Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. *Numerical Recipes*, Boston, MA: Cambridge University Press.

Puri, M.L., and Sen, P.K. (1969). "A class of rank order tests for a general linear hypothesis," *Annals of Mathematical Statistics, 40*, 1325-1343.

Quintana, S. M., and Maxwell, S.E. (1994). "A Monte Carlo comparison of seven ε-adjustment procedures in repeated measures designs with small sample sizes," *Journal of Educational Statistics, 19*, 57-71.

Rogan, J.C., Keselman, H.J., and Mendoza, J.L. (1979). "Analysis of repeated measurements," *British Journal of Mathematical and Statistical Psychology, 32*, 269-286.

Romaniuk, J.G., Levin, J.R., and Hubert, L.J. (1977). "Hypothesis-testing in repeated-measures designs: On the road map not taken," *Child Development, 48*, 1757-1760.

Rouanet, H., and Lepine, D. (1970). "Comparisons between treatment in repeated-measures designs: ANOVA and multivariate methods," *British Journal of Mathematical and Statistical Psychology, 23*, 147-163.

SAS Institute, Inc. (1990). *SAS user's guide: Statistics*, Cary, NC: SAS Institute.

Sawilosky, S.S., Blair, R.C., and Higgins, J.J. (1989). "An investigation of the Type I error and power properties of the rank transform procedure in factorial ANOVA," *Journal of Educational Statistics, 14*, 255-267.

Serlin, R.C., and Harwell, M.R. (1989, April). A comparison of Hotelling's T^2 and Puri and Sen's rank test for the single-factor, repeated measures design. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Utts, J.M., and Hettmansperger, T.P. (1980). "A robust class of tests and estimates for multivariate location," *Journal of the American Statistical Association, 75*, 939-946.

Wilson, R.S. (1975). "Analysis of developmental data: Comparison among alternative methods," *Developmental Psychology, 11*, 676-680.

Vale, C.D., and Maurelli, V.A. (1983). "Simulating multivariate nonnormal distributions," *Psychometrika, 48*, 465-471.

Received September, 1995; Revised August, 1996

COMMUN. STATIST.—SIMULA., 26(2), 619-629 (1997)

KRIGING WITH LARGE DATA SETS USING SPARSE MATRIX TECHNIQUES

Ronald P. Barry
University of Alaska Fairbanks
Dept. of Mathematical Sciences
Fairbanks, Alaska 99775-6660
e-mail: ffrpb@aurora.alaska.edu

R. Kelley Pace
University of Alaska Fairbanks
Dept. of Finance
School of Management
Fairbanks, Alaska 99775-6068
e-mail: ffrkp@aurora.alaska.edu

Key Words: geostatistics, variogram, covariogram.

ABSTRACT

A major impediment to kriging with large data sets is the need to solve matrix equations with the large matrices that result from using variogram-based kriging equations. This is expensive both in computing time and memory. When the range of the variogram is small, use of covariance-based kriging equations and sparse matrix techniques can allow the kriging equations to be solved very efficiently. By fitting the variogram model and then using this to derive the covariance matrix, we keep the better estimation properties of the variogram, and can exploit the sparseness of the covariance matrix. We compare the relative efficiency of variogram-based and covariance-based kriging, using both real and simulated data. We also comment on the use of sparsity in kriging.

1. INTRODUCTION AND NOTATION

Researchers often measure a variable of interest at a set of known locations $\{s_1, ..., s_n\}$ in a region. One method of obtaining an estimated map of the variable at all locations in the region, along with a measure of the accuracy of this map in different parts of the region, is kriging. One of the most often used formulations of kriging makes the following assumptions:

- The variable of interest is a random function $Z(s)$ where s is any location in the region.
- $E(Z(s)) = \mu$ for all s in the region.
- For any two locations s_i, s_j in the region,

$$\mathrm{Var}(Z(s_i) - Z(s_j)) = 2\gamma(s_i - s_j).$$

The function $2\gamma(h)$ is called the variogram of the process. If such a function exists, the process is called intrinsically stationary.

If we are interested in predicting a value $Z(s_0)$ at some location s_0 using a linear combination of the measured values $Z(s_1), ..., Z(s_n)$, along with a requirement that the predictor be unbiased and that it have minimum squared prediction error, then we would use the kriging predictor:

$$\hat{Z}(s_0) = (\gamma + 1\frac{(1 - 1'\Gamma^{-1}\gamma)}{1'\Gamma^{-1}1})'\Gamma^{-1}\mathbf{z},$$

where $\gamma = (\gamma(s_0 - s_1), ..., \gamma(s_0 - s_n))$, $\mathbf{z} = (z(s_1), ..., z(s_n))$, and

$$\Gamma = \begin{pmatrix} \gamma(s_1 - s_1) & \gamma(s_1 - s_2) & \dots & \gamma(s_1 - s_n) \\ \gamma(s_2 - s_1) & \gamma(s_2 - s_2) & \dots & \gamma(s_2 - s_n) \\ \gamma(s_3 - s_1) & \gamma(s_3 - s_2) & \dots & \gamma(s_3 - s_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(s_n - s_1) & \gamma(s_n - s_2) & \dots & \gamma(s_n - s_n) \end{pmatrix}$$

The matrix Γ is both symmetric and conditionally negative definite (Cressie, 1993, p. 60).

The mean squared prediction error at s_0 is

$$\sigma^2(s_0) = \gamma'\Gamma^{-1}\gamma - (1'\Gamma^{-1}\gamma - 1)^2/(1'\Gamma^{-1}1).$$

If the random process is actually second order stationary, then the variogram has a *sill*:

$$\sigma^2_{sill} = \lim_{h\to\infty} \gamma(h)$$

If the random process does have a sill the kriging equations can be written using the matrix Σ of covariances:

$$\Sigma = \begin{pmatrix} \mathrm{Cov}(Z(s_1), Z(s_1)) & \mathrm{Cov}(Z(s_1), Z(s_2)) & \dots & \mathrm{Cov}(Z(s_1), Z(s_n)) \\ \mathrm{Cov}(Z(s_2), Z(s_1)) & \mathrm{Cov}(Z(s_2), Z(s_2)) & \dots & \mathrm{Cov}(Z(s_2), Z(s_n)) \\ \mathrm{Cov}(Z(s_3), Z(s_1)) & \mathrm{Cov}(Z(s_3), Z(s_2)) & \dots & \mathrm{Cov}(Z(s_3), Z(s_n)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(Z(s_n), Z(s_1)) & \mathrm{Cov}(Z(s_n), Z(s_2)) & \dots & \mathrm{Cov}(Z(s_n), Z(s_n)) \end{pmatrix}$$

and the vector $\mathbf{c} = (\mathrm{Cov}(Z(s_0), Z(s_1)), ..., \mathrm{Cov}(Z(s_0), Z(s_n)))$, where

$$\mathrm{Cov}(Z(s), Z(t)) = \sigma^2_{sill} - \gamma(s, t).$$

This gives the covariogram-based version of the kriging equations:

$$\hat{Z}(s_0) = (\mathbf{c} + 1m)'\Sigma^{-1}\mathbf{z},$$

$$\sigma^2(s_0) = \sigma^2_{sill} - (\mathbf{c} + 1m)'\Sigma^{-1}\mathbf{c} + m,$$

where

$$m = \frac{(1 - 1'\Sigma^{-1}\mathbf{c})}{1'\Sigma^{-1}1}$$

(Cressie, 1993, p. 123).

One nice feature of both the variogram- and covariogram-based kriging equations is that the time consuming computation of $\Gamma^{-1}\mathbf{z}$ and $\Gamma^{-1}1$ (or $\Sigma^{-1}\mathbf{z}$, $\Sigma^{-1}1$ and $\Sigma^{-1}\mathbf{c}$) need only be done once even if a large number of predictions $\hat{Z}(s_{01}), ..., \hat{Z}(s_{0p})$ are desired. Unfortunately, the time required to solve a *dense* (few nonzero entries) linear $n \times n$ system grows at order n^3, and the required memory is of order n^2. The usual approach to the

computational problems of large data sets when predicting at s_0 (Deutsch and Journel, 1992, p. 30) is to exclude observations somewhat distant from s_0. In contrast, sparse matrix techniques, by dramatically lowering time and storage costs, can employ all of the data.

2. THE VARIOGRAM VS THE COVARIOGRAM

Strong arguments have been made favoring the use of the variogram over the use of the covariogram in kriging. The two major arguments are that variogram estimation is less biased than covariogram estimation (Cressie, 1993, p. 69-73), and that the variogram is defined for some processes that are not second order stationary, such as the random walk (Cressie, 1993, p. 68). Since the covariance-based kriging equations are more restrictive than the variogram based ones, why should one use the covariances? The answer comes from observing the structure of the matrices Γ and Σ. The variogram matrix Γ, in all but pathological cases, is nonzero everywhere except the diagonal. On the other hand, if the variogram model has a sill ($\sigma^2_{sill} < \infty$), and a range (smallest t such that $\gamma(h) = \sigma^2_{sill}$ for all $h > t$) that is small compared with the size of the region, then the covariance matrix Σ has many zeros, since locations farther apart than the range will be uncorrelated. This results in a sparse matrix. This sparsity allows us to solve the kriging equations involving a very large number of observations. By first estimating the variogram from the data, and then computing the covariance matrix Σ, we can retain the low bias of variogram estimation along with the computational advantages of sparsity in Σ.

3. THE EFFICIENCY OF SPARSE MATRIX TECHNIQUES

The advantages of the covariance-based kriging equations becomes apparent when large data sets are used. For extremely sparse matrices such as Σ, solving the covariogram-based kriging equation $\Sigma^{-1}\mathbf{y} = \mathbf{x}$ can require

far less time and space than that required for solving the variogram-based kriging equation $\Gamma^{-1}\mathbf{y} = \mathbf{x}$. While there is some trade-off between time and space expenditure, a rule of thumb is that storage is a linear function of the number of observations, and, depending on the method of analysis and the structure of Σ, the time required can run from anywhere from linear in the number of observations to the third power of the number of observations (Bruaset, 1995, p. 3).

We conducted an investigation of the relative computational efficiency of working with the variogram-based equations versus using the covariogram-based equations. We generated a series of $p \times p$ regular lattices of locations on a square grid representing $n = p^2$ locations with a spacing of 0.2 unit between grid points. We ordered the observations sequentially from left to right in each row, continuing from the right-hand side of one row to the left-hand side of the next row. We assumed a spherical semivariogram model (Cressie, 1993, p. 61) $\gamma(h) = 10((3/2)(h/R) - (1/2)(h/R)^3$ for $h \leq R$ and $\gamma(h) = 10$ for $h > R$, where the range R of the process was either set to 0.3, 0.6, or 0.9 unit. For each number of observations and each range we obtained (or attempted to obtain) the sparse matrix Σ and the full $n \times n$ matrix Γ. Both the sparse and dense linear systems were solved in MATLAB (Matlab, 1993) using the standard solver (/). When the input is in sparse matrix form, this generic operator automatically uses algorithms that take advantage of sparsity. Since the matrix Σ is symmetric and has positive diagonal elements, MATLAB, by default, first reorders the observations using the symmetric minimum degree algorithm, and then performs a Cholesky decomposition (MATLAB User's Guide, 1993; Gilbert, Moler and Schreiber, 1992). All calculations were done on a 133 MHz Pentium PC with 64 megabytes of RAM. In Table I, the time required to compute the sparse matrix Σ and solve the linear system $\Sigma\mathbf{x} = \mathbf{y}$ is displayed for all three ranges, and from 100 to 10000 observations.

TABLE I

Time, in seconds, to solve dense and sparse matrix equations

n	density (%)	range	dense	sparse
100	7.84	0.30	0.06	0.00
400	2.10	0.30	1.76	0.11
625	1.36	0.30	8.13	0.22
900	0.96	0.30	20.26	0.28
1225	0.71	0.30	61.35	0.44
2500	0.35	0.30	(343)*	0.99
4900	0.18	0.30	(2168)*	2.86
10000	0.09	0.30	(15311)*	11.15
100	20.16	0.60	0.06	0.00
400	5.72	0.60	1.71	0.32
625	3.75	0.60	6.65	0.66
900	2.68	0.60	20.26	1.27
1225	2.01	0.60	61.30	2.08
2500	1.02	0.60	(343)*	5.28
4900	0.54	0.60	(2168)*	20.76
10000	0.26	0.60	(15311)*	60.64
100	44.20	0.90	0.05	0.11
400	14.00	0.90	1.76	1.20
625	9.36	0.90	6.65	2.52
900	6.69	0.90	20.27	4.61
1225	5.01	0.90	50.64	8.29
2500	2.55	0.90	(343)*	24.99
4900	1.33	0.90	(2168)*	73.65*
10000	0.66	0.90	(15311)*	733.26

*- Lower bound obtained via extrapolation.

The density is the proportion of non-zero entries. The locations were ordered first by column number and second by row number, though MATLAB reordered the locations using the symmetric minimum degree algorithm (George and Liu, 1981, p. 92).

When the range of the spherical variogram is 0.3, the covariance matrix is very sparse, with density approaching $9/n$ for large n, since each interior location is within 0.3 unit of 9 locations. Regressing log(time) on log(n) shows a very good fit for a line with slope 1.39 ($r^2 = 0.988$), indicating that the time to solve the equation is increasing on order $n^{1.39}$. All of these sparse computations were performed in RAM, instead of requiring slower virtual memory, so that more system memory would be required to preserve the $n^{1.39}$ order as the number of observations increases. When a range of 0.6 is used, the expected density approaches $25/n$ for large n. Again, regressing log(time) on log(n), we get a slope of 1.63 ($r^2 = 0.998$). For the largest range of 0.9, the density of nonzero elements approaches $69/n$ for n large. In this case, the slope is 1.67 ($r^2 = 0.999$) for n through 4900. At $n = 10000$, the computer used the hard disk for storage during the computation, slowing the computations down. In all cases, the memory requirements of sparse matrix techniques grew with the number of nonzero elements, linearly with n.

Table 1 also shows the time required when dense matrix algorithms are used. A regression of log(time) on log(n) yields a slope of 2.75 ($r^2 = 0.995$), implying an almost cubic increase in time requirements. Obviously the relative efficiency of sparse computations over dense computations increases rapidly with n. For $n = 2500$, 4900 and 10000, the computer was unable to perform the computations in RAM, and had to use the hard disk. This so slowed the computations that we were not able to solve the system of equations. The table contains extrapolated estimates of the computation time for large dense systems, if sufficient RAM were available to allow the computations to occur in system memory. Because of the high storage requirement of dense computations, the resulting memory problems are much worse than in the case of sparse computations. For example, in double precision a 10,000 observation system would require $10000^2 * 8$ bytes, or 800 megabytes to hold a single copy of the matrix. To make matters worse, many matrix computations require the storage of more than one copy of the matrix.

4. GEOCHEMICAL DATA

From the previous simulations, it is clear that using a sparse matrix and a covariance-based approach can lead to computational gains. However, the advantages of sparse techniques do not depend on the use of a regular lattice. The Geological Survey of Canada measured nickel concentrations at 916 sites in Vancouver Island as part of the National Geochemical Reconnaissance. This data set is described in Bailey and Gatrell (1995, p. 150). The sites are irregularly located, as can be seen below:

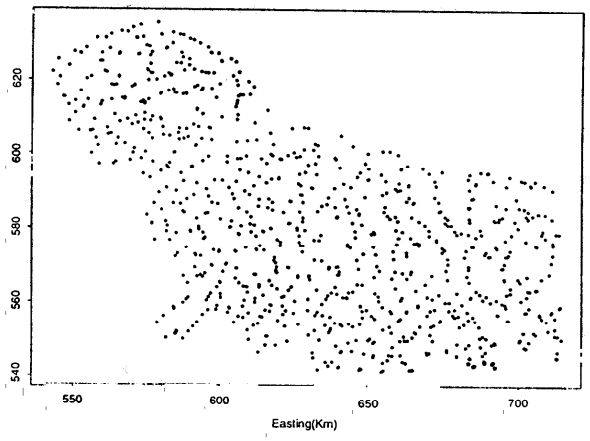

FIGURE I

Locations at which nickel concentration was measured on Vancouver Island.

Bailey and Gatrell fit a spherical variogram to the log of the nickel concentration, obtaining the covariogram model:

$$C(h) = \begin{cases} 0.529(1 - (3/2)(|h|/2.13) + (1/2)(|h|/2.13)^3), & \text{if } 0\text{Km} < |h| \text{ and } |h| \leq 2.13\text{Km}; \\ 0.827, & \text{if } |h| = 0\text{Km}; \\ 0, & \text{if } |h| > 2.13\text{Km}. \end{cases}$$

The resulting covariance matrix is very sparse, with only 2096 nonzero elements (about 29 kilobytes of storage). The full matrix requires 839056 elements (6712 kilobytes of storage). Solution of the linear equation $\Sigma x = 1$ requires 25.92 seconds using full matrix techniques, but only 0.06 second using sparse matrix techniques, the sparse techniques being 432 times as fast! While the dense techniques are still viable for this example, the advantages of sparsity in memory and speed would become even more dramatic as the number of observations grows.

5. DISCUSSION

• We do not estimate the covariogram directly, as variogram estimation has better estimation properties. To obtain sparse matrices, variogram models with a finite range must be used. Models such as the exponential model give non-zero (though often small) correlations for all pairs of points, and are thus unsuitable. The same holds true for the rational quadratic, power, and wave models in two-dimensions (Cressie, 1993, p. 61). The spherical variogram model is an example of a model that has a finite range. Barry and Ver Hoef (1996) have discovered a very flexible family of valid variogram models with finite range that can approximate any smooth variogram with finite range.

• As seen in the sparse matrix computations, smaller ranges lead to systems of equations that can be solved in an amount of time that grows at a slower rate. Thus the advantages of using sparse matrix algorithms should

be greatest for the residuals remaining after regression or detrending, since these tend to have less long-range correlation.

• Sparse techniques also work well with irregular data as long as measurements from most pairs of locations are uncorrelated. Unfortunately there probably will not be a natural ordering as in the regular lattice. It is possible that a poor choice of ordering could decrease the efficiency. A number of automatic reordering algorithms exist that can protect against the effects of a poor choice of order (George and Liu, 1981). In the analysis of regression with spatially autocorrelated errors (simultaneously specified Gaussian) models, exploiting sparsity leads to tremendous gains in computational efficiency, even though the locations are irregular (Pace and Barry, forthcoming).

• For sparse matrices, there are fast iterative methods that are generally even more conservative in their memory usage (Bruaset, 1995; Saad, 1996). Further investigation is needed to determine if iterative algorithms can allow more efficient solution of very large, sparse linear systems arising from geostatistical data.

6. CONCLUSION

The use of short range variograms (when appropriate), covariance-based kriging equations and sparse matrix techniques facilitates to perform kriging on data sets that are much too large for the conventional approach. For models with small range variograms, kriging with a sample size of 100000 may be feasible on a fast PC (in our simulation, for a range of 0.3, extrapolation yields an estimate of slightly more than two hours, and with 900000 nonzero entries the memory requirements should not be impossible to meet).

ACKNOWLEDGMENT

We would like to thank the Editor, the Associate Editor and two anonymous referees for their comments, and the University of Alaska for its generous research support. In addition, Pace would like to acknowledge support from the Center for Real Estate and Urban Economic Studies, University of Connecticut.

BIBLIOGRAPHY

Bailey, Trevor C. and Anthony C. Gatrell (1995). *Interactive Spatial Data Analysis*, Longman Scientific and Technical, Harlow England, 1995.

Barry, Ronald Paul, and Ver Hoef, Jay M. (1996). "Blackbox Kriging: Spatial Prediction without Specifying Variogram Models" *Journal of Agricultural, Biological and Environmental Statistics.* Forthcoming.

Bruaset, Are M. (1995). *A Survey of Preconditioned Iterative Methods*, John Wiley and Sons, Inc., New York.

Cressie, Noel A.C. (1993). *Statistics for Spatial Data, revised edition*, John Wiley and Sons, Inc., New York.

Deutsch, Clayton V., and Journel, André G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, Inc., New York.

George, Alan and Liu, Joseph W-H (1981). *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Gilbert, J.R., Moler, C, and Schreiber, R. (1992). "Sparse Matrices in MATLAB: Design and Implementation" *SIAM Journal on Matrix Analysis*, 13, p. 333-356.

Matlab User's Guide for Microsoft Windows (1993). Mathworks, Natick, MA.

Pace, Kelley, and Barry, Ronald. "Fast Spatial Autoregressions" *Statistics and Probability Letters.* Forthcoming.

Saad, Yousef (1996). *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, MA.

Received February, 1996; Revised July 199