boldsymbol{h}; s_k) = B(s_k) C_I(\boldsymbol{h}; s_k), \quad \forall \boldsymbol{h}, \quad (11)$$

$$C_Y(\boldsymbol{h}; s_k) = B^2(s_k) C_I(\boldsymbol{h}; s_k), \quad \forall \boldsymbol{h} > 0,$$

$$C_Y(\boldsymbol{h}; s_k) = |B(s_k)| C_I(\boldsymbol{h}; s_k), \quad \boldsymbol{h} = 0.$$

The coefficients $B(s_k)$ corresponding to each secondary data s_k is obtained from calibration of the soft y data to the hard i data (Deutsch and Journel, 1998):

$$B(s_k) = m^{(1)}(s_k) - m^{(0)}(s_k) \in [-1, +1] \quad (12)$$

with

$$m^{(1)}(s_k) = E\{Y(\boldsymbol{u}; s_k) | I(\boldsymbol{u}; s_k) = 1\},$$

$$m^{(0)}(s_k) = E\{Y(\boldsymbol{u}; s_k) | I(\boldsymbol{u}; s_k) = 0\}.$$

Consider a calibration data set $\{y(\boldsymbol{u}_\alpha; s_k), i(\boldsymbol{u}_\alpha; s_k), \alpha = 1, \ldots, n\}$, where the soft probabilities $y(\boldsymbol{u}_\alpha; s_k)$ valued in $[0, 1]$ are compared to the actual hard values $i(\boldsymbol{u}_\alpha; s_k)$ valued 0 or 1. $m^{(1)}(s_k)$ is the mean of the y values corresponding to $i = 1$; the best situation is when $m^{(1)}(s_k) = 1$, that is, when all y values exactly predict the outcome $i = 1$. Similarly, $m^{(0)}(s_k)$ is the mean of the y values corresponding to $i = 0$, best being when $m^{(0)}(s_k) = 0$.

The parameter $B(s_k)$ measures how well the soft y data separates the two actual cases $i = 1$ and $i = 0$. The best case is when $B(s_k) = \pm 1$, and the worst case is when $B(s_k) = 0$; that is, $m^{(1)}(s_k) = m^{(0)}(s_k)$.

The case $B(s_k) = -1$ corresponds to soft data predictably wrong and is best handled by correcting the wrong probabilities $y(\boldsymbol{u}_\alpha; s_k)$ into $1 - y(\boldsymbol{u}_\alpha; s_k)$.

If $B(s_k) = 1$, the soft prior probability data $y(\boldsymbol{u}'_\alpha; s_k)$ in Eq. (9) is treated as hard indicator data and therefore not updated. Conversely, if $B(s_k) = 0$, the soft data $y(\boldsymbol{u}'_\alpha; s_k)$ are ignored; i.e. their weights in Eq. (9) become zero.

The sequential indicator simulation principle is an extension of kriging to include all data available within a neighborhood of the location $\boldsymbol{u}$, including the original data and all previously simulated values. The steps in the sequential indicator simulation are as follows:

- In the first step, a grid network and coordinate system is established.

- In the second step, the data are assigned to the nearest grid node. If there are multiple data, only the closest data are assigned to the nearest grid node.
- In the third step, a random path through all grid nodes is determined. For a node in the random path:
  (1) The nearby data and previously simulated grid nodes are searched.
  (2) The conditional distribution is estimated by indicator kriging (Eq. (9)).
  (3) From this distribution a simulated value or category, respectively, is randomly drawn and set as hard data for the simulation at the next node. The next node in the random path is selected and the steps (1)–(3) are repeated. This way, the simulation grid is built up sequentially.
- In the last step, the results are checked. The data and the global proportions have to be honored, and the simulation has to look reasonable.

3.6. Indicator simulation implementation

Fig. 6 shows the dialog box for the module sisim 2.000 in GEOSSAV. The first register (Fig. 6, top left-hand side) is used for hard data input {*Hard Input Data*} and soft data input {*Soft Input Data*}, data output, and output control {*Output*}.

The column numbers for the x, y, and z coordinates and the variable to be simulated have to be specified. One or two of the coordinate column numbers can be set to zero, which indicates that the simulation is 2D or 1D. The range of data values can be reduced, and all values strictly less than the lower trimming limit and strictly greater than the upper trimming limit are ignored {*Trimming limits*}.

Soft information pertaining to continuous variables (cdf data) should steadily increase form 0 to 1, and soft information pertaining to categorical variables (pdf data) must be between 0–1 and the sum equal to 1. If the Markov–Bayes option {*Markov–Bayes simulation*} for cokriging with soft indicator data is activated, then the $B(s)$ calibration parameters have to be specified (see Section 3.5). The soft indicator data, i.e. the prior probability pdfs of type Eq. (7), are derived from calibration scattergram using the module bicalib. Fig. 7 shows the dialog box for the module bicalib 2.000 in GEOSSAV, calling for secondary and calibration data values. The secondary data file {*Secondary data*} contains secondary data that have to be integrated in the stochastic simulation. The calibration data file {*Calibration scatterplot*} contains pairs of primary and secondary data as well as declustering weights. The cutoffs or categories of the primary and the secondary variable has to be specified {*Thresholds*}. The module bicalib computes the prior distributions and the $B(s)$ calibration parameters. The prior distributions are written into an output file {*Output*} and the $B(s)$ parameters are written into a reporting file from which they must be transferred to the sisim parameter file.

The number of realizations {*Number of realizations*} to be generated as well as a random number {*Seed*}, which is used to determine the random path through all of the grid nodes of each realization, is entered in the first register (Fig. 6, top left-hand side). For output control a debugging file is written, depending on the debugging level.

The second register (Fig. 6, top right-hand side) includes the kriging options and the search parameters. The kriging options {*Options*} specify the variable type (continuous or categorical variable), the kriging type (SIK or OIK), and whether a full indicator kriging is performed at each grid node to establish the conditional distribution {*IK flag*}. If a median indicator kriging is performed, then the variogram corresponding to the selected cutoff or category {*Cut/Cat*}, respectively, is used for all cutoffs or categories. Only one kriging system needs to be solved, and, therefore, the computing time is significantly reduced.

The search parameters {*Search Parameters*} specify the maximum number of original data, the maximum number of original data per octant, the maximum number of soft data, and the maximum number of previously simulated nodes that will be used to simulate another grid node. If the data are not assigned to nodes {*Assign data to nodes*}, then the data and previously simulated grid nodes are searched separately. The data are searched with a super block search and the previously simulated nodes are searched with a spiral search. If the data are assigned to nodes, then a spiral search is used. If a multiple grid search will be performed {*Multiple grid search*}, then the grid will be refined depending on the refinement number to be specified, otherwise a standard spiral search will be considered. Details of the search strategies can be found in Deutsch and Journel (1998). In addition, the search parameters also specify the orientation and the radii of the search ellipsoid {*Search Radii/Search angles*} containing the data and the previously simulated nodes used for a node simulation.

In the third register (Fig. 6, bottom left-hand side), the variogram information used for the simulation has to be entered {*Variograms*}. The number of cutoffs or categories, the cutoff values or category codes, and the global cdf or pdf values are specified in a tabular format. For each cutoff or category the isotropic nugget constant and for each nested variogram structure: the structure model (spherical model, exponential model, Gaussian model, power model, and hole effect model), the sill contribution, the anisotropy angles, the maximum and minimum horizontal and the vertical range

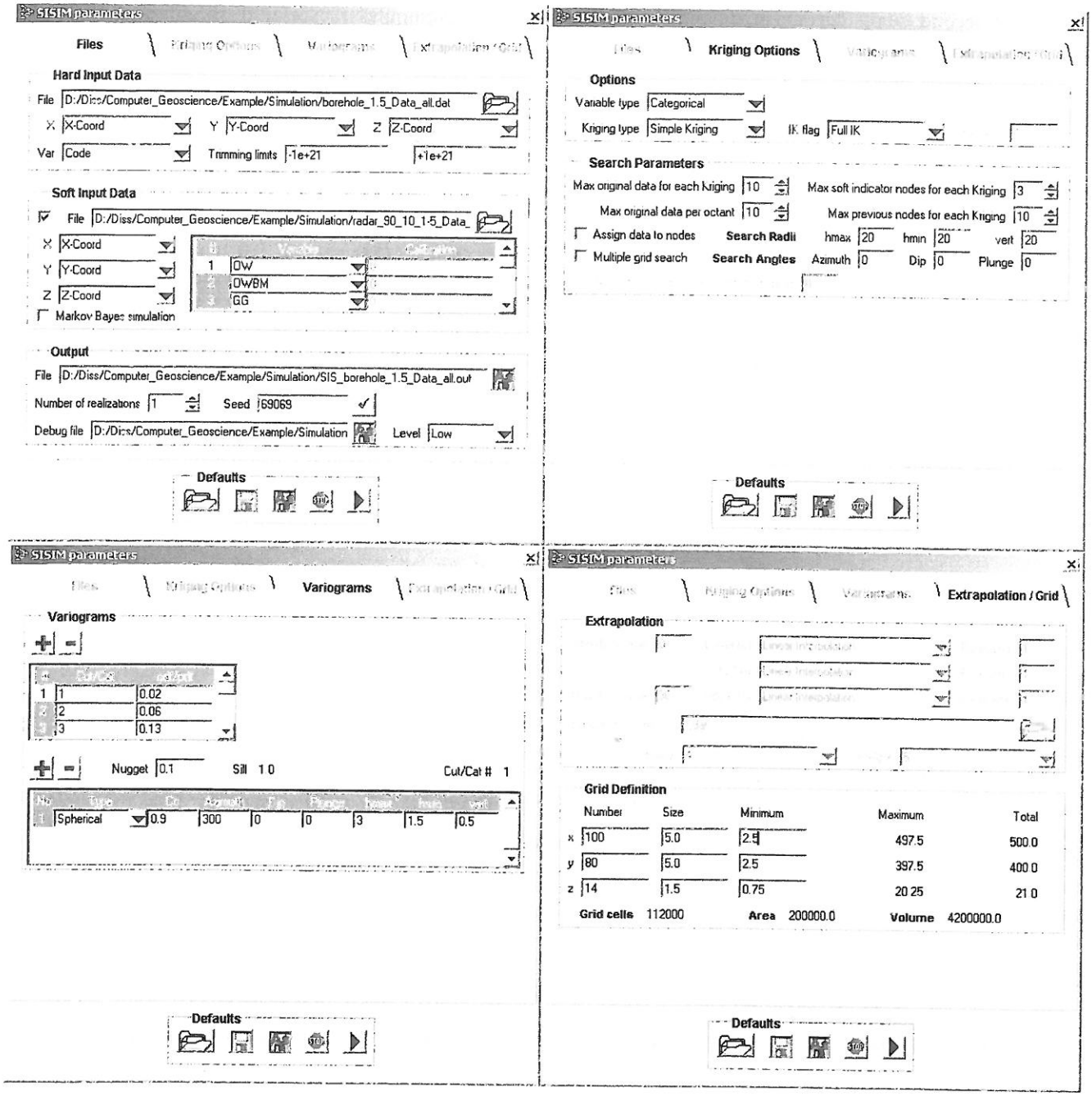

Fig. 6. Dialog box for sequential indicator simulation, used for hard and soft data input, data output, output control, and for specifying kriging options, search parameters, variogram, extrapolation, and grid information.

defining the geometric anisotropy have to be specified (Deutsch and Journel, 1998, Fig. II.4). More structures models can be added using the ± buttons above the table. There is no need to standardize the variogram to a unit sill since only the relative shape of the variogram affects the kriging weights.

The last register (Fig. 6, bottom right-hand side) includes extrapolation information and grid conventions. The extrapolation information {*Extrapolation*} is only used when considering continuous variables. The minimum and maximum data values have to be entered, the extrapolation of the lower and upper tail of the distribution as well as the interpolation within the middle of the distribution have to be specified. The possible types of interpolation/extrapolation schemes used for going beyond a discrete cdf are: (1) linear interpolation between bounds, (2) power interpolation between bounds, (3) linear interpolation between

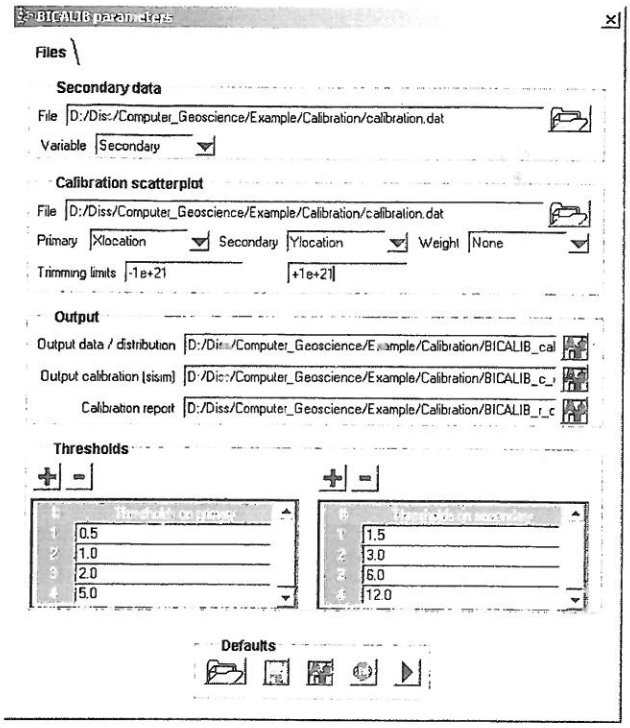

Fig. 7. Dialog box for bivariate calibration of soft data, used for specifying secondary and calibration data required to compute $B(s)$ calibration parameters that are needed in sequential indicator simulation program sisim if Markov–Bayes option is selected.

tabulated quantiles, and (4) hyperbolic extrapolation at the upper tail (Deutsch and Journel, 1998). The parameter refers to the power if the power model interpolation is selected. If linear interpolation between tabulated quantiles is selected for any of the three regions, a file with tabulated values has to be read, specifying the column numbers for the values and declustering weights. If declustering weights are not used, the class probability is split equally between the subclasses defined by the tabulated values.

The grid convention {*Grid Definition*} adapted for the simulation is the following: x-axis is associated with East, grid node indices increase from 1 to nx in the positive x-direction; y-axis is associated with North, grid node indices increase from 1 to ny in the positive y-direction; and z-axis is associated with elevation, grid node indices increase from 1 to nz in the positive z-direction. Consequently, the grid represents a relative coordinate system. The coordinate system is established, specifying the number and size of grid cells in x-, y-, and z-direction, and the coordinates at the center of the first cell. For a site-specific problem, the three axes can be associated to any absolute coordinate system that is appropriate. Therefore, a coordinate transformation must be performed.

4. Visualization methods

The OpenGL API, which is an open standard, is integrated into GEOSSAV for 3D rendering and slicing perpendicular to the main coordinate axis. OpenGL is a software interface that allows the rendering of 2D and 3D graphics images, works independently of the platform (Win32, MacOS, and virtually all variants of Unix), and uses available hardware acceleration as provided by modern graphics adapters. OpenGL is a depth buffer based rendering system for hidden surface removal (Wright and Sweet, 2000). It supports different shading models and texture mapping. The geometrical model has to be built of primitives such as points, lines, and polygons. The surface properties of these primitive objects can be adjusted in terms of color, reflectance, shininess, etc. It incorporates both orthonormal and perspective viewing models for adjusting the virtual camera. The rendered scene is lighted by one or more virtual light sources whose type, color, brightness, etc. can be adjusted. Due to hardware acceleration, even complicated 3D models can be manipulated interactively (rotation, zoom, and pan). Details of OpenGL programming and matrix mathematics can be found for example in Fosner (1997) and Wright and Sweet (2000).

In addition, selected planes perpendicular to the main coordinate axis are visualized in scaled size. This allows a detailed visual analysis of single slices and, by continuous slicing through the simulated volume, of the completely simulated property field.

5. Data export options

The export modules generate files that specify the spatial distribution of hydraulic parameters (e.g., hydraulic conductivity and porosity) as either: (1) ASCII matrix files which characterize properties of single model layers or (2) bcf package files for the complete 3D flow model setup for MODFLOW-based (Harbaugh and McDonald, 1996) groundwater simulation systems. The ASCII matrix and the bcf files can be loaded into 2D or 3D finite-difference groundwater model systems (e.g., ASMWIN, Chiang et al., 1998); GMS, [2]; PMWIN, Chiang and Kinzelbach, 2001) and can be used for flow and transport simulations.

5.1. Data export implementation

Fig. 8 shows the options for generating export files depending on the variable type used in the stochastic simulation, and Fig. 9 shows the corresponding dialog box for the export of bcf files. The simulation input file {*Simulation Input*} contains the output data from the stochastic simulation: in case of continuous variables, already the distribution of parameter values under study

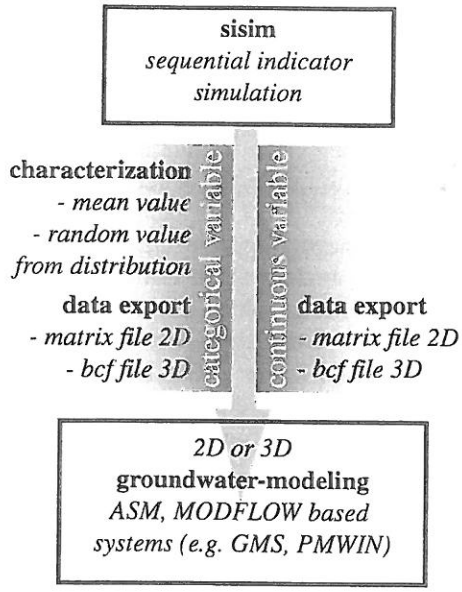

Fig. 8. Options for generating export files depending on variable type used in stochastic simulation. Categorical variables (lithofacies) have to be replaced by property under study (e.g., hydraulic conductivity), each lithofacies characterized by mean value and standard deviation. Export programs generate either (1) ASCII matrix files for the characterization of single model layers or (2) bcf package files for characterization of complete 3D flow model setup for MODFLOW-based groundwater simulation systems.

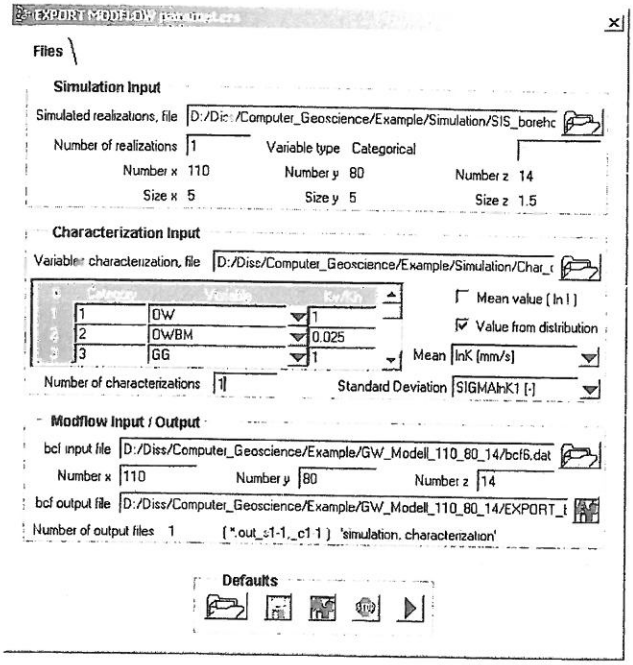

Fig. 9. Dialog box for export of bcf package files. An existing bcf file is taken to create several bcf files corresponding to simulation and characterization input parameters.

(e.g., hydraulic conductivity and porosity); in case of categorical variables, the distribution of simulated categories such as lithofacies types. If the variable type is continuous, no replacement has to be made. The parameter values of the stochastic simulation are written directly either into ASCII matrix or into bcf files depending on the selected export module. If the variable type is categorical, first the spatial distributed lithofacies types have to be replaced by hydraulic parameter values (e.g., hydraulic conductivity and porosity); subsequently the corresponding values will be written either into ASCII matrix or into bcf files {*Characterization Input*}.

The replacement of the spatial distributed lithofacies types by hydraulic parameter values is done using a random generator from Press et al. (1988). The input mean values and standard deviations of the hydraulic parameters and lithofacies types have to be specified as logarithm values. The simulated lithofacies types can be replaced either by the mean values or by random values generated from the defined distributions. The output values are not given as logarithm values. For data export an existing bcf file is taken to create several bcf files corresponding to the data of the stochastic simulation and the export parameters {*Modflow Input/ Output*}.

For subsequent external flow simulations, MODFLOW requires either hydraulic conductivity or transmissivity values depending on the layer type of each model layer (Harbaugh and McDonald, 1996). For models with more than one model layer, MODFLOW requires the input of the vertical conductance term between two model layers, known as vertical leakance (vcont). In order to export bcf files, the vcont arrays have to be calculated for all layers except the bottom layer, because MODFLOW assumes that the bottom layer is underlain by impermeable material. Vcont is calculated using the following equation (Chiang and Kinzelbach, 2001):

$$vcont = \frac{2}{(\Delta V_l/(K_z)_{j,i,l}) + (\Delta V_{l+1}/(K_z)_{j,i,l+1})}, \tag{13}$$

where ΔV_l and ΔV_{l+1} are the thicknesses of layers l and $l+1$, and $(K_z)_{j,i,l}$ and $(K_z)_{j,i,l+1}$ are the vertical hydraulic conductivities of column j, row i, and layers l and $l+1$, respectively. If the variable type is categorical, the ratio of vertical to horizontal hydraulic conductivity of each single category is considered. If the variable type is continuous, a global ratio of vertical to horizontal hydraulic conductivity is considered. The remaining data of the read bcf file are taken on as it is.

When selecting ASCII matrix files for 3D data export, properties of the subsurface are exported in single model layers. Vertical and horizontal information has to be exported separately and the vertical conductance has to be calculated running MODFLOW.

6. Example

6.1. Data

GEOSSAV was applied to generate the aquifer structures of a well capture zone (in the order of several hundreds of meters). The structures were simulated on a 550 m × 400 m × 22 m grid and subsequently exported to a finite-difference groundwater flow and advective transport model to simulate a river restoration pilot project in the region of Basel, Northwestern Switzerland. Particularly, the groundwater simulation of this portion of the Rhine/Wiese aquifer, described in Regli et al. (2003), includes simulation of changing well capture zones depending on subsurface heterogeneity, hydrologic variations, water supply operation data, and progress of river restoration.

The study site is located in the area of the ancient confluence of the main river Rhine (with flow to the Northwest) and its tributary Wiese (with flow to the Southwest). The average discharge of the river Rhine over the last 110 years amounts to 1052 m^3/s and is therefore around 90 times larger than the average discharge of the tributary Wiese with 11.4 m^3/s over the last 68 years (Bundesamt für Wasser und Geologie, 2001).

Drill-core data from five boreholes and georadar data from 14 vertical georadar sections (total length of all sections 3040 m) have been examined. The unconfined aquifer consists of Quaternary unconsolidated coarse alluvial deposits. Tertiary marls underlie these gravels and are considered impermeable for the purposes of the model. The aquifer thickness varies between 13 and 18 m. The lower 80% of the aquifer consists of Rhine gravel and the upper 20% of Wiese gravel (Zechner et al., 1995). This may be explained as due to the reworking of the Wiese gravel by the river Rhine under landscape-shaping conditions whereby the top sequence of Wiese gravel would be preserved until the next shift of the active channel area of the river Rhine. The Rhine and Wiese sediments are easily distinguished lithologically because the sediments come from different source areas with distinct geological units. Also, within these two stratigraphic units a number of sedimentary structures are recognized that were generated by sedimentary processes in the braided fluvial system. Lithofacies types associated with the sedimentary structures for this area include (Regli et al., 2002): open-framework gravel (OW), open-framework/bimodal gravel couplets (OW/BM), gray gravel (GG), brown gravel (BG), alternating gray and brown gravel layers (GG/BG), horizontally layered or inclined, silty gravel (SG), sand lenses (SA), and silt lenses (SI).

For the georadar investigations, described in Regli et al. (2002), a pulseEKKO IV georadar system with a 1000 V transmitter was used (Sensors & Software Inc., 1993). The transmitting and receiving antennae were separated by 2 m and the recording step size was 0.25 m. The 50 MHz antennae used for this study allow recognition of the basal aquiclude surface, the main erosion boundaries within the coarse alluvial deposits, and the larger sedimentary structures down to the aquiclude at approximately 13–18 m depth. According to Jol and Smith (1991) and Huggenberger (1993), the vertical resolution depends on the georadar-wave frequency and is equal to a quarter of the wavelength, or 0.5 m in this case. Due to the relatively long wavelengths of the 50 MHz antennae, few transitions of alternating sequences of open-framework and bimodal gravel may be distinguished on the georadar sections. However, the main sedimentary structures as described above could be delineated.

The drill-core and georadar data are interpreted based on Regli et al. (2002). This lithofacies-based interpretation of hard outcrop, soft drill-core and georadar data respects differences in data uncertainty and provides lithofacies type probabilities for points along boreholes and grid nodes with arbitrary mesh sizes in horizontal and vertical direction along georadar sections. The locations of the boreholes and the traces of the georadar sections used for the simulation are shown in the visualization window of Fig. 2. The sampled data from the georadar sections are at nodes of grids separated by 10 m × 1.5 m. In this case, the proportion of sampled aquifer material represented by drill-core compared with georadar data is about 1:90.

6.2. Variogram computation and indicator simulation

The variogram computation is based on the drill-core and georadar data. The data interpretation method of Regli et al. (2002) provides lithofacies type probabilities for drill-core and georadar data. The indicator transform at grid node location is set to 1 for the lithofacies types with the greatest probability values, 0 otherwise. The experimental indicator variograms of open-framework/bimodal gravel couplets, gray gravel, brown gravel, silty gravel, and sand are shown in Fig. 10. The directional variograms, given for specified azimuth and dip, characterize the geometric anisotropy of the sedimentary structure types. They indicate orientations and ranges corresponding to maximum and minimum horizontal and vertical distances of spatial correlation. Variogram information for the nine sedimentary structure types in the study area is given in Table 1. The lithofacies proportions are based on the data density representing a specific structure type. The values written in italics are estimated because the corresponding structure types never have the greatest probabilities and, therefore, the indicator transform is always set to 0 by default for these structure types. The orientation of the sedimentary structure types represents the

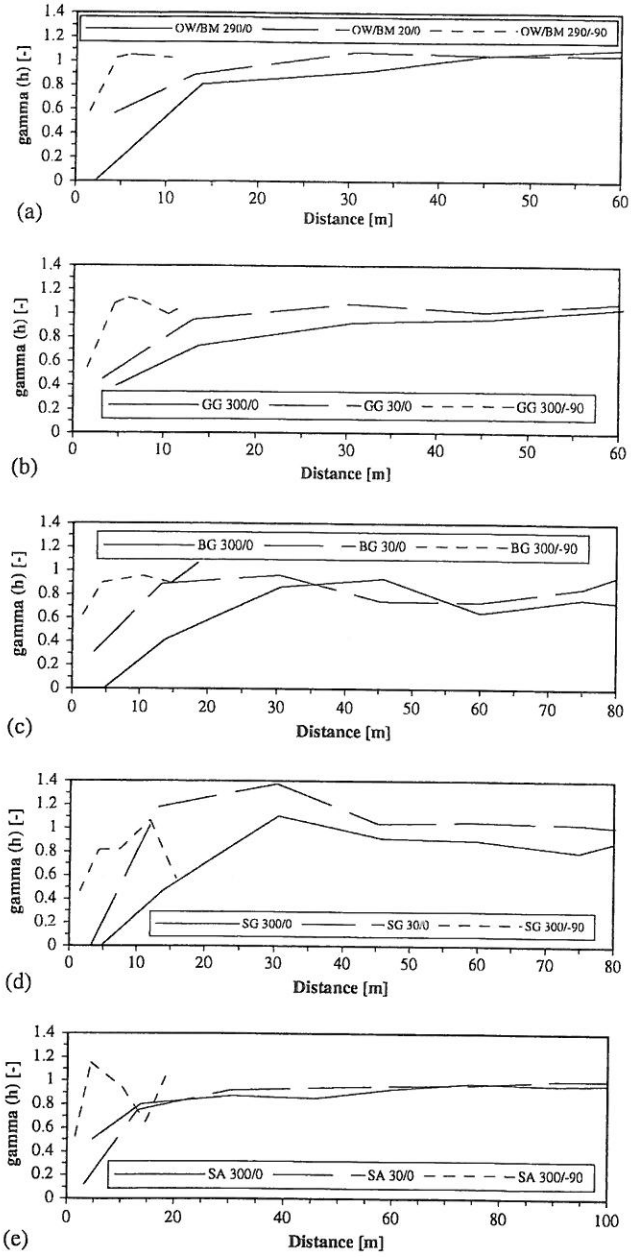

Fig. 10. Experimental indicator variograms of (a) OW/BM: open-framework/bimodal gravel couplets, (b) GG: gray gravel, (c) BG: brown gravel, (d) SG: silty gravel, and (e) SA: sand. Directional variograms are given for specified azimuth and dip (e.g., 240/0) which characterize geometric anisotropy of sedimentary structure types. Parameters of model variograms are given in Table 1.

dominance of the main flow direction of the river Rhine. The relatively large ranges of spatial correlation may be significantly influenced by the resolution of the georadar system and by the density of the sampled data taken from the georadar sections.

Fig. 11 shows a realization of the sequential indicator simulation, visualized within the OpenGL window (Fig. 11, above on the left-hand side), and a sectional detail of the XY-plane (Fig. 11, above on the right-hand side). The visualization is managed by selecting the left mouse button for 3D rendering, the right mouse button for zooming, the checkboxes for plane selection within the OpenGL window, the sliders for continuous slicing through the simulated volume, and the selection boxes for plane views perpendicular to coordinate axes.

Some of the parameter values used for this realization can be found in Fig. 6 and Table 1. The regular model grid is defined by $110 \times 80 \times 14$ cells with cell sizes of $5\,\mathrm{m} \times 5\,\mathrm{m} \times 1.5\,\mathrm{m}$. The simulated model reproduces the lithofacies proportions with an accuracy of $\pm 10\%$. Since the conditional probabilities were estimated by indicator kriging with given indicator variograms, the simulated values taken altogether will reproduce those variograms. The lithofacies types are replaced by hydraulic parameters, where the values are generated randomly based on the defined statistical description. Finally, files are generated and exported for use in groundwater flow and transport simulation models.

6.3. Discussion

When using the indicator kriging approach for generating parameter distributions, prior data should have regular spatial density. A local accumulation of data points within a particular sedimentary structure type leads to a biased amplified proportion of this type. In such cases, one should first decluster the data (e.g., with the module 'declus', Deutsch and Journel, 1998) or choose a uniformly grid pattern for sampling. The grid should be as regular as possible.

If there are many lithofacies types of very small proportions, the reproduction of such small lithofacies proportions is difficult in simulated models. Therefore, the criterions to reject a simulated model have to consider this problem.

In the example shown in Fig. 11, the current main direction of groundwater flow, the directions of hydraulic gradients and the direction of the river Wiese are oriented approximately parallel to the x-axis. The geological stuctures, however, show an anisotropy of about 35° to the x-axis. Such information is important, e.g., for modeling river–groundwater interaction and determination of well capture zones.

The geometric anisotropy of the sedimentary structures was characterized based on data of the entire aquifer. However, changes in orientation and ranges of sedimentary structures may occur due to the interactions of the two rivers over time. Those modifications could be included in the model by partitioning the aquifer vertically into two hydrostratigraphic units. This is actually done in Regli et al. (2003).

Table 1
Variogram information of Rhine and Wiese gravel used for sequential indicator simulation to define geometric anisotropy of sedimentary structure types: OW, open-framework gravel; OW/BM, open-framework/bimodal gravel couplets; GG, gray gravel; BG, brown gravel; GG/BG-horizontal, alternating gray and brown gravel, horizontally layered; GG/BG-inclined, alternating gray and brown gravel, inclined; SG, silty gravel; SA, sand; SI, silt

Sedimentary structure type variogram parameter	OW	OW/BM	GG	BG	GG/BG horizontal	GG/BG inclined	SG	SA	SI
Probability density function[a]	0.019	0.053	0.094	0.158	0.578	0.044	—	0.050	0.004
Probability density function	0.02	0.06	0.13	0.05	*0.50*	*0.05*	0.03	0.15	*0.01*
Azimuth (deg)	300	290	300	300	*300*	*300*	300	300	*300*
Max. horiz. range (m)	3	40	53	60	*80*	*8*	40	115	*3*
Min. horiz. range (m)	1.5	26	22	35	*40*	*4*	15	70	*1.5*
Vertical range (m)	0.5	4	4	10	*10*	*2*	9	4	*0.5*

Note: Values in italics are estimates since adequate data for reliably computing proportions were not available for these lithofacies.
[a] Data from Rauber et al. (1998), valid for Rhine gravel aquifers in Northeastern Switzerland; isotropic nugget constants are 0.1; the Cc values (variance contribution of nested variogram structures) are 0.9; variogram models are exponential; dips and plunges are 0.

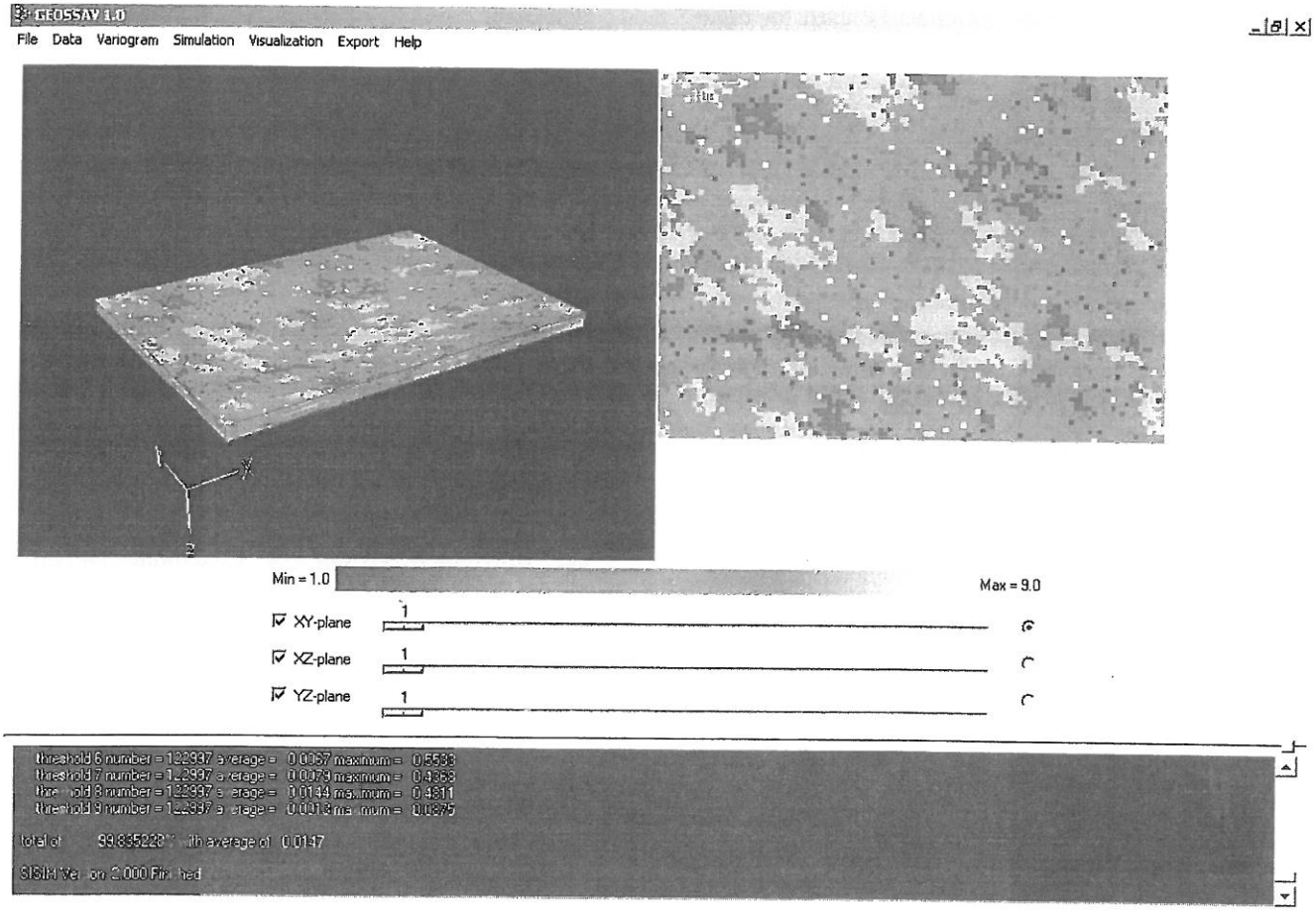

Fig. 11. Realization of sequential indicator simulation, visualized within OpenGL window, and a sectional detail of *XY*-plane. Visualization is managed by control panel, and by left and right mouse buttons.

7. Conclusions

GEOSSAV is described as a tool for the integration of hard and soft geological and geophysical data in the stochastic simulation and visualization of hydrogeological properties of the subsurface. The implemented software packages (bicalib, gamv, vargplt, and sisim from GSLIB plus OpenGL) support subsurface characterization from data processing through simulation and visualization to final control and data export to

external finite-difference groundwater simulation systems (e.g., ASM- and MODFLOW-based systems).

In particular, GEOSSAV allows one to model categorical variables such as lithofacies types and continuous variables such as distributions of hydraulic properties. The distinguished lithofacies can be explained based on depositional processes (e.g., fluvial sediment sorting processes). For many practical hydrogeological problems such as determination of capture zone of wells or risk estimation of groundwater pollutants due to contaminated sites, not only statistical distributions of lithofacies are required but also site-specific structural information. The presented methods allow integration of site-specific structural information into a framework of stochastic description of geological structures.

Although the given example includes a characterization of the simulated geological structures by hydraulic parameters, GEOSSAV can definitely be used for other applications such as groundwater and/or soil contamination, site remediation, air pollution, and ecology. In fact, whenever stochastic solutions are preferred to solve site-specific heterogeneity problems, GEOSSAV can be used as a user-friendly and adequate software solution.

GEOSSAV has been successfully tested on Microsoft Windows NT 4.0/2000/XP and SuSE Linux 7.3, and has already been applied successfully to hydrogeological problems. Due to the Tcl/Tk integration platform and development environment as well as the modular setup, GEOSSAV can be easily upgraded and adapted to specific problems, which may be solved using geostatistical methods. The current version of GEOSSAV is available at http://www.unibas.ch/earth/pract.

8. Hardware and software requirements

GEOSSAV may be operated on Windows and on most kinds of Unix based operating systems. The following hardware components are required:

- PC with at least a Pentium II processor running Windows NT 4.0 or higher, or SuSE Linux 7.3 or higher;
- 24-bit color graphics card or better;
- 64 MB or more Ram; and
- 5 MB or more hard disk space.

Tcl/Tk is the software environment that embeds the implemented applications for hydrogeological analysis and allows an integrated use of them. The following open-source software is required to run GEOSSAV:

- Tcl/Tk, release 8.4. Tcl is an extensible scripting language and library. Tcl was chosen due to its easy development environment, its cross-platform deployment, and its ability to embed codes written in other programming languages. Tk is a graphical user interface and is shipped as an extension with all distributions of Tcl. To enable the writing of graphical applications at the script level, the stand-alone interpreter wish (windowing shell) is provided.
- [incr Tcl], release 3.2. The [incr Tcl] extension is an object-oriented extension for Tcl. Objects are organized into classes with identical characteristics that can inherit functionality from one another. This helps to organize the code into modules that are easier to understand and maintain, and, therefore, larger Tcl/Tk applications can be built.
- [incr Tk], release 3.2. The [incr Tk] extension is an object-oriented framework for building mega-widgets using the [incr Tcl] object system. Mega-widgets are high-level widgets such as a file browser or a tabbed notebook.
- [incr Widgets], release 3.0.0. The [incr Widgets] extension is a set of mega-widgets that are included with the [incr Tcl] object system and the [incr Tk] mega-widget framework. This extension comes bundled with the [incr Tcl] distributions.
- TkTable, release 2.7. The TkTable extension is a full-featured 2D table widget extension for Tk. It supports images, embedded windows, and much more.
- OpenGL, release 1.1. OpenGL is an open, standardized API for 3D rendering, possibly with hardware acceleration. It was chosen because of its cross-platform functionality. The OpenGL code has been integrated into GEOSSAV with a Tcl/Tk extension written in C.

The distributed GEOSSAV executable (which is platform dependent with different versions for the Win32 and Linux platform but sharing the same code base for all platforms) automatically calls all the necessary Tcl/Tk scripts.

9. Distribution information

GEOSSAV is distributed electronically at no cost. The current version is available at http://www.unibas.ch/earth/pract. The package includes everything needed to run GEOSSAV including the current version of GEOSSAV and all the necessary open-source Tcl/Tk interactive-environment software. Although the authors cannot provide any professional support service, they welcome comments and will attempt to respond to questions regarding the software and its application. The authors neither guarantee the integrity nor the proper performance of the software.

10. Further developments

GEOSSAV is useful for many earth sciences applications and other subsurface investigations because its main target is the simulation and visualization of heterogeneous subsurface properties using hard and soft data. Consequently, we plan to upgrade GEOSSAV with: (1) additional methods for data analysis (e.g., change of coordinate systems and declustering of data); (2) interactive curve fitting for selecting variogram models based on experimental variograms generated from field data; (3) subsurface characterization (e.g., Monte Carlo analysis); and (4) the option to export additional data files for finite-difference groundwater flow and transport systems. In addition, we anticipate the implementation of a design tool in GEOSSAV that will allow fully object-oriented visualization of model data and simulated structures, which can be added in any order to the display.

Acknowledgements

Numerous members of the Applied and Environmental Geology of the Geological Institute of Basel University and of Rauber Consulting at Zürich have tested the current version of GEOSSAV and have found it easy to use and suitable for field examples. The combined effort of geologists with their geostatistical knowledge and computer scientists with their abilities to fulfill the geologists' wishes led to the present solution. We thank W. Barrash and K. Bernet for reviewing and highly contributing to the manuscript. Special thanks are addressed to two anonymous reviewers for valuable critiques. Their comments have significantly improved the manuscript. This work is part of a Ph.D. Thesis completed by Ch. Regli in 2002 at Basel University and was financially supported by the Swiss National Science Foundation, grant nos. 21-49272.96 and 20-56628.99.

References

Ayyub, B.M., Gupta, M.M., 1997. Uncertainty Analysis in Engineering and Sciences—Fuzzy Logic, Statistics, and Neural Network Approach. Kluwer Academic Publishers, Boston, 400pp.

Beres, M., Huggenberger, P., Green, A.G., Horstmeyer, H., 1999. Using two- and three-dimensional georadar methods to characterize glaciofluvial architecture. Sedimentary Geology 129, 1–24.

Bundesamt für Wasser und Geologie, 2001. Hydrologisches Jahrbuch der Schweiz 2000. Eidgenössisches Departement für Umwelt, Verkehr, Energie und Kommunikation, Bern, 435pp.

Chiang, W.-H., Kinzelbach, W., 2001. 3D-Groundwater Modeling with PMWIN—a Simulation System for Modeling Groundwater Flow and Pollution. Springer, Berlin, 346pp.

Chiang, W.-H., Kinzelbach, W., Rausch, R., 1998. Aquifer Simulation Model for Windows—Groundwater Flow and Transport Modeling, an Integrated Program. Gebrueder Borntraeger, Berlin, 137pp.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide, 2nd Edition. Oxford University Press, Oxford, 369pp.

Fosner, R., 1997. OpenGL Programming for Windows 95 and Windows NT, 2nd pr. Addison-Wesley Developers Press, Reading, MA, 259pp.

Furger, G., 1990. Von der Geologie zum Stofftransportmodell. Dissertation Nr. 9356, Eidgenössische Technische Hochschule, Zürich, 164pp.

Harbaugh, A.W., McDonald, M., 1996. User's documentation for MODFLOW-96, an update to the US Geological Survey modular finite-difference ground-water flow model. US Geological Survey Open-File Report 96-485, Reston, VA, 56pp.

Harrison, M., 1997. Tcl/Tk Tools. O'Reilly & Associates, Inc., Cambridge, 653pp.

Hess, K.M., Wolf, S.H., Celia, M.A., 1992. Large-scale natural gradient test in sand and gravel, Cape Cod, Massachusetts, 3. Hydraulic conductivity and calculated macrodispersivities. Water Resources Research 28 (8), 2011–2027.

Huggenberger, P., 1993. Radar facies: recognition of facies patterns and heterogeneities within Pleistocene Rhine Gravels, NE Switzerland. In: Best, C.L., Bristow, C.S. (Eds.), Braided Rivers. Geological Society, Special Publication 75, London, pp. 163–176.

Jol, H.M., Smith, D.G., 1991. Ground-penetrating radar of northern lacustrine deltas. Canadian Journal of Earth Sciences 28 (12), 1939–1947.

Ousterhout, J.K., 1994. Tcl and the Tk Toolkit, 9th pr. Addison-Wesley Professional Computing Series, Reading, MA, 458pp.

Pannatier, Y., 1996. Variowin—Software for Spatial Data Analysis in 2D. Springer, New York, 91pp.

Poeter, E.P., McKenna, S.A., 1995. Reducing uncertainty associated with ground-water flow and transport predictions. Ground Water 33 (6), 899–904.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vellerling, W.T., 1988. Numerical Recipes in C: the Art of Scientific Computing. Cambridge University Press, Cambridge, 735pp.

Rauber, M., Stauffer, F., Huggenberger, P., Dracos, T., 1998. A numerical three-dimensional conditioned/unconditioned stochastic facies type model applied to a remediation well system. Water Resources Research 34 (9), 2225–2233.

Regli, Ch., Huggenberger, P., Rauber, M., 2002. Interpretation of drill-core and georadar data of coarse gravel deposits. Journal of Hydrology 255, 234–252.

Regli, Ch., Rauber, M., Huggenberger, P., 2003. Analysis of aquifer heterogeneity within a well capture zone, comparison of model data with field experiments: a case study from the river Wiese, Switzerland. Aquatic Sciences 65, 111–128.

Schafmeister, M.-Th., 1999. Geostatistik für die hydrogeologische Praxis. Springer, Berlin, 172pp.

Schroeder, W.J., Avila, L.S., Martin, K.M., Hoffman, W.A., Law, C.C., 2001. The Visualization Toolkit User's Guide. Kitware, Inc., Clifton Park, New York, 380pp.

Sensors & Software Inc., 1993. PulseEKKO Software User's Guide. Sensors & Software Inc., Mississauga, Ont., 120pp.

Sudicky, E.A., 1986. A natural-gradient experiment on solute transport in a sand aquifer: spatial variability of hydraulic conductivity and its role in the dispersion process. Water Resources Research 22 (13), 2069–2082.

Weissmann, G.S., Carle, S.F., Fogg, G.E., 1999. Three-dimensional hydrofacies modeling based on soil surveys and transition probability geostatistics. Water Resources Research 35 (6), 1761–1770.

Wingle, W.L., Poeter, E.P., McKenna, S.A., 1997. UNCERT: a geostatistical uncertainty analysis package applied to groundwater flow and contaminant transport modeling. Colorado School of Mines, Golden, CO, 464pp.

Wright, R.S., Sweet, M., 2000. OpenGL Superbible, 2nd Edition. Waite Group Press, Indianapolis, IN, 696pp.

Zechner, E., Hauber, L., Noack, Th., Trösch, J., Wülser, R., 1995. Validation of a groundwater model by simulating the transport of natural tracers and organic pollutants. In: Leibundgut, Ch. (Ed.), Tracer Technologies for Hydrological Systems, Vol. 229. International Association of Hydrological Sciences Publication, Wallingford, Oxfordshire, United Kingdom, pp. 57–64.

Further reading

[1] Isaaks & Co., 2001. SAGE2001. A Spatial and Geostatistical Environment for Variography. http://www.isaaks.com.

[2] Environmental Modeling Systems Inc., 2002. GMS: Ground water Modeling System. http://www.ems-i.com.