Spatial Interpolation of Forest Conditions Using Co-Conditional Geostatistical Simulation

H. Todd Mowrer

Abstract.—In recent work the author used the geostatistical Monte Carlo technique of sequential Gaussian simulation (s.G.s.) to investigate uncertainty in a GIS analysis of potential old-growth forest areas. The current study compares this earlier technique to that of co-conditional simulation, wherein the spatial cross-correlations between variables are included. As in the earlier study, uncertainties were assessed across 500 independent spatial Monte Carlo realizations for each of three variables of interest (quadratic mean stand diameter, age of dominant and co-dominant trees, and percent canopy cover). Potential old-growth for the study area was estimated for each set of these perturbed realizations using a simple GIS analysis. An uncertainty histogram was created by adding the 500 realizations on a cell-by-cell basis. Results were compared using empirical confidence regions from the upper percentiles of the histogram for each study. For uncertainty assessment using these particular co-located spatial data, co-conditional simulation creates intuitively less desirable results, and does not appear to provide any advantage over independent realizations using s.G.s.. For other data sets, where all variables are not measured at each location, improvements may result.

The advent of inexpensive georeferencing through the use of global positioning systems (GPS), and visualization and manipulation of spatial data with geographic information systems (GIS) have greatly facilitated the application of geostatistical techniques to forestry. Moreover, the exact positioning of measurement locations within forest landscapes allows this point information to be analyzed on a continuum ("continuous model of spatial variation," c.f. Heuvelink, 1998) instead of averaging plot measurements within stand polygons that are assumed to be homogeneous ("discrete model of spatial variation," ibid.). Geostatistical techniques take advantage of spatial auto-correlation between variables measured at multiple locations, as quantified by the variogram (Isaaks and Srivastava 1989). To obtain unbiased estimates at intermediate unmeasured points, the model variogram is then used in the kriging process to determine the minimum variance weighted linear combinations of measured values from nearby locations. Sequential Gaussian simulation (s.G.s.) (Deutsch and Journel 1992) calculates the kriging mean and variance at each unmeasured location, and draws a random deviate from the resulting distribution to simulate the value at that point. Mowrer (1994) applied sequential Gaussian simulation to estimate forest crown cover uncertainties for propagation through raster-based GIS. Subsequently, Hershey *et al.* (1995) compared ordinary kriging, indicator kriging, and s.G.s. for mapping tree species distributions across the State of Pennsylvania. In 1997, Mowrer applied the s.G.s. process to create independent realizations for three input variables (quadratic mean diameter, mean age of dominant and co-dominant stems, and percent crown cover). A simple GIS analysis demonstrated the propagation of uncertainties from interpolation of plot location data on estimates of potential old-growth stands. The results of 500 realizations for each variable were propagated through the GIS analysis, and the results were added on a cell-by-cell basis to create an uncertainty surface for the resulting old-growth areas.

Cokriging (Carr *et al.* 1985) may improve the estimation process over kriging by including cross-correlations. This enabled the development of co-conditional simulation (Carr and Myers 1985), which first creates independent realizations across grids for each variable of interest, and then uses cokriging to include the auto- and cross-correlations between variables in the conditioning process. The current study used the Carr and Myers (1985) program COSIM, with modifications, to repeat the process from the earlier s.G.s. study. For brevity sake, the term co-simulation will be used in the subsequent discussion to mean co-conditional simulation. The goal of the current study is to determine if the inclusion of spatial cross-correlations improves estimates with respect to the earlier analysis.

METHODS

The data for the analysis consisted of measurements on eighty-two 1/50-acre georeferenced plots distributed across a 300-acre subalpine watershed on the Fraser

Research Forester (Mensuration), USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA.

Experimental Forest, Fraser, CO, USA. All stems greater than 1 inch diameter at 4-1/2 feet above the ground were measured and cored through the pith for age. Percent crown cover was calculated from two perpendicular crown radii on each tree. An extensive description of data collection techniques is provided in Mowrer (1997). The current analysis consisted of four separate steps: (1) estimation of the variogram and cross-variogram model structure from the georeferenced plot data; (2) modification of the COSIM program and creation of 500 Monte Carlo cross-correlated realizations each, for quadratic mean diameter (q.m.d.) and age of dominant and co-dominant trees, and, separately, 500 independent realizations for percent crown cover; (3) repetition of the GIS analysis for each of the 500 sets of realizations of input values, creating 500 realizations of potential old-growth locations for the analysis area; (4) creation of empirical confidence regions and comparison to the earlier analysis.

Variogram Modeling

For all pairs of data values (v_i, v_j) of variable $\boldsymbol{v}(v_1, v_2, \ldots, v_n)$ separated by directed vector $\mathbf{h}_{i,j}$, the number of which is denoted N(h), the *sample variogram* is calculated over all (i,j) data pairs separated by distance **h**, using the following equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{i,j}=h}^{N(h)} (v_i - v_j)^2 .$$

When calculating a variogram from the sample data, a fixed increment to **h** is selected to maximize the structure in the variogram. This distance increment is usually called the lag spacing. Thus, the variogram value from the above formula is calculated for increasing increments of **h** reflecting larger and larger distances between v_i and v_j. The sample variogram is shown by dots in individual variograms of the matrix in figure 1. Variogram values ($\gamma|h|$) are plotted on the vertical axes, while the horizontal axes measure the magnitude of **h**. The *model variogram* is developed through the selection of one of several positive definite model forms (e.g., exponential, spherical, Gaussian, and power function). Positive definiteness is necessary to ensure the uniqueness of the kriging and cokriging results, discussed below. The selected variogram model form is then fitted (visually and mathematically) to the structure of the sample variogram, by varying the nugget, range, and sill parameters for the model (Isaaks and Srivastava 1989, Goovaerts 1996). Interactive variogram estimation software developed by Pannatier (1996) was used to calibrate the variograms for the current study.

The model forms for the three variograms (model type, range, nugget, sill, and isotropy) were maintained from the previous analysis to maximize comparability. Thus, because the quadratic mean diameter and age variogram models were exponential, their cross-variogram must be exponential, also. The equation for the exponential model (Pannatier 1996) is:

$$\gamma(h) = c\left[1 - \exp\left(-\frac{3h}{a}\right)\right]$$

Where a is the *range* parameter, the distance at which the average squared difference values reach a plateau and no longer increase. The *sill*, c, is the maximum variogram value on the vertical axis for this asymptote. A *nugget effect*, if it were present, would be indicated by a vertical model discontinuity at zero. The crown cover variogram model was maintained as spherical. The equation for the spherical model (Pannatier 1996) is:

$$\gamma(h) = \begin{cases} c\left(1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right) & \text{if } h \le a \\ c & \text{otherwise} \end{cases}$$

The model variograms are shown by solid lines in the individual variograms in figure 1.

Cross-Variogram

For all pairs of data values of variable $\mathbf{u}$ ($u_1, u_2, \ldots, u_n$) and of variable $\boldsymbol{v}$ ($v_1, v_2, \ldots, v_n$) separated by directed vector $\mathbf{h}_{i,j}$, the number of which are denoted N(h), the sample cross-variogram is calculated from the following equation:

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{(i,j)|h_{i,j}=h}^{N(h)} (u_i - u_j)\cdot(v_i - v_j).$$

The model cross-variogram is fitted to the sample cross-variogram in the manner described above for the variogram. In the current study, a cross-variogram for q.m.d. and mean plot age of dominant and co-dominant trees was calculated for use in the subsequent cokriging process in COSIM.

Simply stated, the idea of including cross-correlations in spatial analyses can be likened to explicit estimation of off-diagonal elements in the multivariate correlation or covariance matrix in classical statistics, instead of assuming independence (e.g., that off-diagonal elements are zero). In the geostatistical case, the diagonal elements of the spatial variogram matrix are more commonly referred to as spatial auto-correlations and are estimated by the variogram, while the off-diagonal elements are the

cross-variograms. Using Moran's bivariate spatial statistic permutation test (Czaplewski and Reich 1993), age of dominant and co-dominant trees was found to be significantly spatially cross-correlated with q.m.d. Figure 1 displays the variogram matrix for the current analysis, with the cross-variogram for age of dominant and co-dominant trees and q.m.d. as the off-diagonal element from the variograms for these two variables. Percent crown cover is spatially independent of the others, and thus is represented by the single variogram on the diagonal. In general, one must estimate $n(n-1)/2$ cross-variograms for n variables, because the variogram matrix is symmetric. Certain rules must be followed in determining the model structure of cross-variograms, to ensure that the overall structure remains positive semi-definite (Goovaerts 1997). This can make cross-variogram estimation for more complex models a formidable task.

Cokriging

Now consider for $\boldsymbol{u}$ $(u_1, u_2, \ldots, u_n)$ and $\boldsymbol{v}$ $(v_1, v_2, \ldots, v_m)$, that m may be greater than or equal to n. In this case, $\boldsymbol{u}$ is considered the primary variable of interest, and $\boldsymbol{v}$ the secondary variable. The additional values of $\boldsymbol{v}$ for which values of $\boldsymbol{u}$ are not measured may improve the estimates of $\boldsymbol{u}$ at all estimated locations. For each estimation location, $\hat{u}$, the value to be estimated is obtained from the minimum variance weighted combination of the values at

Figure 1.—*Variogram matrix, with the variograms for quadratic mean diameter (q.m.d.), age of dominant and co-dominant trees (AgeDC), and percent crown cover across the diagonal, and the cross-variogram for age and quadratic mean diameter (crown cover is considered to be independent). The vertical axes measure the variogram value ($\gamma|h|$) for increasing distance between sample locations (in feet) along the horizontal axis. The sample variograms are shown by dots and the model variograms by solid lines. Note that the overall matrix for variograms and cross-variograms is symmetric.*

nearby points. This estimation process and the necessary constraints on the values of the weights ($a_1, a_2, \ldots, a_n$) and ($b_1, b_2, \ldots, b_m$) are as follows:

$$\hat{u} = \sum_{i=1}^{n} a_i \cdot u_i + \sum_{j=1}^{m} b_j \cdot v_j$$

where,

$$\sum_{i=1}^{n} a_i = 1, \sum_{j=1}^{m} b_j = 0.$$

The cokriging estimator was used in the COSIM program as described in the next section.

COSIM Computer Program

The COSIM computer program developed by Carr and Myers (1985) first creates a grid of random deviates through a non-conditional Monte Carlo simulation process called "turning bands." The grids (one for each variable) are then "conditioned," or adjusted to the spatial correlation structure described by the variograms and cross-variograms using the cokriging algorithm.

Non-conditional Simulation

The first step of the non-conditional part of the COSIM program is the creation of standard normal unidimensional random vectors with a given spatial correlation. Each vector is split up into a number of line segments of equal length, each assigned a standard normal random value. This random value is then assigned to a three-dimensional band, defined by planes perpendicular to the end points of each of these short line segments along the vector. Fifteen of these vectors are created, along the lines connecting midpoints of opposite edges of a regular icosahedron (a 20-sided figure, each side consisting of an equilateral triangle). These 15 vectors therefore pass at equal angles through the center of a sphere, which is the center point for the simulated area. The vectors are simulated in five groups of three. An initial set of three vectors is rotated (by way of matrix multiplication) through five consecutive sequences to create the requisite 15 vectors, each with their associated bands of random values, thus the term "turning bands."

A user-specified grid is established that passes through the center of the sphere. The random contribution from each band passing through the node of each cell in this grid is summed to provide the random deviate for that grid cell, thus completing the Monte Carlo portion of the simulation. A linear transformation is then used to create independent realizations with the same overall mean and variance as the original data. (Readers interested in a more detailed explanation of the non-conditional portion of this simulation are referred to Journel and Huijbregts (1978, p. 498.)) In COSIM, this process is repeated independently for each variable in the simulation (in the current analysis, two: quadratic mean diameter and mean age of dominant and co-dominant stems).

Cokriging

Cokriging is then used to give these independent realizations the same overall spatial structure as the original data, reflected in the variograms and the cross-variograms. As described above, the variograms and cross-variograms are used in the cokriging process to determine the minimum variance linear combination of weights for the conditioning data. A weighted moving average of conditioning data (the actual values of quadratic mean diameter and mean age at nearby plot locations) is thereby used to estimate values at unmeasured locations on the grid. This same set of weights is also applied to the random deviates (created in the step above) closest to the location of each conditioning datum. Thus, the COSIM program creates two estimates for each cell in the grid: the cokriging estimate and a second estimate created by applying the same set of weights to the set of simulated random deviates closest to each data location. To create the conditioned (spatially correlated) random deviate at each unmeasured grid cell location, the latter estimate was subtracted from the original (independent) random value for that grid cell, and the cokriging estimate added to the result.

COSIM Program Modifications

It was necessary to modify the COSIM program to accept the exponential variogram and cross-variogram models. Fortunately, Journel and Huijbregts (1978) derived formulae for the exponential variogram model that could be used as a theoretical basis for modifying the FORTRAN code in COSIM for the exponential models. The program was then run to create 500 spatially correlated realizations for these two variables, a process taking approximately 44 hours on a 400 MHz. personal computer. (By comparison, 500 realizations with the Deutsch and Journel (1992) s.G.s. program takes approximately 38 minutes on the same machine.) The original COSIM model was also run to create 500 independent realizations using a spherical variogram model for crown cover.

GIS Analysis

As mentioned above, the GIS analysis followed the same procedure as that outlined in Mowrer (1997). The goal of the analysis was to delineate areas that simultaneously satisfied three sets of conditions: quadratic mean stand diameter greater than 10 inches at breast height (4.5 feet above the ground), average age of dominant and co-dominant trees greater than 220 years, and percent crown

cover greater than 40 percent. The study area was divided into a grid of 30-foot square cells: 116 in the north-south direction, and 160 in the east-west direction. Cells outside the delineated watershed polygon were deleted from the analysis. Each of 500 sets of three realizations (for quadratic mean diameter, age, and percent crown cover) from the co-simulation process was used to initialize the GIS analysis. For comparison, one realization for age from the current COSIM process is shown in figure 2a, and one realization for age from the s.G.s. process is shown in figure 2b. The GIS analysis created binary masks for each realization, whereby cells satisfying the previous criteria were given a value of one, and other cells were given a value of zero. By multiplying the binary cell values for these three images on a cell-by-cell basis, those cells with a resulting value of one would represent an area where estimated conditions satisfied the three criteria simultaneously. Otherwise a zero cell value would result. Summing the 500 output images resulted in an "uncertainty image" where the absolute number in each cell reflected the level of certainty of old-growth occurrence. These overall uncertainty images, resulting from the current co-simulation process and the previous s.G.s. process are shown in figures 3a and 3b, respectively. Figure 4 shows a comparison of the average cumulative variance across both the co-simulation and the

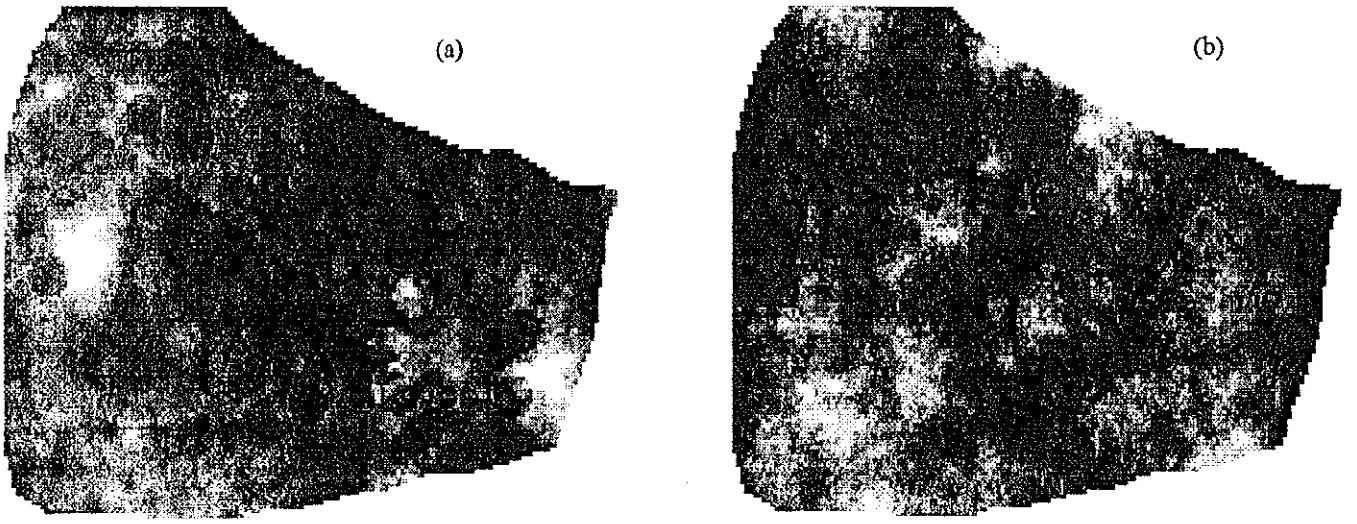

Figure 2.—*A single realization for age of dominant and co-dominant trees on the sample watershed from the conditional co-simulation program (a), and from the sequential Gaussian simulation program (b). Note striations across figure (a) and definite circles at sampled points. Darker gray-tones indicate higher values for age.*

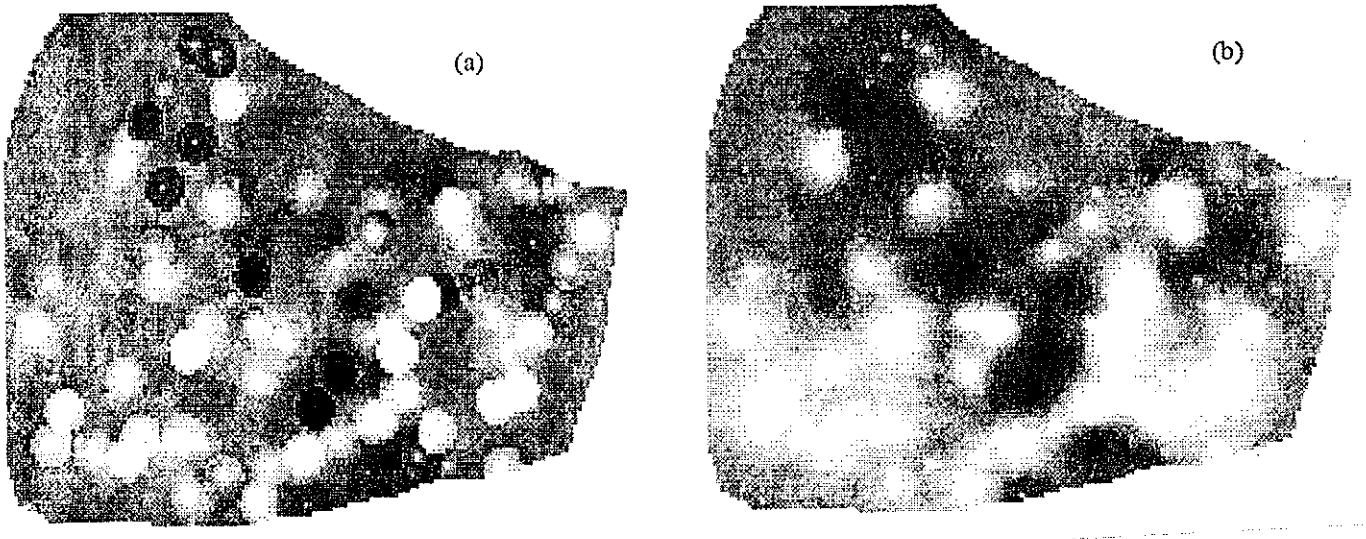

Figure 3.—*"Uncertainty images" for potential old-growth areas on the study watershed from the conditional co-simulation program (a) and from the sequential Gaussian simulation program (b). Darker gray-tones indicate higher probabilities. Note that while striations are absent from figure (a), the definite circles around the sampled points remain.*

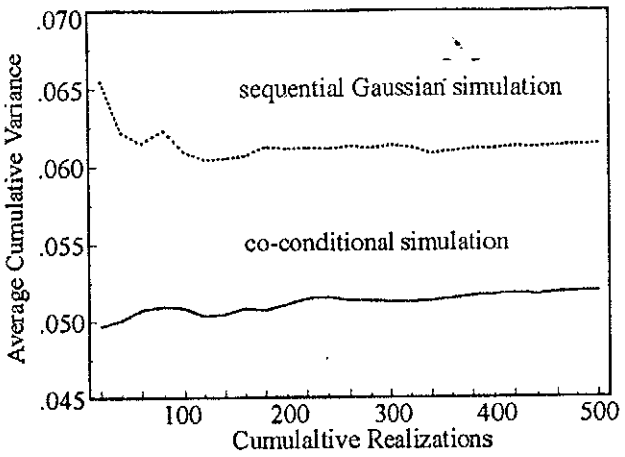

Figure 4.—Average cumulative variance calculated across identical pixel locations and summed over multiple realizations of the co-simulation and the s.G.s. Monte Carlo processes.

s.G.s. realizations. Both have reached an adequately smooth, level value at 500 realizations, indicating stable estimation.

COMPARISON OF RESULTS

To compare the results more closely, "empirical spatial confidence regions" for the two methods are presented in figures 5a (co-simulation) and 5b (s.G.s). These were constructed by selecting cells for the top 90, 95, and 99 percent of the image histogram, and shading them with increasingly darker shades of gray. The delineation of the confidence regions for the co-simulation algorithm is much more geometric than that for s.G.s.. The circles of higher probability surrounding certain measurement points are very pronounced, and similar to the appearance one obtains by simply depicting kriging variances. Visual inspection of an individual turning bands realization (fig. 2a, for example) indicated notable striations, which one would not expect for forestry applications. These have "averaged out" in the cumulative image (fig. 5a).

A puzzling artifact present in the co-simulation image is the inclusion of one or more pixels of low probability in the center of a circle or ring of higher probability pixels. This is difficult to explain in either a geostatistical or ecological context, and appears to be a result of the way the simulation results are combined in the COSIM program, as described above. After careful and protracted scrutiny of the COSIM FORTRAN code, no method was apparent whereby the original data values were retained at their original locations. This is certainly contrary to current descriptions of conditional simulation whereby "resulting realizations honor data values at their locations" (Deutsch and Journel 1998). By contrast, the s.G.s. analysis provides "empirical spatial confidence regions" that are continuous, and understandable from both an ecological and geostatistical perspective.

CONCLUSIONS

While the notable striations, apparent through visual inspection of an individual turning bands realization (fig. 2a) are impossible to justify for forestry applications, they might be desirable for geological applications where linear strata are common. Deutsch and Journel (1998)

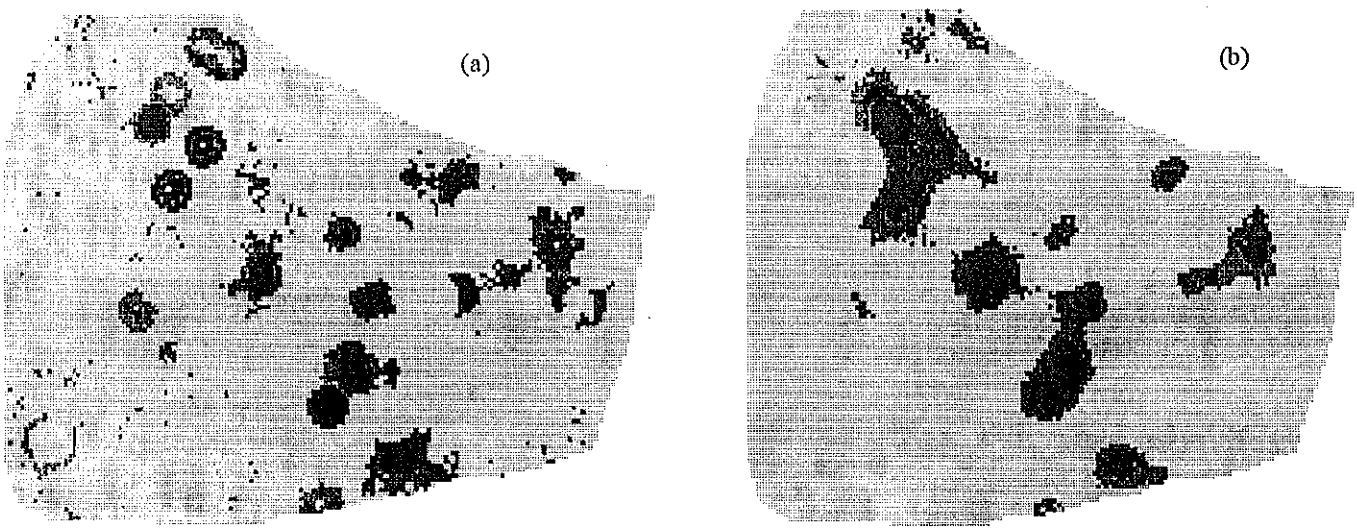

Figure 5.—*Empirical spatial confidence regions for old-growth locations on the study watershed created from the images in figure 3 by selecting the pixels from the image histogram with the top 99 percent (black), 95 percent (dark gray), and 90 percent (medium gray) from the conditional co-simulation program (a) and from the sequential Gaussian simulation program (b).*

note that the striations or "artifact banding" of the turning bands algorithm is due to the coarse partitioning by 15 bands. However, in the current study these striations do not reflect any characteristics of the data. The fact that some of these lines appear consistently lighter or darker than those surrounding them indicates that the turning bands Monte Carlo process does not create completely independent sets of random deviates along these lines. Improving the random number generator and/or increasing the number of lines in the turning bands algorithm would likely minimize these striations. However, in light of the symmetric appearance and seemingly contradictory overall appearance of the empirical confidence regions for the co-simulation technique, efforts to improve the COSIM program would likely not be worthwhile.

The overall symmetrical and circular dispersion of the resulting areas of old-growth is also difficult or impossible to justify ecologically. These seem more the result of the fact that conditioning through cokriging takes place after the randomization process in the co-conditional simulation algorithm. The results from s.G.s. are much more satisfying intuitively and seem much more likely to reflect approximate conditions than are the symmetric circles in the co-simulation analysis. For these data, where all variables are measured at each plot location, the co-simulation algorithm has little to offer. At this point there is no generally available software that provides co-conditional sequential Gaussian simulation. Considering the advantage that cokriging has to offer with regard to improved estimation through extensively measured covariates, such software would be a valuable addition to the geostatistical techniques.

ACKNOWLEDGMENTS

The author wishes to thank the following reviewers for their helpful comments in improving this paper: Ronald E. McRoberts, North Central Research Station, St. Paul, MN, USA, and Gerhard Hunner, Department of Forest Sciences, Colorado State University, Fort Collins, CO, USA.

LITERATURE CITED

Carr, J.R.; Myers, D.E. 1985. COSIM: A FORTRAN IV program for coconditional simulation. Computers and Geosciences. 11(7): 675-705.

Carr, J.R.; Myers, D.E.; Glass, C.E. 1985. Cokriging—a computer program. Computers and Geosciences. 17(5): 111-127.

Czaplewski, R.L.; Reich, R.M. 1993. Expected value and variance of Moran's bivariate spatial autocorrelation statistic for a permutation test. Res. Pap. RM-309. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 13 p.

Deutsch, C.V.; Journel, A.G. 1992. GSLIB: Geostatistical software library and user's guide. New York, NY: Oxford University Press. 340 p.

Deutsch, C.V.; Journel, A.G. 1998. GSLIB: Geostatistical software library and user's guide, 2d ed. New York, NY: Oxford University Press. 369 p.

Goovaerts, P. 1996. Geostatistics for natural resources evaluation. New York, NY: Oxford University Press. 442 p.

Hershey, R.R.; Ramirez, M.R.; Drake, D.A. 1995. Exploring the geostatistical techniques available to create a map of tree species distribution: an example using forest inventory and analysis (FIA) data. In: Analysis in support of ecosystem management, proceedings of a workshop; 1995 April 10-13; Fort Collins, CO. Washington, DC: U.S. Department of Agriculture, Forest Service, Ecosystem Management Analysis Center: 20-30.

Heuvelink, G.B.M. 1998. Error propagation in environmental modeling. London: Taylor and Francis. 127 p.

Isaaks, E.H.; Srivastava, R.M. 1989. An introduction to applied geostatistics. New York: Oxford University Press. 561 p.

Journel, A.; Huijbregts, C.J. 1978. Mining geostatistics. London: Academic Press. 600 p.

Mowrer, H.T. 1994. Spatially quantifying attribute uncertainties in input data for propagation through raster-based GIS. In: Proceedings of GIS 94; 1994 February 21-24; Vancouver, B.C., Canada. Vancouver, B.C. Canada: Polaris Conferences: 371-382.

Mowrer, H.T. 1997. Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. Ecological Modeling. 98(1997): 73-86.

Pannatier, Y. 1996. VARIOWIN: Software for spatial data analysis in 2D. New York, NY: Springer-Verlag. 91 p.