3B2v8.06a/w (Dec 5 2003).51c + model | CAGEO : 1702 | Prod.Type:FTP pp.1–13(col.fig.:NIL) | ED:BibhuPrasad PAGN:Raj SCAN:

ARTICLE IN PRESS

ELSEVIER

Computers & Geosciences ▮ (▮▮▮▮) ▮▮▮–▮▮▮

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

A sequential indicator simulation program for categorical variables with point and block data: BlockSIS ☆

Clayton V. Deutsch*

Centre for Computational Geostatistics (CCG), Department of Civil & Environmental Engineering, University of Alberta, 3-133 NREF Bldg. Edmonton, AB Canada T6G 2W2

Received 28 January 2006; received in revised form 13 March 2006; accepted 13 March 2006

Abstract

Stochastic simulation of facies or geologic units is important before the assignment of continuous rock properties. Sequential indicator simulation (SIS) remains a reasonable approach when there are no clear genetic shapes that could be put into object-based modeling. Constraining SIS to soft secondary data coming from geological interpretation or geophysical measurements is important. There are a number of techniques including indicator kriging (IK) with a local mean, collocated cokriging, Bayesian updating, permanence of ratios, block kriging and block cokriging. `BlockSIS` implements all of these and more (nine all together). The images may also be cleaned using maximum a -posteriori selection.

Keywords: Geostatistics; Cokriging; Kriging; Soft data

1. Introduction

Geostatistical realizations are being used increasingly for uncertainty quantification. There are many important decisions in geostatistical modeling. Choosing the volume to estimate within is arguably the most important. Decisions of stationarity are also important. It is common to create models of a categorical variable that represents facies or rock type. Stationarity is assumed within the different categorical variables. Such categorical variable models are sometimes built deterministically relying on expert judgment; however, in many cases there is inadequate data to permit reliable modeling. A stochastic modeling algorithm is used to construct multiple realizations. Sequential indicator simulation (SIS) is a widely used technique for categorical variable models.

There are legitimate criticisms against SIS. The models can appear very patchy and unstructured; indicator variograms only control two-point statistical measures. Object-based or process-based models provide more structural control. SIS also often leads to uncontrolled and geologically unrealistic transitions between the simulated categories; the cross correlation between multiple categories is not explicitly controlled. Truncated

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

*Tel.: +1780 492 9916; fax: +1780 492 0249.
E-mail address: cdeutsch@ualberta.ca.

0098-3004/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2006.03.005

pluriGaussian models provide a more straightforward approach to handle multiple category interactions.

Despite these criticisms, there are many good reasons to consider SIS. The required statistical parameters are easy to infer from limited data. The models are reasonable in settings where there are no large-scale curvilinear features. The algorithm is robust and provides a straightforward way to transfer uncertainty in categories through to the resulting numerical models.

A number of SIS variations have evolved over the years. Most of the variations relate to the use of soft secondary data arising from geological interpretation or geophysical measurements. This paper collects the most common variations together in clean GSLIB-like code.

2. Methodology

Consider K different categories. They are mutually exclusive—only one category can exist at a particular location. They are exhaustive—one of the categories must exist at all locations. The categorical variable is expressed as a series of K indicator variables:

$$i(\mathbf{u};k) = \begin{cases} 1 & \text{if category } k \text{ prevails at location } \mathbf{u} \\ 0 & \text{otherwise} \end{cases}, \quad k = 1,\ldots,K. \tag{1}$$

An indicator variable is often interpreted as the probability for a category to prevail at a particular location: the probability is 1 if it does prevail and 0 if it does not. Hard local measurements are coded into 1s and 0s. Imprecise or soft measurements may be coded into continuous probabilities between 0 and 1. Classical (geo) statistical inference is undertaken using indicator data including declustering for representative proportions and variography to understand the spatial continuity of each of the K indicator variables.

The local hard and soft indicator data are used to estimate the distribution of uncertainty at an unsampled location. Commonly, kriging is used for this estimation hence the name indicator kriging (IK). The K estimates $i^*(\mathbf{u};k)$, $k = 1,\ldots,K$ vary continuously between 0 and 1; they depart from the global proportions in presence of relevant local data. Regardless of how the indicators are interpolated, they do not always satisfy the order relation requirement for a closed set of probabilities, that is, they should be non-negative and sum to 1.0. Standard practice is to reset negative estimates to zero and then standardize the K estimates to sum to one by dividing through by the sum. Order relation deviations always exist. They are more severe when the K estimates are inconsistent with each other.

Sequential indicator simulation applies IK in a sequential fashion where a precise category is drawn by Monte Carlo simulation at each location. All locations are visited sequentially with an increasing level of conditioning. A random order is followed to avoid artifacts. The simulated realization reproduces the indicator variograms according to the same simple kriging principle used in continuous variables. Order relation deviations lead to a lack of reproduction.

The main purpose of modeling a categorical variable such as facies and rock type before continuous petrophysical properties is to permit a more reasonable decision of stationarity, that is, to provide subdivisions that are more geologically and statistically homogeneous. Interestingly, however, it is common for the categories themselves to have trends and regions of higher and lower proportion. Information on trends is available from geological mapping or geophysical measurement. Geological trends are mapped as proportions or probabilities.

Geophysical measurements could be treated as a secondary variable for cokriging, which does not require the seismic data to be explicitly calibrated to facies proportions; the original seismic units could be retained and the calibration enters through the cross covariances between the hard indicator data and soft seismic data. Notwithstanding this flexibility of cokriging, calibrated probability values are preferred because: (1) the calibrated probability values are in units we understand, and (2) cross variograms or covariances are often difficult to infer in presence of sparse well data. This calibration is discussed in many sources including Deutsch, (2002).

Based on direct mapping or calibration, the soft secondary data takes the form $p(\mathbf{u};k)$, $k = 1,\ldots,K$ where p is the proportion or probability of category k at location u. The locally varying proportions may be at a larger scale; often, they represent vertical averages over a particular stratigraphic interval. A number of different (co) kriging options have evolved to handle soft secondary data. Following are the kriging options available in the `BlockSIS` program:

Option	Description
0	Stationary simple kriging
1	Ordinary kriging
2	Nonstationary simple kriging using residuals from local mean
3	Nonstationary simple kriging assigning one minus the sum of the weights to the local mean
4	Collocated cokriging
5	Bayesian updating
6	Updating of probabilities with permanence of ratios
7	Block kriging with probability representing the Z thickness
8	Block cokriging with probability representing the Z thickness

Each of these kriging options will be described below. Goovaerts's book is a good place to find a more thorough description of the various kriging equations. The focus here is on the differences between the options and the implementation in `BlockSIS`. Option 0 is stationary simple kriging, which is required by theory for CDF and variogram reproduction. In presence of n nearby data the estimator is written:

$$i^*_{\mathrm{SK}}(\mathbf{u};k) - p_k = \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \bullet \left[i(\mathbf{u}_\alpha;k) - p_k \right],$$

$$i^*_{\mathrm{SK}}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k) + \left[1 - \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \right] \bullet p_k. \quad (2)$$

The second equation is a reorganization of the first. The simple kriging weights are solved by the well known kriging equations; the notation for the kriging weights is somewhat confusing, but it is required to be clear that the weights are for each data (α subscript), by simple kriging (the SK superscript), and they relate to the location being estimated and the particular category (the $(\mathbf{u};k)$) parenthetical parameters. The p_k values are the global declustered mean probability for each category. The K estimates are performed independently and order relation deviations are corrected. Note that the kriging variance is not used; we only need the weights. We also note that if the kriging weights were to sum to one, the global mean would not be used in the equation.

Option 1 is ordinary kriging. The kriging weights are constrained to sum to one. The estimator simplifies to

$$i^*_{\mathrm{OK}}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{OK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k). \quad (3)$$

Ordinary kriging is not recommended in sequential Gaussian simulation because of variance inflation; however, ordinary indicator kriging does not have the same problem because the kriging variance is not used. Ordinary kriging in sequential simulation does not always work as expected. Previously simulated grid nodes are used in the kriging, which spreads the influence of the original data to a much larger region than expected.

The remaining seven options in `BlockSIS` are various ways to use secondary data. Option 2 uses nonstationary simple kriging with residuals from the locally varying mean probabilities:

$$i^*_{\mathrm{LVM}_1}(\mathbf{u};k) - p_k(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \bullet \left[i(\mathbf{u}_\alpha;k) - p_k(\mathbf{u}_\alpha) \right]. \quad (4)$$

The simple kriging weights are the same. We cannot reorganize this equation as we did in Eq. (2) because the locally varying mean is different at the location being estimated and the data locations.

Option 3 is a slight modification to the generalized nonstationary simple kriging. One minus the sum of the weights is assigned to the local mean:

$$i^*_{\mathrm{LVM}_2}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k) + \left[1 - \sum_{\alpha=1}^{n} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u};k) \right] \bullet p_k(\mathbf{u}). \quad (5)$$

The differences between Options 2 and 3 are minor when the locally varying mean values are smooth, but the differences can be more significant when there is greater variability in the local probability values, for example, when they come from seismic data.

Option 4 is a collocated cokriging. A simplified cokriging system is solved to get the $n+1$ weights that apply to the n hard indicator data and the collocated probability:

$$i^*_{\mathrm{CCK}}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{CCK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k) + \lambda^{\mathrm{CCK}}_{n+1} \bullet p_k(\mathbf{u}) + \left[1 - \sum_{\alpha=1}^{n} \lambda^{\mathrm{CCK}}_{\alpha}(\mathbf{u};k) - \lambda^{\mathrm{CCK}}_{n+1}\right] \bullet p_k. \quad (6)$$

$$i^*_{\mathrm{PR}}(\mathbf{u};k) = \frac{(1 - p_k/p_k)}{(1 - p_k/p_k) - (1 - i^*_{\mathrm{SK}}(\mathbf{u};k)/i^*_{\mathrm{SK}}(\mathbf{u};k)) - (1 - p_k(\mathbf{u})/p_k(\mathbf{u}))} \quad (8)$$

A correlation coefficient is required to build and solve the simplified cokriging system of equations. A different correlation coefficient could be used for each category; however, care should be taken to ensure consistency. The correlation coefficient could not be high for one category and low for another category. They can be increasingly different as there are more categories. Unbiasedness requires that one minus the sum of all weights be applied to the global mean. Collocated cokriging assumes a Markov-type model of coregionalization between the soft probability and the hard indicators.

Option 5 is Bayesian updating. There are interesting theoretical links between Bayesian Updating and collocated cokriging. The implementation, however, is quite different and it is worthwhile to have this option implemented differently from the previous option. The simple kriging estimate is calculated Eq. (2) and postprocessed as follows:

$$i^*_{\mathrm{BU}}(\mathbf{u};k) = i^*_{\mathrm{SK}}(\mathbf{u};k) \bullet \frac{p_k(\mathbf{u})}{p_k}, \quad (7)$$

where $p_k(\mathbf{u})$ is the locally varying probability and p_k is the global probability for the kth category. The updated probability for one or more of the categories could exceed 1.0, which is non-physical. The resulting probabilities would have to be corrected. A standard order relations correction is to (1) reset all negative probabilities to zero, then (2) divide the probabilities by the sum of all probabilities. There is no need for a correlation coefficient; the information content in the secondary data is contained in the difference of $p_k(\mathbf{u})$ from p_k. There is an implicit assumption of conditional independence in this Bayesian updating formalism.

An alternative updating procedure dubbed *permanence of ratios* was proposed by André Journel (Journel, 2002). The permanence of ratios method is equivalent to the naïve Bayes model in machine learning. Many people consider this method to provide an improved scheme for merging two estimates of conditional probability (the SK estimate and the secondary data estimate in our context). Option 6 uses this procedure. The equation for the estimate is as follows:

where $p_k(\mathbf{u})$ is the locally varying probability and p_k is the global probability for the kth category. There is a different implicit assumption of dependence in this formalism.

Option 7 implements block kriging for the local probability. An assumption is made that the secondary data represents the value over the entire vertical extent of the model, which is appropriate in many stratigraphic settings. The estimator:

$$i^*_{\mathrm{BK}}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{BK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k) + \lambda^{\mathrm{BK}}_{n+1} \bullet p_k(\mathbf{u}) + \left[1 - \sum_{\alpha=1}^{n} \lambda^{\mathrm{BK}}_{\alpha}(\mathbf{u};k) - \lambda^{\mathrm{BK}}_{n+1}\right] \bullet p_k. \quad (9)$$

The average covariances between every data and the assumed block data $p_k(\mathbf{u})$ is calculated by numerical integration. There is an assumption that the block data is a true average of the smaller scale values. A block average of probabilities is not the probability of the block; it is a composite or average of point scale probability values.

Option 8 implements block *co*kriging for the local probability. This is similar to Option 7, but block cokriging is performed instead of block kriging. A Markov-type model is assumed whereby the cross variogram at a point-scale is derived from the indicator variogram and the correlation coefficient. A similar numerical integration approach is used to calculate the require block cross covariances. The estimator:

$$i^*_{\mathrm{BCK}}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda^{\mathrm{BCK}}_{\alpha}(\mathbf{u};k) \bullet i(\mathbf{u}_\alpha;k) + \lambda^{\mathrm{BCK}}_{n+1} \bullet p_k(\mathbf{u}) + \left[1 - \sum_{\alpha=1}^{n} \lambda^{\mathrm{BCK}}_{\alpha}(\mathbf{u};k) - \lambda^{\mathrm{BCK}}_{n+1}\right] \bullet p_k. \quad (10)$$

The correlation coefficient can depend on the category, but in general they should be similar.

2.1. Choosing the right option

The nine different options available in the BlockSIS program can be confusing. There are a number of general remarks about when the different options should be chosen:

- Simple kriging (Option 0) is the best approach when there is no secondary data and there is no evidence of significant nonstationarity. Simple kriging is required by theory for statistical parameter reproduction.
- Ordinary kriging (Option 1) is a good approach when there are many local data and there is some evidence of non stationary areas.
- The locally varying mean approach (Options 2 and 3) is suitable for local mean values coming from geological interpretation. They could be used for geophysical-derived values. The first local mean option (2) is the most correct by theory; however, the simplified option 3 places slightly more emphasis on the local hard data instead of the local mean values.
- Collocated cokriging (Option 4) is suitable when the scale of the secondary data is similar to that being modeled and there is a clear statistical correlation between the hard data and the secondary data. Collocated cokriging gives explicit control over the weight placed on the secondary data.
- The updating approaches (Options 5 and 6) are suitable when there are few hard data and it is not straightforward to establish a correlation coefficient. The correlation is somehow embedded in the local probability values, but it is not explicitly specified. The Bayesian Updating approach will give more weight to the secondary data. The permanence of ratio approach will, in general, give less weight to the secondary data because its redundancy with the hard data is captured better.
- The block (co)kriging approaches (Options 7 and 8) are appropriate for cases where the secondary data represents a vertical average over the vertical extent of the model. Block kriging is correct when the secondary probability values are reliable. The cokriging approach can be used to downweight the secondary data.

In the end, there is no recipe for correct application. The results should be carefully checked to ensure that there are no biases in the local or global proportions and that the patterns of spatial variation appear reasonable.

2.2. Image cleaning

One concern with sequential indicator simulation is that the realizations often show short-scale variations, which appear geologically unrealistic. In some cases, such variations affect subsequent processing and predicted reserves; a justifiable reason to consider realization cleaning algorithms. A second concern is that the category proportions often depart from their target input proportions. In particular, facies types with relatively small proportions (5–10%) may be poorly matched. In indicator simulation, the main source of this discrepancy is the order relations correction (the estimated probabilities are corrected to be non-negative and sum to 1.0). There is no evident alternative to the commonly used order relations correction algorithms; post-processing the realizations to honor target proportions is a convenient and attractive solution.

For convenience, the maximum a posteriori selection or MAPS technique has been implemented within the BlockSIS program. The basic MAPS algorithm amounts to replacing the categorical code type at each location by the most probable category based on a local neighborhood. The probability of each category, in the local neighborhood, is based on (1) closeness to the location being considered; (2) whether or not the value is a conditioning datum; and (3) mismatch from the target proportion.

Four variations are implemented in BlockSIS: no cleaning, light cleaning, heavy cleaning, or superduty cleaning. The basic structure and weighting of the cleaning is based on the covariance table that is constructed for the kriging. Larger cleaning windows are considered when the level of cleaning is increased. The user should choose the level of cleaning carefully; unnecessary cleaning can impose too much continuity and unreliable distributions of uncertainty.

3. Program

The BlockSIS program follows standard GSLIB conventions. Most of the functions are available in GSLIB. Two source code files are required: BlockSIS.for and BlockSISsubs.for; the subroutines have been collected to facilitate compilation if the compiled GSLIB library is not available. There are no hard-coded limits in the

BlockSIS program; dynamic memory allocation is used throughout. The parameters for the program:

```
Line  START OF PARAMETERS:
1     8                       0 = SK,1 = OK,2 = L1,3 = L2,4 = CC,5 = BU,6 = PR,7 = BK,8 = BC
2     1                       Clean: 0 = none, 1 = light, 2 = heavy, 3 = super
3     2                       number of categories
4     0  1                      categories
5     0.70  0.30                global proportions
6     0.50  0.50                correlation coefficients for soft data
7     well.dat                file with local data
8     1  2  3  4                columns for X,Y,Z, and category
9     lvm.dat                 file with gridded prior mean values
10    1  2                      columns for each category
11    3                       2-D areal map (2) or 3-D cube (3)
12    keyout.dat              file with keyout array
13    1                       column for keyout indicator
14    1                       debugging level: 0,1,2,3,4
15    BlockSIS.dbg            file for debugging output
16    BlockSIS.out            file for simulation output
17    1                       number of realizations
18    100  0.00  150.0        nx,xmn,xsiz
19    150  0.00  150.0        ny,ymn,ysiz
20    50  0.00  1.0           nz,zmn,zsiz
21    69069                   random number seed
22    12                      maximum original data for each kriging
23    12                      maximum previous nodes for each kriging
24    1                       assign data to nodes? (0 = no,1 = yes)
25    0                       maximum per octant (0 = not used)
26    5000.  5000.  10.       maximum search radii
27    30.  0.  0.             angles for search ellipsoid
28    101  101  101           size of covariance lookup table
29    1  0.0                  Cat 1: nst, nugget effect
30    1  1.0  30.  0.0  0.0     it,cc,ang1,ang2,ang3
31    5000.  5000.  10.         a_hmax, a_hmin, a_vert
32    1  0.0                  Cat 2: nst, nugget effect
33    1  1.0  30.  0.0  0.0     it,cc,ang1,ang2,ang3
34    5000.  5000.  10          a_hmax, a_hmin, a_vert
```

The kriging option is specified on Line 1. The cleaning option is specified on Line 2. The number of categories and the category codes are specified on lines 3 and 4. The global proportions (expressed as a fraction) are specified on Line 5. They are required regardless of the kriging option chosen. The correlation coefficients between the hard indicators and soft probabilities for each category are specified on Line 6. They are required regardless of the kriging option chosen.

The local data are specified on Lines 7 and 8. Standard GSLIB conventions are used for the data file and the column specifications. Unconditional realizations are created if the file does not exist. The locally varying probabilities are specified on Lines 9, 10 and 11. A probability is required for each category even though they must all sum to 1 and there are only K-1 degrees of freedom. Line 11 specifies whether the locally varying probabilities represent an areal 2-D map or a full 3-D grid. The 2-D values will be replicated to the 3-D grid if required. A mask or keyout array the size of the grid may be considered; Lines 12 and 13 specify the

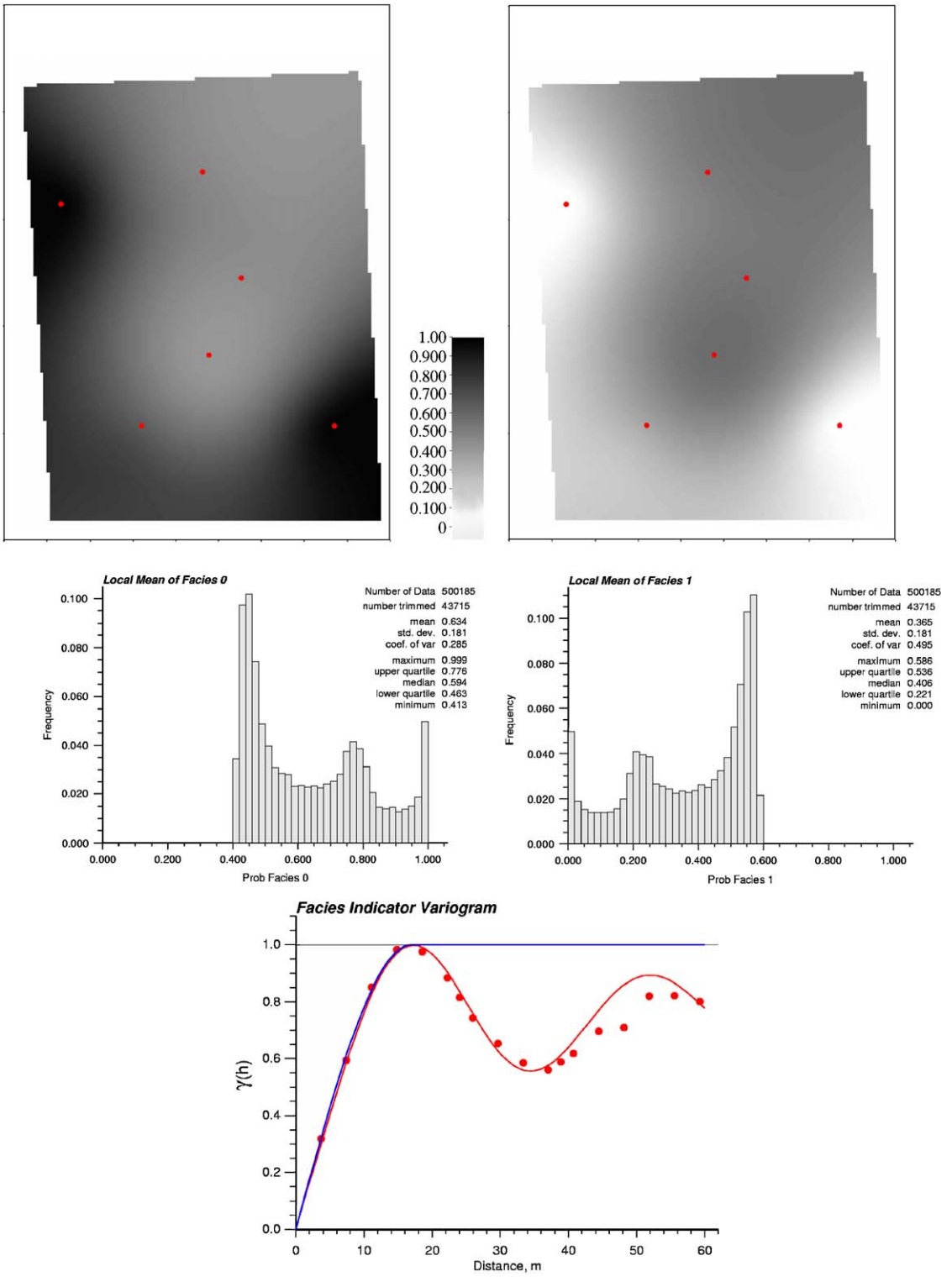

Fig. 1. Maps and histograms of local probability values for facies types 0 (left) and 1 (right). Red dots are well locations. Histograms of prior mean values and indicator variogram are also shown.

keyout file. Locations where the keyout value is 0 will not be simulated; locations where the value is 1 will be simulated. The entire grid will be simulated if there is no keyout file.

The debugging level and a file for the debugging output are specified in Lines 14 and 15. The output file is specified on Line 16. Standard GSLIB conventions are followed. The number of realizations is specified in Line 17. The grid in standard GSLIB conventions is specified on Lines 18, 19 and 20. The random number seed is in Line 21.

Lines 22 through 28 specify search parameters. Line 24 specifies whether or not the data are assigned to the grid node locations. It is more CPU efficient to assign the data, but that may not be acceptable given the spacing of the data and the grid size. The number of original data in Line 22 is only used when the data are not assigned to grid

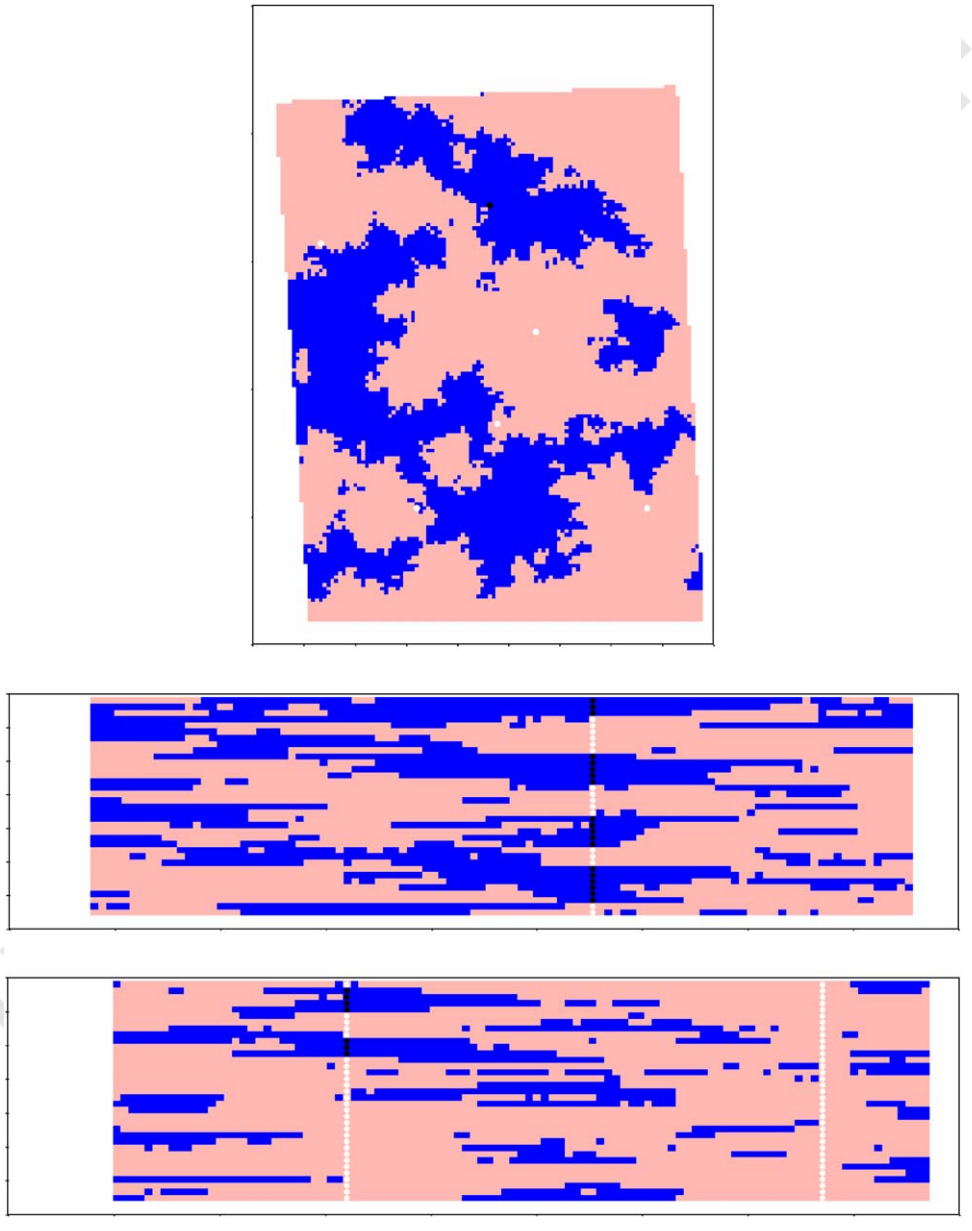

Fig. 2. Slices through SIS (SK) realization with no cleaning. Horizontal slice (at top) is through center of model. Two *XZ* cross sections are at index 76 (near center) and 30 (toward South).

nodes. The number of previously simulated grid nodes, specified on Line 23, is used for original data assigned to grid nodes and to previously simulated grid nodes. A maximum number per octant is specified on Line 25; it is used if it is greater than 0. The search radii and orientation is specified on Lines 26 and 27. The size of the covariance lookup table is on Line 28. The lookup table should be set large enough to avoid calculating the covariances every time; this is particularly true if block kriging. The vertical size of the covariance lookup table should be set to *2nz+1*.

Lines 29 and greater specify 3-D variogram models in standard GSLIB conventions for each category. A variogram model is required for each category – even if there are only two.

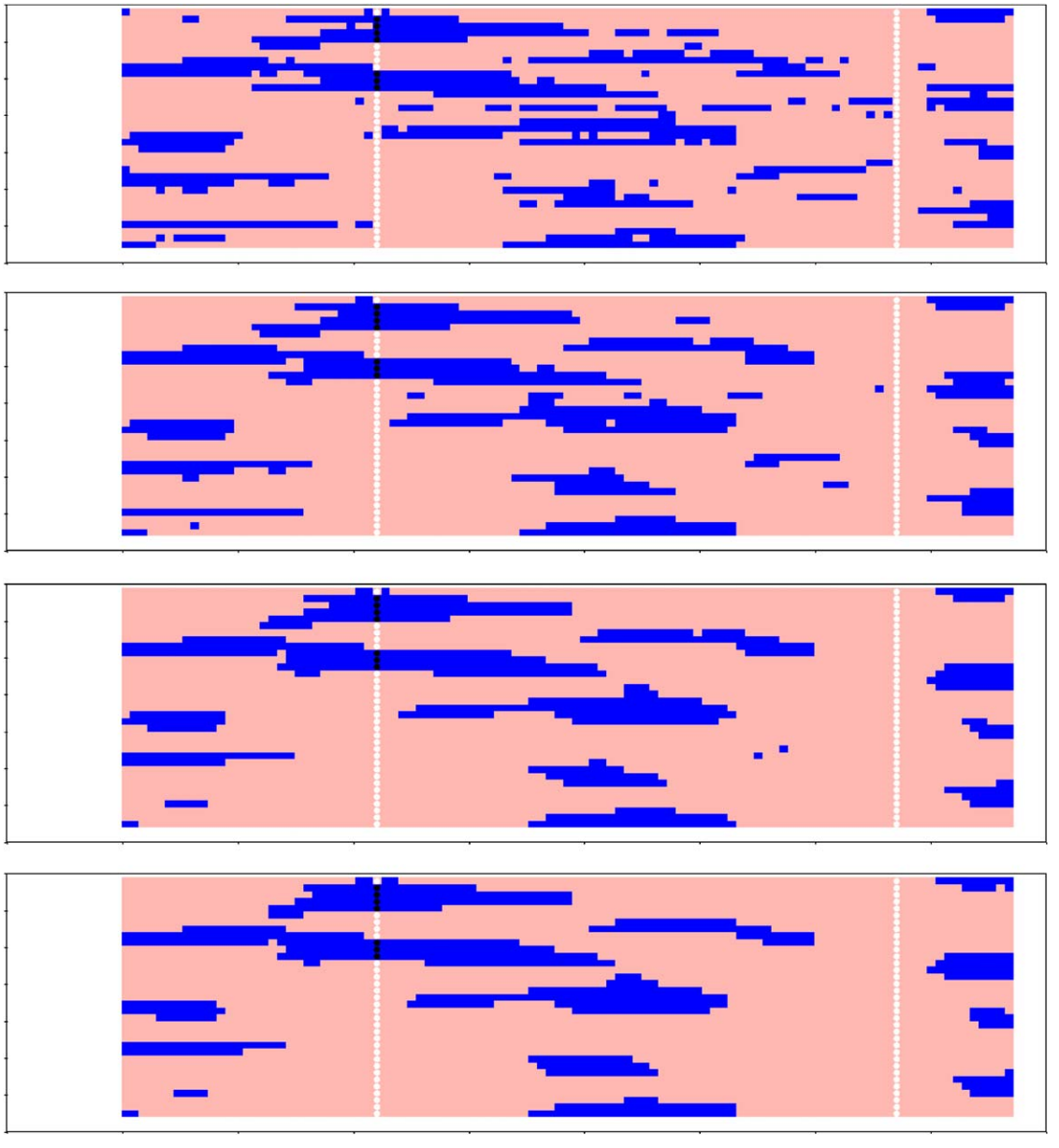

Fig. 3. Slice through SIS (SK) realization with different cleaning options: no cleaning (top), light, heavy, and super cleaning (from top to bottom). All realizations reproduce well data.

4. An example

We will use a small 2-category example for illustration. The first category (facies) "0" is non-net reservoir and the second facies "1" is net sand. Fig. 1 shows the location of six wells and locally varying mean values coming from mapping. We see that the proportion of facies 0 is 0.635 and the proportion of facies 1 is 0.365. The proportions from the well data are 0.705 and 0.295. We should, of course, use the proportions from the locally varying mean which are declustered to represent the area of interest. There is no vertical resolution in the locally varying mean values; they represent the

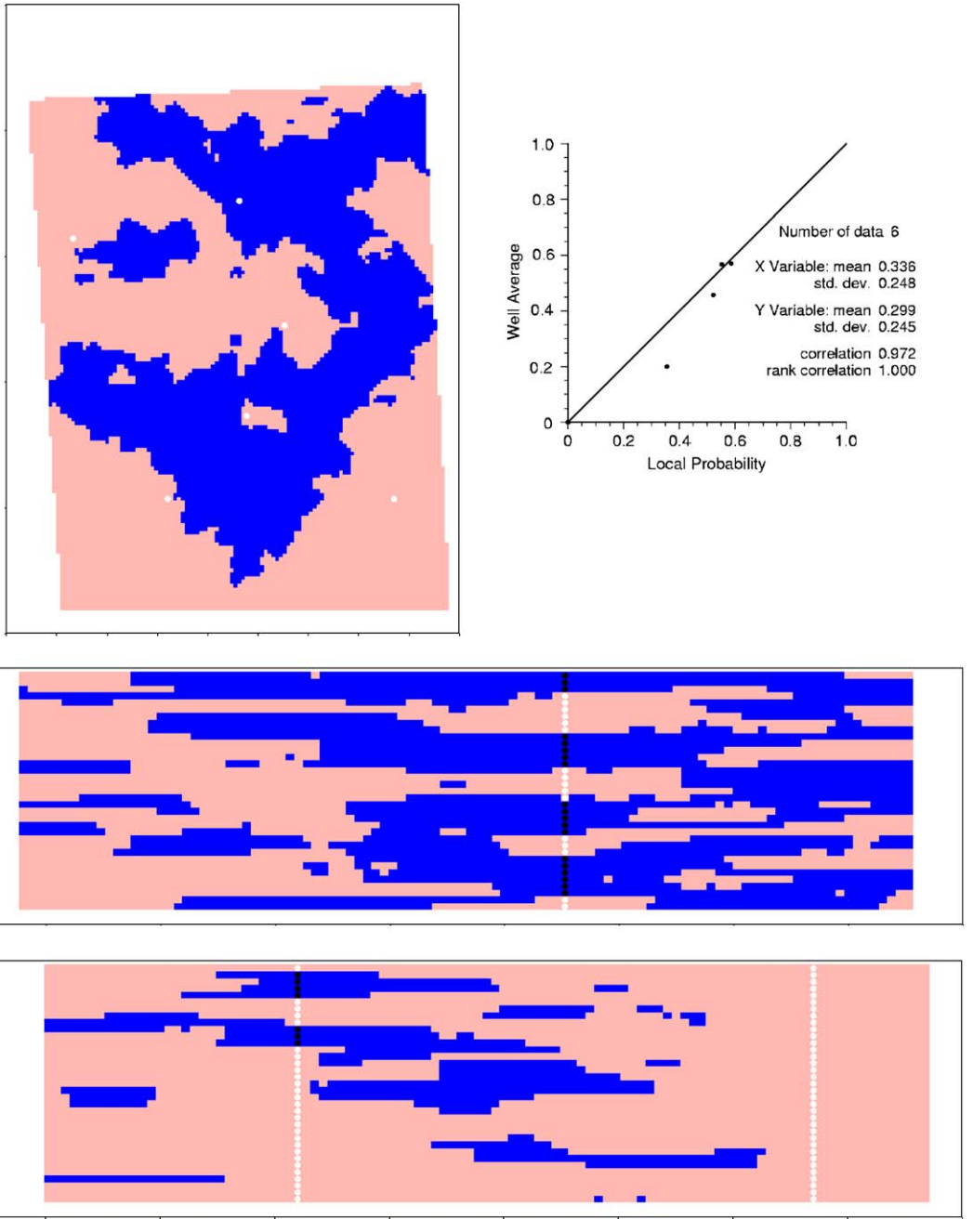

Fig. 4. Slices through SIS (Block Cokriging) realization with light cleaning. Horizontal slice (at top) is through center of model. Two *XZ* cross sections are at index 76 (near center) and 30 (toward South).

entire vertical thickness of the stratigraphic layer. 8% of the grid is keyed out—notice the edges.

The vertical indicator variogram from the well data is shown at the bottom of Fig. 1. The blue curve is a single spherical structure with a range of 17 m, which is the simplest reasonable variogram model to consider. A two-structure variogram model with a dampened hole effect is shown as the red curve. The range on these variograms is likely too large; it is being affected by the non-net wells. The range could be decreased. That detailed analysis is outside the scope of this note on the `BlockSIS` paper.

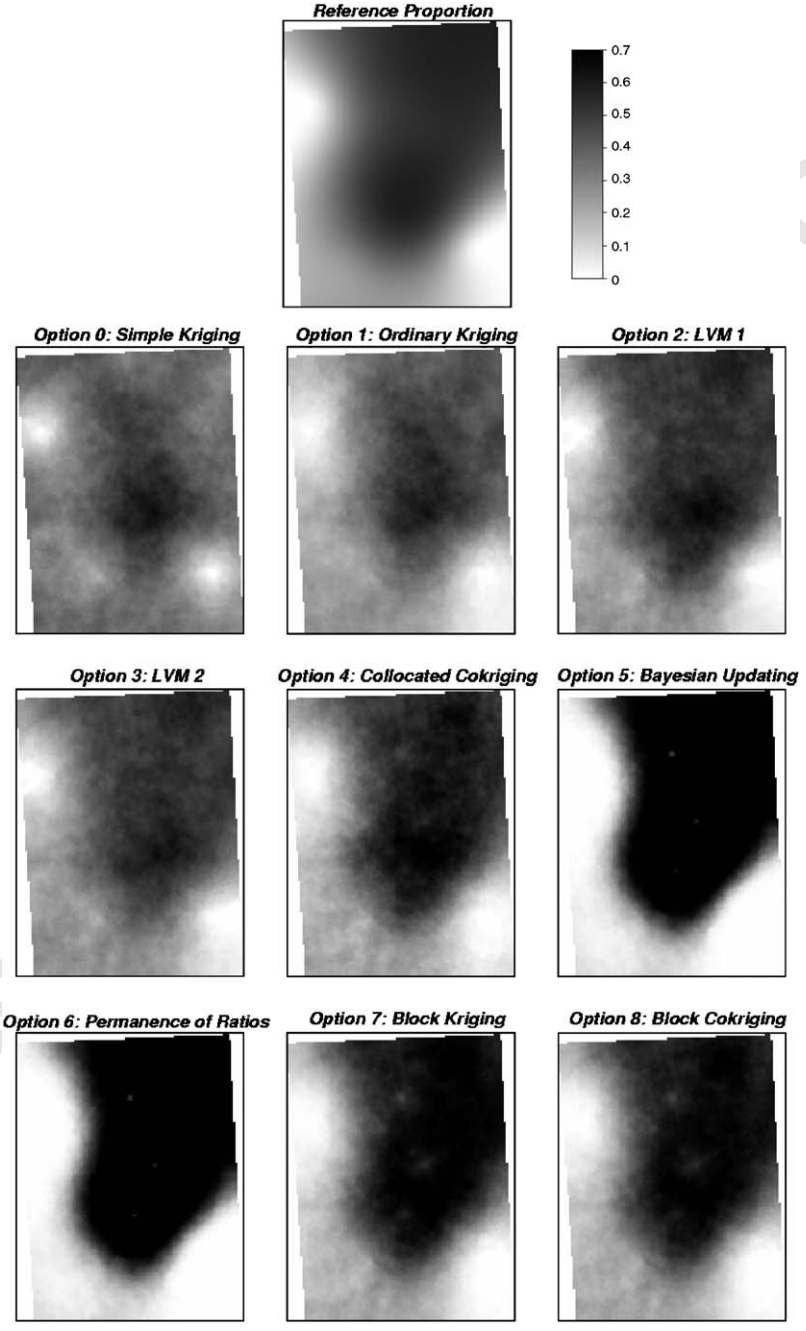

Fig. 5. Areal proportions of facies 1. Reference is at top center. All options averaged over 10 realizations are shown below.

The grid is 111 by 140 by 35 grid nodes (543900 total). The indicator variograms will be fixed with a very small nugget effect, a single spherical structure, a vertical range of about 3 grid blocks and a horizontal range of about 30 grid blocks.

Fig. 2 shows three slices through an SIS realization with Option 0 – SK and no cleaning. The global proportions and variograms are reasonably reproduced. The local well data are exactly reproduced. The local probabilities are not used in the modeling and are not reproduced. Fig. 3 shows one of the slices through SK-based realizations with different levels of cleaning. The facies are indeed smoother and less erratic as the cleaning level increases.

Realizations were generated with all options; however, to save space, the results are not shown. The OK option preserves trends away from the well data better than the SK option. The two LVM realizations look very similar. The collocated cokriging option led to results very similar to the LVM. The Bayesian updating and the permanence of ratios options let to results that were almost exactly the same. These two models gave too much influence to the secondary variable. Fig. 4 show the results for block cokriging. The block size is the same as the simulation scale in the horizontal plane and the entire model size in the vertical direction. This mimics the case of seismic data in a relatively thin stratigraphic interval. The results of block kriging and cokriging are very similar in this case because the correlation coefficient in the block kriging is very high.

Fig. 5 shows the reproduction of the areal proportion in category 1 over all slices (averaged over 10 realizations for numerical stability). The reference proportion is shown at the top. The influence of the wells can be seen in all cases. In the simple kriging model (upper left), we only see the influence of the wells since the trend model is not used; the reproduction of the trend model is quite poor. The ordinary kriging model (upper center) also only uses the wells, but the local estimation of the mean leads to fairly good results. We note that ordinary kriging works quite well with indicators because the kriging variance is not used (as it is in Gaussian simulation). The LVM, cokriging and block kriging models all look quite good. The Bayesian updating and the permanence of ratios have too many category 1 blocks in the central area.

There is a large amount of sensitivity in the selection of the *best* option. The results of block kriging look good in this case. The secondary data is not always so smooth; another algorithm could work better in the case of a secondary variable arising from geophysical measurements. An advantage of the `BlockSIS` program is that many different algorithms can be easily tried.

5. Conclusions

Sequential indicator simulation is a useful categorical variable simulation tool. There are a number of important implementation choices, particularly in presence of soft secondary data. The `BlockSIS` program implements most of the available techniques including block (co)kriging. This program also integrates a common image cleaning algorithm using the covariance lookup table and conditioning data that are readily accessible inside the simulation algorithm.

Appendix A. Supplementary Materials

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cageo.2006.03.005.

References

Deutsch, C.V., 2002. Geostatistical Reservoir Modeling. Oxford University Press, New York 376pp.

Journel, A.G., 2002. Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses. Mathematical Geology 34 (5), 573–596.

Further reading

Alabert, F.G., 1987. Stochastic imaging of spatial distributions using hard and soft information. Master's Thesis, Stanford University, Stanford, CA.

Deutsch, C.V., 1998. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. Computers & Geosciences 24 (6), 551–562.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library: and User's Guide, second ed. Oxford University Press, New York, NY, 369 pp.

Doyen, P.M., et al., 1994. Bayesian sequential indicator simulation of channel sands from 3-D seismic data in the Oseberg Field, Norwegian North Sea. In: Proceeedings of 69th Annual Technical Conference and Exhibition, New Orleans, LA, SPE paper # 28382, pp. 197–211.

Galli, A., et al., 1994. The pros and cons of the truncated Gaussian method. In: Armstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulations. Kluwer Academic Publisher, Dordrecht, pp. 217–233.

Goovaerts, P., 1994. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. Mathematical Geology 26 (3), 385–410.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, NY 496 pp.

Gómez-Hernández, J.J., Srivastava, R.M., 1990. ISIM3D: An ANSI-C three dimensional multiple indicator conditional simulation program. Computers & Geosciences 16 (4), 395–410.

Journel, A.G., 1993. Nonparametric estimation of spatial distribution. Mathematical Geology 15 (3), 445–468.

Journel, A.G., Isaaks, E.H., 1984. Conditional indicator simulation: application to a Saskatchewan uranium deposit. Mathematical Geology 16 (7), 685–718.

Journel, A.G., Alabert, F.G., 1988. Focusing on spatial connectivity of extreme valued attributes: stochastic indicator models of reservoir heterogeneities, SPE Paper 18324.