USING SPATIAL STATISTICS TO LOCATE TREES WITH HIGH COMMERCIAL VALUE

Susan L. King[1]

ABSTRACT.—Geostatistics is a branch of applied statistics that models and estimates spatial patterns. The distribution of trees by species and size across a region, such as a state, form a spatial pattern. Based on the species, size, and quality of the tree, a dollar value is assigned to each tree. One hypothesis is that trees of a specific dollar value are located near other trees of similar economic value. This might occur due to similar growing conditions and land use histories. Because it is impossible to sample every tree on every acre, statistical interpolation procedures are used to estimate the value of the trees between sample points. Four different procedures were used: a moving window average, ordinary kriging, indicator kriging, and sequential gaussian simulation. Each procedure is best suited for a particular set of circumstances. Of the four procedures, sequential gaussian simulation is more versatile, providing information on medians and ranges as well as the average.

INTRODUCTION

Time and cost constraints do not allow for an exhaustive inventory on a few hundred acres of forest land much less across an entire state. The alternative to an exhaustive inventory is a sample. Sample plots are selected throughout the area of interest to obtain estimates of the forest resource. All of the trees within the boundaries of the particular plot are measured. A plot total for any tree attribute is the sum of that attribute for each individual tree on the plot. In this context, these totals are point samples. However, point samples provide limited information. They do not provide information on the attribute or phenomenon of interest between the sample points. Interpolation procedures provide these estimates at unsampled locations.

Geostatistics is a branch of applied statistics that concentrates on the description of spatial patterns and estimating values at unsampled locations. The basic premise of geostatistics is that, on average, observations closer together are more similar than those farther apart. Neighboring trees share similar soils, elevation, aspect, moisture levels, and land use histories. Thus, neighboring trees should be of similar value.

Value is a function of the species, size, and quality of the tree. Value is important to both forest planners and to timber companies. Obviously, timber companies are interested in the location of trees with a particular economic value in order to harvest and locate mills. Forest planners are also interested in the location of trees with a particular economic value for managing the forest resource. Four different procedures were compared to estimate the values of trees between sample points: a moving windows average, ordinary kriging, indicator kriging, and sequential gaussian simulation.

DATA

The data set was collected by the Northeastern Forest Inventory and Analysis (NEFIA) unit of the USDA Forest Service. The data set is from the 1989 forest inventory for Pennsylvania and is part of the Eastwide Forest Inventory Data Base. Forest inventories are designed to estimate the amount of forest area and volume within a

[1]Susan L. King, Operations Research Analyst, USDA Forest Service, Northeastern Forest Experiment Station, 100 Matsonford Road, 5 Radnor Corp Ctr Ste 200, Radnor, PA 19087-4585

specified statistical error. The data from these inventories can be used for other studies. The on-the-ground sample has both forest and nonforest plots. This study used 2,967 forested plots. The average distance between the plots was approximately 3 km.

The Pennsylvania Woodland Timber Market Report (1993 - 1996) gives the stumpage prices paid for pulpwood and sawtimber in the Commonwealth of Pennsylvania for each quarter of the year. Figure 1 shows the division of the state into four timber regions: Southwest (SW), Northwest (NW), Southeast (SE), and Northeast (NE). Prices are reported for 10 species groups in each region. The timber markets within each region are unique. For example, the quality and quantity of black cherry in the NW region surpasses that in the other three regions, and is some of the most valuable black cherry in the world. The number of reporting mills varies each quarter. There is also a seasonal variation in prices. To account for these differences, an average of the stumpage prices weighted by the number of reporting mills is taken over 13 quarters. Table 1 gives the price of each of the 10 species, by region. Species that were not members of one of the 10 species groups were classified with the group having the most similar growing and competitive characteristics. Only trees larger than or equal to 5 inches in diameter at breast height (dbh) were included in this study. Value is the product of the price, net board-foot volume for the sawlog portion of the tree and the net cubic-foot volume for the pulpwood portion of the tree, and any required conversion factors. The net cubic-foot volume of the tree is the difference between the gross volume and the amount of cull in the tree. Thus, the net cubic-foot volume reflects the quality of the tree.

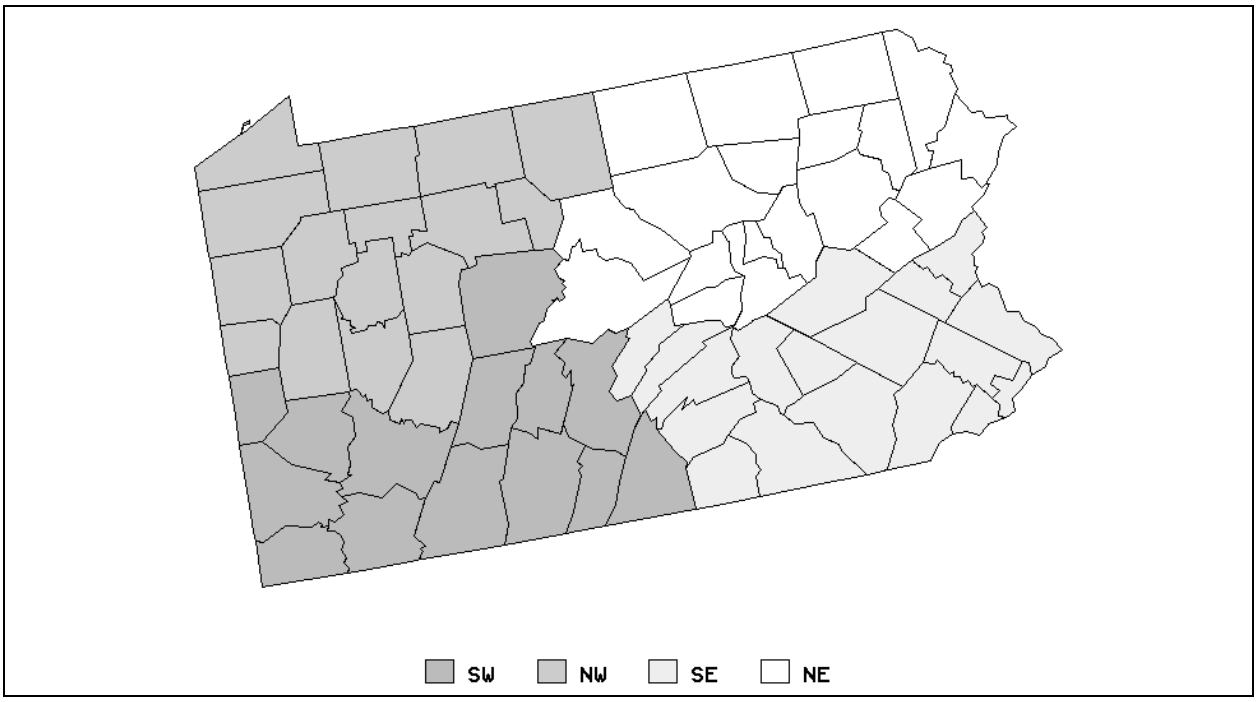

Figure 1.—Timber regions of Pennsylvania

Table 1.—Dollars/MBF by region and species

Species Group	SW	NW	SE	NE
Northern red oak	381.74	498.13	417.94	374.00
White oak	264.68	236.68	323.40	178.97
Mixed oak	244.79	224.68	338.77	175.84
Black cherry	387.13	817.34	280.55	695.66
White ash	280.91	277.45	384.66	311.42
Hard maple	160.46	200.53	218.51	203.51
Soft maple	136.49	121.80	159.75	118.67
Yellow-poplar	170.25	112.22	220.98	125.16
Misc. hardwoods	125.61	63.09	160.31	73.36
Pine-hemlock	90.46	41.81	117.79	63.60

For a sawlog tree, the value of the sawlogs plus the value of the pulpwood of the other parts of the tree were compared with the value of the tree as pulpwood alone. The value of the tree is the maximum of these two values. The value of a pole size tree is its pulpwood value. Each tree was weighted by its per-acre expansion factor. The value of all of the trees on a plot was summed. Figure 2 shows the location of the plots by value classes for the state. The classes are widely distributed across the state. The sampling is more intense in the NW region in the Allegeheny National Forest (ANF) than in the other regions of the state. The number of forested plots is lower in the southeastern part of the state due to the Philadelphia metropolitan area and the abundance of agricultural land. However, the volume per acre is high in this area due to low levels of harvesting. The forest land is fragmented and primarily privately owned. Yellow-poplar, northern red oak, and other oaks are the most valuable species in the SE region. This area was not as affected by the gypsy moth outbreak in the 1980's as was other parts of the state.

Figure 2.—Spatial distribution of the species

The ANF in the NW region covers the counties of Warren, McKean, Forest, and Elk. It is managed land. Widmann (1995) states that due to the management practices, the ANF has 15 percent of its stands in the sapling-seedling class, 12 percent in the poletimber class, and 73 percent in the sawtimber class. Poletimber stands were thinned to hasten the shift of these stands into the sawtimber class. Black cherry, red maple, and sugar maple are the top sawtimber species in the ANF. Much of the black cherry is growing in even-aged, overstocked stands, approaching 90 years of age. Black cherry, northern red oak, and red maple are the predominant species to the south and west of the ANF in the NW region.

The SW region contains the most rugged and highest mountains in the state. The mountains form a pattern of ridge and valleys that extend into the Pocono Mountains in the SE and NE regions. The steep slopes are often forested and state owned, while the valleys are agricultural areas. Although, black cherry is a top sawtimber species in the far southwestern counties, the ridge and valley region is dominated by oaks. The oaks were heavily hit by drought and gypsy moth in the 1980's. This led to salvage harvesting. The Pocono region contains the major anthracite coal fields of the country. The logging and coal industry wreaked heavy damage in this region around the turn of this century. Small trees were repeatedly harvested for mine props.

Powell and Considine (1982) state that the extreme part of the NE region was glaciated. Northern hardwoods rather than oaks grow in former glaciated regions. The top sawtimber species are sugar maple, red maple, and eastern hemlock. White ash is also a predominate species in this region.

Both the activities of man and nature affect the distribution of the species. In most situations, nature provides a gradual change over the landscape, but the activities of man cause more abrupt change. This impacts the basic assumption of geostatistics that neighboring sample plots are similar. Therefore, not all of the local variation can be modeled at the current sampling intensity of the FIA plot samples.

Figure 3 shows the distribution of the plots by $500 stumpage classes in Pennsylvania. The species are heavily skewed toward the lower dollar classes. The plots in the extreme higher dollar classes are in the NW region and have large black cherry trees.

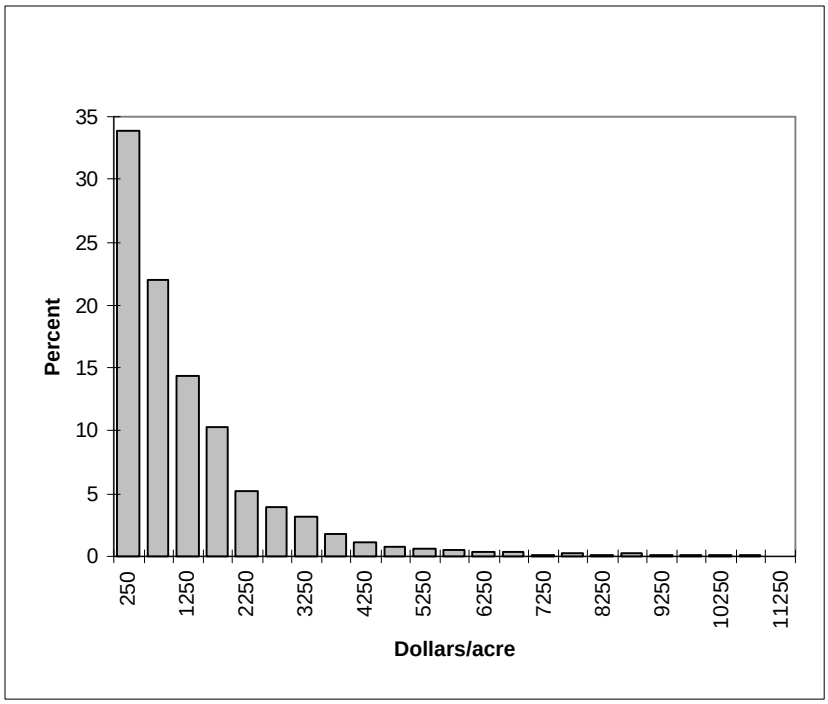

Figure 3.—Histogram of the sample data in $500 classes

INTERPOLATION AND SIMULATION TECHNIQUES

A moving window statistic is a descriptive interpolation technique. A state or region is subdivided into windows or blocks of equal size. The size of the window should be large enough to capture several observations and produce reliable statistics, yet small enough to capture the local variation. These are conflicting goals. Usually, there is not enough data for every single window to have a sufficient number of observations and produce reliable statistics. The solution is to extend the window so that adjacent windows share some of the same data. Any descriptive statistic such as the mean, median, variance, standard deviation, and ranges can be calculated and mapped. Figures 4a and b show the moving average and variance for the valuable species in Pennsylvania. (Maps are placed together for comparison: so, the figures are not referenced in numerical sequence.) The window size was 5 km x 5 km and the search radius was 20 km. Knowing and mapping the uncertainty of any estimate is critical before making any management decision. Are the areas of high variability due to natural variation in the phenomenon or are they due to very few observations in the region? The high variability in the ANF region is due to plots with large valuable black cherry trees being located near other plots that are less valuable. The variability in the southeastern part of the state is due to the relatively few forested plots.

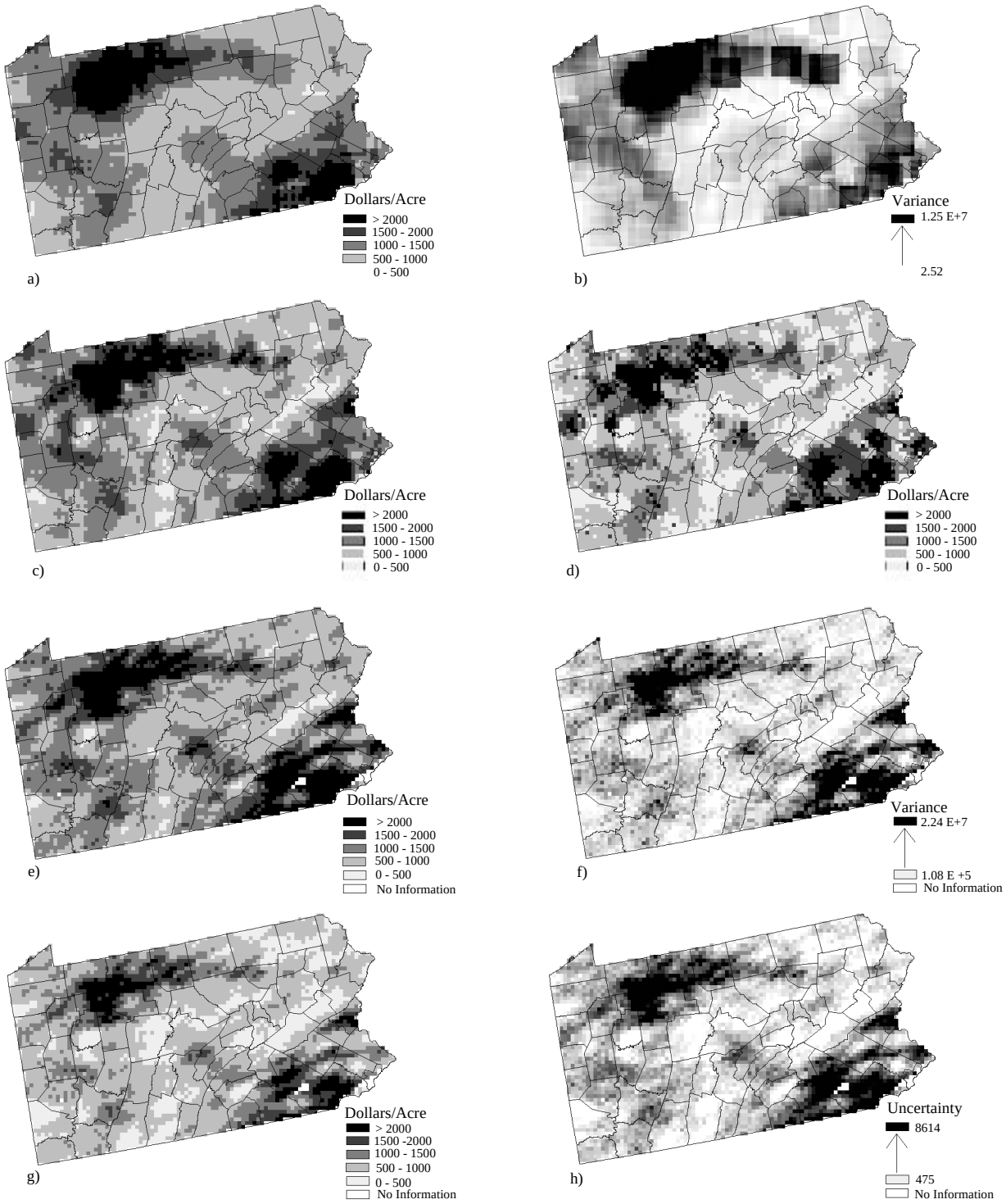

Figure 4.—(a) Moving window mean, (b) Moving window variance, (c) Ordinary kriging, (d) Indicator kriging, (e) SGS mean, (f) SGS variance, (g) SGS Median, (h) Range of uncertainty about the median.

Ordinary kriging, indicator kriging, and sequential gaussian simulation all require a model of spatial continuity. The tools for modeling spatial continuity are the variogram, correlogram, and covariance. Isaaks and Srivastava

(1989) and Rossi et al. (1992, 1994) are three excellent references on variogram models. An experimental variogram models the average degree of similarity between samples as a function of the separation distance and direction. The “nonergodic” or spatial covariance model accounts for the differences in local means. The “nonergodic” correlogram filters the differences in both the local mean and the local variance. All three models are usually plotted as a function of distance alone. Both the covariance and the correlogram models may be transformed to look like variograms. Any of the three models may be used in the kriging or simulation routines. Developing a theoretical and an experimental variogram is perhaps the most critical and certainly the most time consuming aspect of geostatistics. In the first phase, an experimental variogram is built from the sample data. In the second phase, a theoretical variogram is built using linear combinations of the exponential, gaussian, and spherical functions. In kriging, there is a requirement that the system of equations be positive definite. These functions and their linear combinations meet this requirement.

Figures 5 and 6 are variograms of the value attribute for all species in the state. A variogram, covariance, or correlogram rises and then levels off at its sill. The sill is the population variance. Often as in Figures 5 and 6, the variogram is standardized by dividing each variogram value by the population variance. The distance at which the variogram reaches its sill, 147 km in Figure 5, is called the range. Beyond the range, the samples are no longer correlated and are independent. The intercept of the variogram on the vertical axis is called the nugget. Ideally, the nugget should be zero. A nonzero nugget indicates either sampling error or that there is spatial variability below the minimum lag, which cannot be measured with the current sampling scheme. A lag is a distance class. Correct selection of the lag is critical in variography. A good starting point is the average distance between the samples. However, many factors such as the total number and arrangement of samples, the goals of the study, and the number of pairs available at small lags affect the lag selection.

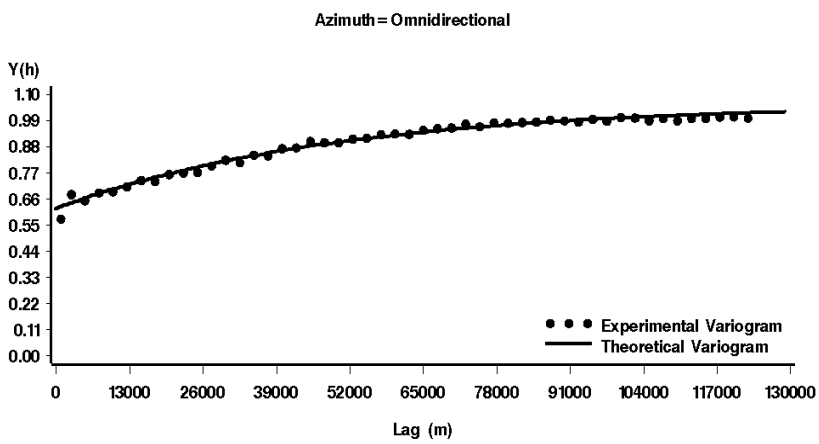

Figure 5.—Omnidirectional variogram for all species

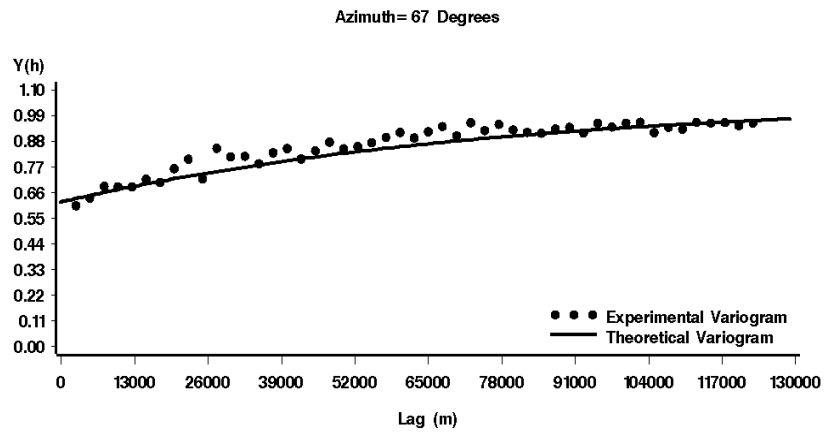

Figure 6.—Variogram for all species - shows anisotropy

The nugget in Figure 5 is 0.6. This means that 60 percent of the variability is unexplained. In practice, it is desired that the nugget be below 0.5. Still, there is enough of the spatial variability explained in the model to warrant the use of geostatistical techniques. If the variogram is constant, then using geostatistical techniques should be equivalent to a moving average procedure. Figure 6 illustrates another point. This variogram takes into account direction, whereas Figure 5 is an omnidirectional variogram. The range is 232 km for the directional variogram. The difference in ranges between the two variograms indicates geometric anisotropy and must be taken into account in kriging and simulation.

All ordinary kriging (OK) routines are weighted moving averages. OK routines assign a value to a particular location or a block. The observations closer to the estimated point in both distance and direction are weighted more heavily. Rossi et al. (1994) states that OK has three advantages over more traditional point estimation techniques such as inverse weighting, triangulation, and moving window statistics. First, OK can provide estimates, which are either higher or lower than the sample values for a specific location. Second, OK takes advantage of both distance and direction among the samples to determine its weights. Third, OK attempts to minimize the variance of the expected error. Figure 4c is the OK map for the estimated timber values in Pennsylvania. The block size is 5 km x 5 km and the search radius is 20 km. The kriging variance provided by the algorithm is in general independent of the data values and cannot be used as a measure of estimation accuracy. The only exception is if the data are from a multivariate normal distribution.

Indicator kriging (IK) is a variant of OK. IK divides the data set into two classes. One class is above the cutoff of interest and the other class is below that cutoff. A cutoff of interest in this project might be to model the location of trees worth more than $2,000/acre. At least 10 percent of the population must be on one side of the cutoff. IK requires a special variogram called an indicator variogram. The indicator variogram is binomial, assuming the values of 0 or 1. IK provides an estimate of the probability that the location's value is 0 or 1. IK uses either OK or another type of kriging, simple kriging (SK), to estimate the probabilities. An advantage of both IK and OK is that no specific distribution of the data is required. However, neither technique produces a variance estimate. The IK map for the valuable species in Figure 4d was made by overlaying an IK map for each cutoff: 500, 1,000, 1,500, and 2,000.

Sequential Gaussian Simulation (SGS) is a stochastic conditional simulation, also known as stochastic imaging technique. Rossi et al. (1993) provides an excellent review of stochastic conditional simulation, sequential gaussian simulation, and indicator sequential simulation. The goal of stochastic conditional simulation is to create multiple, equally probable images of the variable or phenomenon of interest. Each image is called a realization. Each realization should recreate the "vital" statistics of the sample population, i.e. the sample data's moments. The basis of stochastic conditional simulation is Monte Carlo simulation and Bayes' theorem. SGS applies the sequential

simulation procedure using the multivariate normal distribution. An assumption is made that the multivariate normal distribution accurately describes the spatial characteristics of the data. Typically, the original data are not normally distributed. The original data must be normal scored before the SGS process, and the results are then back-transformed. SGS also assumes the transformed data are multivariate normal. Bivariate normality may be checked, but multivariate normality must be assumed.

Using SGS, 200 realizations were created for the estimation of the location of trees with high commercial value in Pennsylvania. Each 5 km x 5 km block has 200 estimates of the dollars per acre of the trees in that block. These estimates form a distribution. For each block, summary statistics such as the mean, median, variance, standard deviation, and ranges are calculated and mapped. Figures 4e and 4f show the mean and variance for the location of valuable trees in Pennsylvania. Figures 4g and 4h show the median and the difference between the 84th and the 16th percentile. The difference between the two percentiles is a measure of the variability.

RESULTS AND CONCLUSIONS

All of the procedures show the same large-scale trend. The trees of high commercial value are in southeastern Pennsylvania and in the ANF region. The ridge and valley area has trees of lower commercial value. These results correspond to the land use and species distributions as previously discussed.

All of the procedures smooth the data. In any smoothing process, the high-population values are underestimated and the low-population values are overestimated. The goal for the moving windows average, OK, and SGS mean is to be correct on average. IK has a different objective. IK attempts to retain the same proportion of estimated data points above and below each cutoff as in the original data. Table 2 compares the descriptive statistics for the interpolation procedures with the original data. The mean of the 200 SGS realizations, 1,299.36, is closest to the mean of the original data, 1,289.05. A comparison of the moments illustrates the smoothing. OK detects more extreme values than SGS mean. Another way to compare the procedures is to use the block estimates for each procedure as input for an experimental variogram. These new experimental variograms are compared with the original experimental variogram in Figure 7. A single realization of the SGS procedure best reproduces the original variogram. Unfortunately, a single realization is only one possible picture. Many realizations are required before any level of uncertainty can be assigned to each block. SGS mean and OK equally re-created the original variogram. Due to the smoothing, it was not expected that any of the procedures would closely re-create the variogram of the original data. The ultimate test in comparing the procedures is to examine the maps. The moving average map is more smoothed than the maps from OK and SGS mean. However, the map captures the major trends and does not require any variography. An estimate of the variance is also available. Both OK and SGS require variography, which is time consuming. SGS is more computer intensive than OK, requiring more temporary storage space and computation time. However, with today's computers these are not significant factors in dismissing SGS. SGS produces a variance and other output statistics. OK only gives an average. From the maps in Figures 4c and 4e, SGS maintains more local variability than OK. SGS detected the absence of forested plots in Philadelphia County and in the center of Lancaster County. Thus SGS is the better procedure. Riemann Hershey et al. (1997) also concluded that SGS was superior to OK and multigaussian kriging in their study of tree species distribution in Pennsylvania.

Table 2.—Comparison statistics for the original data and the averaging techniques

	Original data	Moving average	OK	SGS mean
Mean	1289.09	1234.87	1235.33	1299.36
Skew	3.03	1.62	1.84	1.76
Variance	2218370	393941	481590	478747
CV	115.54	50.83	56.18	53.25
Maximum	19088.31	4746.73	6631.79	6484.76
Q3	1707.35	1474.78	1540.46	1529.37
Medium	843.79	1082.29	1062.01	1115.56
Q1	339.00	800.43	757.74	833.67
Minimum	0.00	7.76	6.48	254.08

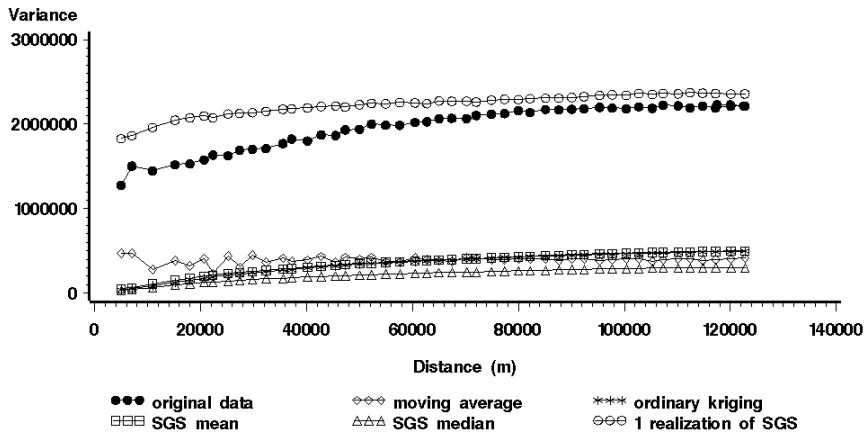

Figure 7.—Comparison of variograms from estimated data sets

Both the IK map and the SGS median map provide different views of the phenomenon. In appearance, these two maps are more closely related than the maps for SGS mean, OK, and the moving window average. The maps are not as smoothed. The far western counties receive a lower estimate with these procedures. Rossi et al.(1993) state that when the conditional probability distribution at each simulation point is not normal or symmetric, a more robust statistic is the SGS median. The ultimate decision on which map is better rest with the user and his or her objectives in using the map.

The current maps do not differentiate between the trees of high commercial value in the ANF and in southeastern Pennsylvania. One area is heavily forested and the other area is sparsely forested. The interpolation procedures assign values to areas that are not forested. One solution is to overlay in a GIS the map of trees of high commercial value with a kriged map of the forest/nonforest areas. The forest/nonforest map could be from the FIA photo interpretation points or from satellite imagery such as Landsat Thematic Mapper (TM) or Advanced Very High Resolution Radiometer (AVHRR). The FIA photo interpretation points are at a finer scale than the on-the-ground sample data. The nonforest locations are then masked. This map would provide a more useful decision tool.

There is enough spatial structure to support the hypothesis that trees of a specific economic value are located near other trees of similar economic value. The maps make sense. The results would be locally more accurate if data were available at a smaller scale.

LITERATURE CITED

ISAAKS, E.H., AND R.M. SRIVASTAVA. 1989. *An introduction to applied geostatistics*. Oxford University Press.

PENNSYLVANIA STATE UNIVERSITY. 1993-1996. Pennsylvania Woodlands Timber Market Report (First Quarter 1993 - First Quarter 1996). University Park, PA: The Pennsylvania State University, School of Forest Resources, Cooperative Extension Service..

POWELL, D.S., AND T.J. CONSIDINE, JR. 1982. *An analysis of Pennsylvania's forest resources*. Resour. Bull NE-69. Broomall, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

RIEMANN HERSHEY, R., M.A. RAMIREZ, AND D.A. DRAKE. 1997. Using geostatistical techniques to map the distribution of tree species from ground inventory data. In: *Modeling longitudinal and spatially correlated data: methods, applications, future directions*. eds. T. Gregoire et al. 187-198. Santa Clara, CA. Lecture Notes in Statistics. vol. 122. Springer-Verlag.

ROSSI, R.E., P.W. BORTH, AND J.J. TOLLEFSON. 1993. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. *Ecological Applications* 3(4):719-735.

ROSSI, R.E., J.L. DUNGAN, AND L.R. BECK. 1994. Kriging in the shadows: geostatistical interpolation for remote sensing. *Remote Sens. Environ*. 49:32-40.

ROSSI, R.E., D.J. MULLA, A.G. JOURNEL, AND E.H. FRANZ. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs* 62(2):277-314.

WIDMANN, R.H. 1995. *Forest resources of Pennsylvania*. Resour. Bull NE-131. Radnor, PA: U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 41p.