USING A GEOGRAPHIC INFORMATION SYSTEM AND GEOSTATISTICS TO ESTIMATE SITE INDEX OF *PINUS RADIATA* FOR KAINGAROA FOREST, NEW ZEALAND

B. K. HÖCK, T. W. PAYN,

New Zealand Forest Research Institute,
Private Bag 3020, Rotorua, New Zealand

and J. W. SHIRLEY

Forestry Corporation of New Zealand,
P.O. Box 1748, Rotorua, New Zealand

(Received for publication 23 February 1993; revision 4 March 1994)

ABSTRACT

Site index is used as a measure of productivity for large plantation forests. Although site index had been calculated in less than half of the compartments in Kaingaroa Forest, data were fairly evenly spread. A Geographic Information System (GIS) was used to produce a contour map of site indices associated with compartment centre points. The limitations of estimation techniques within the GIS were highlighted by the difficulty of predicting values between contour lines. Instead, geostatistics, a statistical interpolation method, was adopted as it can estimate local values from data that varies spatially.

The variogram for site index in Kaingaroa Forest was fitted by a linear model up to 25 km. The parameters of this model were used in estimation (kriging) procedure. Values for 757 compartments were predicted, ranging from 18.8 to 34.3 m. The standard error ranged from 1.6 to 3.6 m, with a mean of 1.7 m. A jack-knifing procedure showed estimates to agree well with actual values.

It was concluded that linking a GIS with geostatistics allowed more effective use to be made of the GIS.

Keywords: site index; Geographic Information System; geostatistics.

INTRODUCTION

Kaingaroa Forest (Fig. 1), at 144 000 ha, is one of the largest planted forests in the Southern Hemisphere. To allow long-term planning of wood flows from the forest, its productivity must be known. One common index of forest productivity is the site index ($SI_{[20]}$)—the top height, measured in metres, that a stand of trees attains at age 20 years.

A site index map constructed in the 1970s by New Zealand Forest Service staff has remained substantially unchanged since then. The pedigree of this map, the data contributing

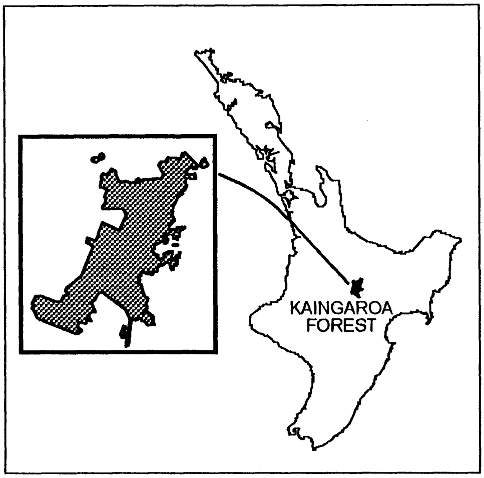

FIG. 1–Locality map.

to it, and the procedures used to derive the site indices are unknown (J.W.Shirley unpubl. data). A review of these site indices in early 1992 (T.M.Dale unpubl. data) indicated that some of them were not very reliable. This, together with the availability of many new data, prompted the call to update the site index map of Kaingaroa.

An initial assessment of the new data showed that more than half of the 1643 compartments in the forest still had no associated site index. In order to avoid costly field measurements, a method was sought for efficiently estimating values of site index for those compartments. This paper describes our approach to producing a new and more accurate map of site index for the whole forest. The emphasis of the forest owner was to produce a workable, best estimate, site index map at a low cost.

Several approaches to estimation are possible. Some simple options appealed to the forest owners and were considered, as described below.

A straightforward technique for estimating missing values is to divide the forest into large but moderately uniform blocks and to use the block mean to approximate all missing site indices within each block. The map output from such an approach would show the forest divided into a few large blocks, each covering several compartments. The associated error of estimation would be the variance of the population. This method, however, does not take local trends and variations into account. Within-block variation can be large. For example, site indices in a Kaingaroa block can vary by 6 to 8 m over distances of a few kilometres.

A Geographic Information System (GIS) can be used to produce a map of the compartments shaded in colour to match the associated site index values. The site indices were known to follow a gradual trend across the forest from north to south and this technique shows such a trend well. However, this is essentially a mapping exercise. Missing values estimated from this map will be approximations that are subjective and non-reproducible, and unreliable where the coverage of known values is irregular.

Another technique suggested was to use a GIS or other software to contour the known site indices and to estimate unknown site indices from surrounding contours. The estimation process was to be based on a combination of the values of the surrounding contours proportional to their distance from a missing site index. Although the technique has the potential to be biased and has no estimate of error (Webster & Oliver 1990), the approach was intuitively reasonable to the forest owners and so was considered.

There is a large body of literature on interpolation (for example, Oliver & Webster 1990; Lam 1983). Techniques range from global approaches such as trend surface analysis, to local methods such as low order polynomial, and weighted moving averages. None of these methods, however, determines whether its assumption of spatial dependence in the data holds, nor do they provide any estimates of the errors of estimation (Oliver & Webster 1990). They are based on modelling a continuous spatial variable by mathematical functions. The desire to use a more realistic model to describe the spatial relations that lie behind a map such as the site index map led to the appraisal of geostatistics.

Background to Geostatistics

Geostatistics was developed initially within the mining industry by Matheron (Matheron 1965) and Krige (1966) among others, and in recent years the technique has been applied in other disciplines, notably soil science (Webster & Oliver 1990) and in one recent instance in forestry (Samra *et al.* 1989). The concept underlying geostatistics is fairly simple and is described below. For more details, *see* Oliver & Webster (1990) and Webster & Oliver (1990).

Most natural properties vary continuously. Although from place to place the variation may be quite erratic, values close together in space are more likely to be alike than those further apart, i.e., they depend on one another in a statistical sense. Geostatistics exploits this autocorrelation and uses it in the estimation by assigning greater weights to data closest to the point being estimated.

The procedure involves two steps, the first being to describe the autocorrelation of the data. This is done using regionalised variable theory which uses a stochastic approach to account for the spatial dependence of natural properties. For simple applications this theory assumes a constant local mean and a stationary variance of the differences between places separated by a given distance and direction. This variance of the differences, usually denoted γ, the semi-variance, is half the expected squared difference between two values. Formally,

$$\mathrm{var}[z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})] = \mathrm{E}[\{z(\mathbf{x}) - z(\mathbf{x}+\mathbf{h})\}^2] = 2\gamma(\mathbf{h}) \tag{1}$$

where $z(\mathbf{x})$ is the value of a property z at position $\mathbf{x}$, a vector, and $z(\mathbf{x}+\mathbf{h})$ is the value at $(\mathbf{x}+\mathbf{h})$. The semi-variance depends on the separation $\mathbf{h}$, the lag, in both distance and direction, not on the actual positions of the data.

The function which relates γ to $\mathbf{h}$ is the variogram. The equation for calculating it from sample data is:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2M(\mathbf{h})} \sum_{i=1}^{M(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2 \tag{2}$$

where $M(\mathbf{h})$ is the total number of pairs of points separated by the lag $\mathbf{h}$. By changing $\mathbf{h}$ an ordered set of values is obtained, which is termed the experimental variogram, and the shape of the plot of this variogram describes the degree of autocorrelation of the variable. In some instances γ may depend not only on the separation distance, but also on the direction of the pairs of points.

A mathematical model is fitted to the experimental variogram and from it the model parameters for the variogram are estimated. Various functions are acceptable, e.g., spherical, linear, bounded linear, exponential, and Gaussian (McBratney & Webster 1986). They fall into two broad groups: the unbounded and bounded models. Normally, when fitting the function a non-linear regression approach with weights in proportion to the number of paired comparisons contributing to γ is recommended (Oliver & Webster 1991). The variogram parameters derived for bounded models are:

- the upper bound known as the *sill*
- the intercept on the ordinate (C_0)—an intercept greater than zero, known as the *nugget variance*, signifies that variation is present at a lag distance less than the minimum lag measured
- the range of the autocorrelation (a), i.e., the lag distance above which the samples are independent of each other
- the slope of the function.

The parameters of the model fitted are then used in the second stage, the estimation itself (kriging). Kriging estimates by local weighted averaging:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{n} \gamma_i \, z(\mathbf{x}_i) \tag{3}$$

where $\hat{z}(\mathbf{x}_0)$ is the estimate at $\mathbf{x}_0$, and γ_i are weights assigned to the surrounding known data based on their distance and possibly direction from $\mathbf{x}_0$. In order for the estimate to be unbiased the weights must sum to 1 and, subject to this, they are chosen to minimise the estimation variance. If the data are autocorrelated, the weights for known values close to $\mathbf{x}_0$ will be larger than those further away. In addition, clustered data will have correspondingly smaller individual weights and samples shielded from $\mathbf{x}_0$ by closer data will also have smaller weights (Oliver & Webster 1990). If the data are not autocorrelated, the weights of all data used in the estimation will be the same irrespective of distance from $\mathbf{x}_0$. As the weights must sum to 1, the weights will be equal to 1/n where n is the number of data involved in the estimation, and so in this case:

$$\hat{z}(\mathbf{x}_0) = \frac{1}{n} z(\mathbf{x}_1) + \frac{1}{n} z(\mathbf{x}_2) \ldots + \frac{1}{n} z(\mathbf{x}_n) \tag{4}$$

One of the advantages of kriging over traditional methods of interpolation is that the estimates are unbiased and are of minimum and known variance. Since the estimation variances can be calculated and mapped, the confidence that can be placed in the estimates can be determined.

METHODS

Site Index

To compute a site index, the mean top height of selected trees on a site and the age of those trees need to be known. Height models have been developed for several areas in New Zealand as part of the *Pinus radiata* D.Don growth modelling research programme. The Pumice Plateau growth model was used for Kaingaroa Forest (A.G.Dunningham & M.E.Lawrence unpubl. data). The model uses top height at a given age to predict the height at age 20, the site index. Four sources of estimates of top height at known ages were used for Kaingaroa Forest.

- Permanent sample plot (PSP) system (J.D.Dunlop unpubl. data).
  This database contains nearly 16 000 measurements from many *P. radiata* plots, of which about 1300 had height/age pairs between ages 10 and 50 years. For each measurement pair the site index was calculated. All values of site index for a plot were averaged.
- Pre-harvest and mid-rotation inventory plots.
  Height/age pairs were available for about 5000 temporary plots measured between the ages of 20 and 50 years since 1984. Site index was calculated for each measurement pair.
- Age 13 inventory plots.
  Height/age pairs from 358 stands (5600 temporary plots) were measured during 1974 to 1978, and were used to estimate site index for each stand (a number of stands make up a compartment).
- Extraction thinning quality control measurements (M.W.Deadman & C.Pilaar unpubl data).
  Height/age pairs were available from thinning quality control measurements in 391 stands between the ages of 13 and 22 years.

Although there were about 11 000 site index estimates available, there were 899 compartments out of the total of 1643 without any associated site index. The limited availability of digital spatial data, as discussed below, restricted the estimation to the compartment level rather than the more accurate individual plot level. For those compartments with site index values, compartment means were calculated. This was done by first averaging multiple measurements within one of the above sources—for example, by averaging the site indices of several age 13 inventory plots in a compartment. Then, if there was more than one source of site index for a compartment, these indices were averaged too (Shirley unpubl. data).

GIS Methods

To use a GIS, the spatial data need to be digitised and the attributes appended. For this study, the 1643 forest compartment boundaries were available in digital format, with each compartment uniquely identified by forest code and compartment number. This was the most detailed level of digital spatial information available. The data were imported into the GIS (ARC/INFO). The compartment identifiers were the links to other data—in this exercise the identifiers were used to store the 744 known compartment site indices into the GIS.

The GIS was used to produce colour shading plots. Site indices were grouped into intervals of 2 m and coloured from red (low value), through orange, yellow, etc., to dark blue (high value). The resulting map showed some potential invalid site indices—for example, a value seemingly very different from others in the same area. With no more information to go on than the feeling that a value was not correct, it was decided not to drop any of the known values until they could be investigated further.

There were a number of influences on the data other than geographical ones. These included:

- The history of site preparation and silviculture;

- The number of rotations on a compartment—parts of the forest are under second or even third rotation, with the more recent plantings being genetically improved trees;
- The values themselves—they were from many sources and had been collected over many years.

The first two factors created valid outliers while the last one increased the likelihood of errors. However, given that it was sufficient for this project to produce a workable, best estimate, site index map at a low cost, it was considered acceptable to use the data as they were.

The GIS was also used to contour the site indices. ARC/INFO uses a TIN (triangulated irregular network, ESRI 1989) algorithm for contouring. No estimation of unknown values was made from the contour maps as it would have been prone to errors and impractical for the amount of estimation required.

Finally, the GIS was used as a link between the site index data and the geostatistics method. Site indices were stored by compartment, while geostatistics requires the co-ordinates of the compartment centres and the associated values of site index. The GIS was used to output a data set of site indices with the centre point co-ordinates of their compartments, as well as the co-ordinates of the centre points of the 899 compartments for which site index estimates were required. The GIS was also used to find the average compartment area for the forest. With these data the geostatistics could be run. The results of the geostatistics method, i.e., the estimated site indices and their co-ordinates, were entered into the GIS again. The GIS was used with these co-ordinates to assign values of site index to their appropriate compartments.

Geostatistics Methods

The known site indices, as exported from the GIS, were initially analysed for normality using the SAS statistics package (SAS Institute 1985). The summarised data were:

Mean	29.5
Standard Deviation	3.31
Skew	–0.54
Kurtosis	0.48

Analysis of the measured site indices showed them to be normally distributed about the mean, and therefore no transformation was required prior to construction of the variograms.

Variograms were constructed using the program "Semivar" (Robertson 1987), with lag distances of 1 km intervals. This was followed by an assessment of anisotropy (i.e., directional rather than distance-related variation). For this assessment, variograms were computed and plotted in four directions (0°, 45°, 90°, and 135° with a tolerance of 10° either side of those directions).

Models were fitted to the isotropic variogram data, using a non-linear weighted least squares approach with the SAS package (SAS Institute 1985). Weights were assigned according to the number of points contributing to the estimate of variance. Only acceptable model types were evaluated (spherical, exponential, bounded linear, unbounded linear, and Gaussian—McBratney & Webster 1986).

Site indices for compartments with no measurements were estimated using the program "Punctual" (Robertson 1987), with the variogram given above. Site indices were estimated at the compartment centre co-ordinates calculated within the GIS. Estimation variances were also calculated and converted to standard errors (√estimation variance), and are presented as such.

The kriging was validated using a jack-knifing technique, which re-estimated the known site indices. This was done for all 744 data points, by dropping a single known value and using the remaining data to estimate the dropped value, until all site indices had been re-estimated. Population statistics for the known and estimated data were then compared to determine the efficiency of kriging.

The resulting sets of co-ordinates, actual and estimated site index, and the standard errors of the estimates for each compartment were transferred to the GIS as ASCII files.

RESULTS AND DISCUSSION

GIS Analysis

A sample contour map with compartment boundaries and known site index values is shown in Fig. 2. It gives an overall picture of variation in site index, but has shortcomings. The site index values are compartment means, not point values as assumed by the contouring algorithm. Most of the computed contour lines run through parts of compartments, making the actual site index value unclear. There are some practical difficulties with estimating site

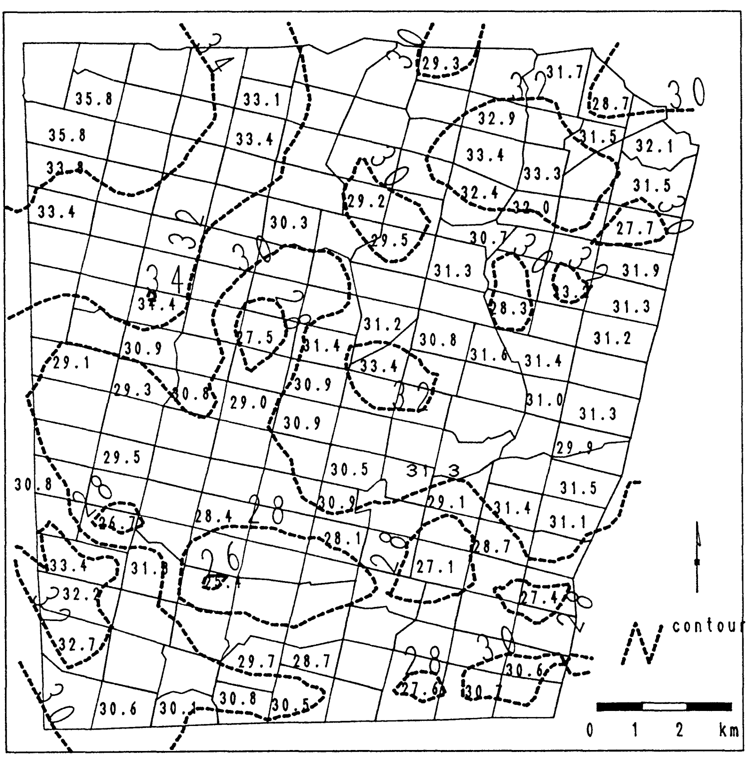

FIG. 2–Example of a contour map of site index (m) for a block of compartments in Kaingaroa Forest. Compartment boundaries are shown, plus measured values of site index where available.

index from contours. Assuming a straight line gradient between adjacent contour lines would be only an approximation. If surface trends are considered, estimating missing values becomes complex. In addition, if two adjacent contour lines have the same value it can be difficult to estimate the height or depth of the "hill" or "valley" between them. Currently this estimation technique would be manual, with any distances required having to be read off the map.

A map produced by graduated shading of compartments based on their site index value (e.g., Fig. 3) shows the full area of the forest covered by each value, removing the uncertainty caused by contours bisecting compartments. Although this type of map gives less insight into values for unmeasured compartments than does the contour plot, it is useful for clearly highlighting compartments which have values of site index quite different to those surrounding them. If a single compartment stands out, the value may be incorrect and should be investigated. If a group of compartments is highlighted, site factors or the history of the area may indicate why this is so. For example, a group of compartments in northern Kaingaroa Forest had unexpectedly small site indices. Investigation of stand records suggested that the stands may have been subjected to an unusually intense pre-establishment burn (the previous crop was *P. ponderosa* P. et C.Lawson which is known to have high slash). This may have contributed to the lower-than-expected site indices in the subsequent crop.

It seems that, for our purposes, graduated-shading compartment maps based on the results of an estimation technique would be the preferred option for both visual assessment of site index variation and production planning.

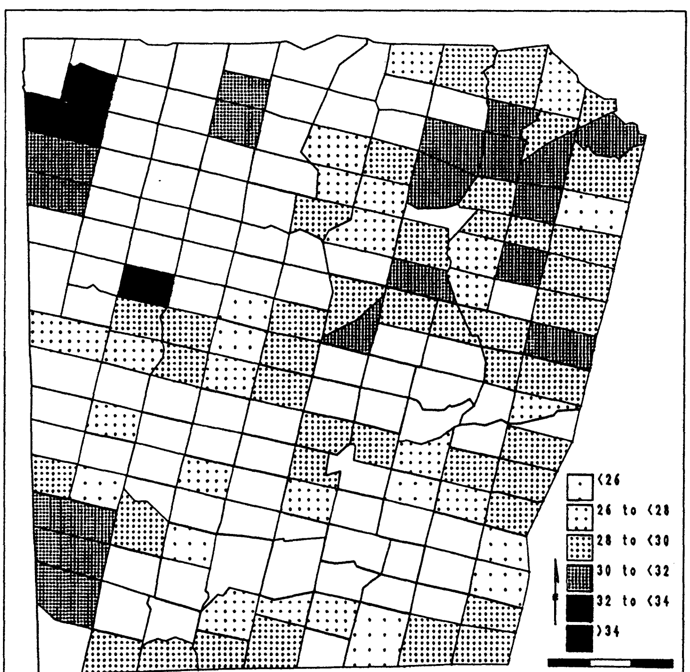

FIG. 3–Example of graduated shading map of site index (m) for a block of compartments in Kaingaroa Forest.

Geostatistical Analysis

Variogram analysis

The variogram was constructed up to a maximum lag distance of 25 km, taking into consideration the approximate width of the forest (Fig. 4).

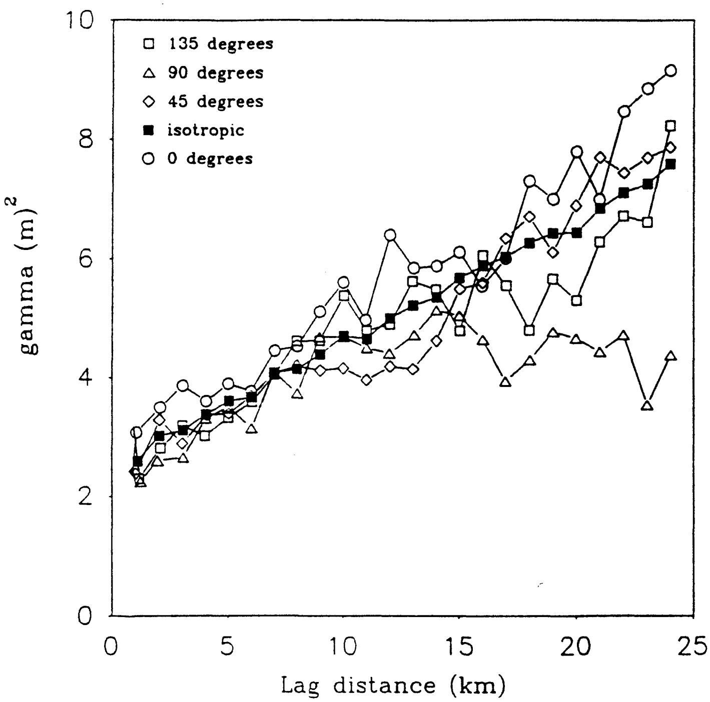

FIG. 4–Isotropic variogram plus variograms constructed in four directions to evaluate any directional variation in spatial dependence of site index.

The variogram was assessed for anisotropy. The plotted variograms for the four directions 0°, 45°, 90°, and 135° are shown in Fig. 4 with the isotropic variogram. As the variograms were much the same in all directions, it was considered appropriate to use the isotropic variogram for the kriging to capture the benefits of unbiased estimates with minimum error.

A linear model was fitted to the variogram. The model is presented in Equation 5:

$$\hat{\gamma}(\mathbf{h}) = 2.482 + 0.21\ \mathbf{h} \tag{5}$$

The positive intercept signifies the degree of variation in site index present at a smaller scale than the 1 km minimum lag distance, and in this case would represent the within-compartment variation in site index.

Kriging results

Estimates were produced for missing site indices where the surrounding data were dense enough to ensure 16 known values within a radius of 25 km. The standard errors associated with the estimates were also calculated. The estimated site indices ranged from 18.1 to 34.2 m, with a mean of 28.1 m. The standard errors ranged from 1.6 to 3.6 m, with a mean of 1.68 m. The majority of these standard errors were half the value of 3.3 m that would apply if data independence was assumed and the error of the estimate was the standard error of the total population. The standard errors increased with distance from known site indices. For example, compare points A and B in Fig. 5 with point C. The values of the standard error are smallest in A and B (1.62 m), where the compartment is bordered by compartments with known site index, while point C has a standard error of 1.70 m and surrounding known values are further away. Over most of the forest where values were estimated the standard error was less than the mean of 1.68 m. This was due to the fairly even spread of data points through the forest. Higher estimation errors occur in outlying areas of the forest, and where data are sparse.

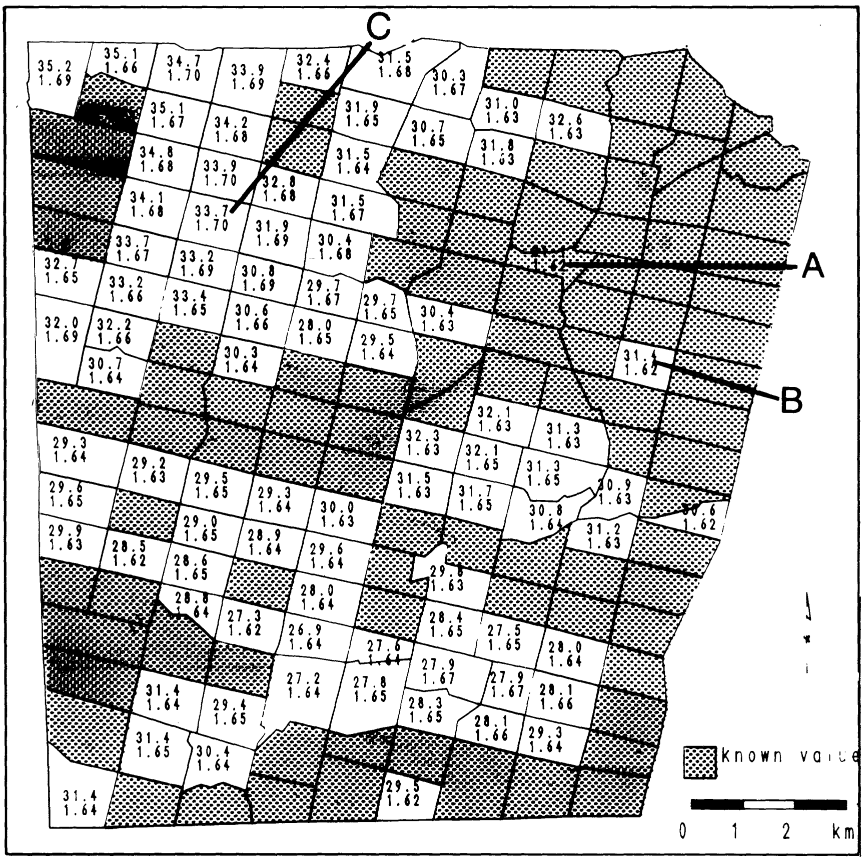

FIG. 5–Map of estimates of site index of those compartments without measured values for a block of compartments in Kaingaroa Forest. Errors associated with the estimates are also shown. Compartments with known values are shaded.

Evaluation of the Kriging Procedure

The results of the jack-knife analysis, i.e., the known *versus* estimated site indices, were plotted in Fig. 6. The correlation between actual values and estimates yielded an r^2 of 0.6249, p=0.0001. The average difference between known and estimated site index was 0.005 m

(Table 1). The largest difference was 9.4 m, but the vast majority of the differences were much smaller. The standard error associated with the estimates was 1.9 m. Kriging smooths surfaces and, as would be expected, the technique over-estimated for compartments with a lower site index and under-estimated for compartments with higher site index values. However, the estimation procedure is unbiased overall and, again as expected, the mean difference was very close to zero. Several compartments whose known values appeared to be markedly different to values recorded in surrounding compartments will be field checked and their site index values corrected if necessary. After this, the map will be updated.

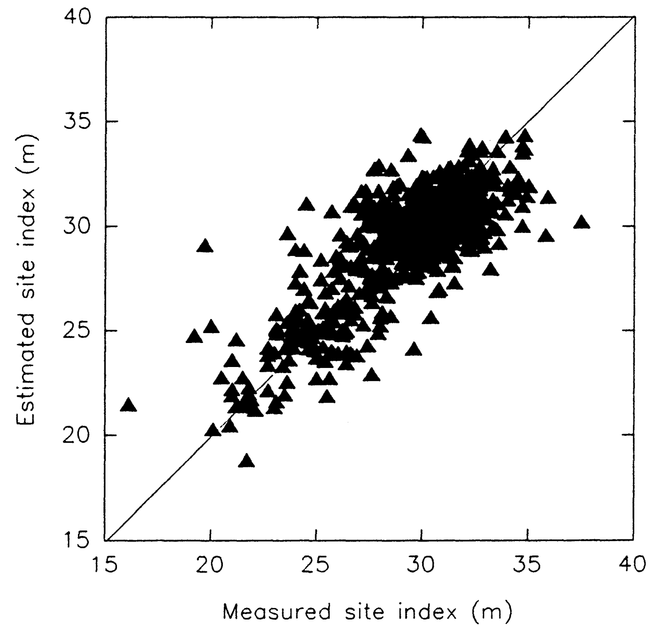

FIG. 6–Comparison of known site index with estimates made using geostatistics (jack-knifing procedure).

TABLE 1–Comparison of the measured site index and values estimated using geostatistics for the same compartments (jack-knifing).

	Measured site index (m)	Estimated site index (m)	Difference (m)
Mean	29.5	29.2	–0.005
Std error	3.3	2.76	1.91
Minimum	16.1	18.2	–7.27
Maximum	37.7	34.4	9.40

As expected, standard deviation increased with distance from known values. This was clearly illustrated (Fig. 7) by the group of compartments with estimated site index values of about 20 m but with increasing values of standard error associated with the estimates. The compartments involved comprise a long thin "spit" of forest extending from the main body

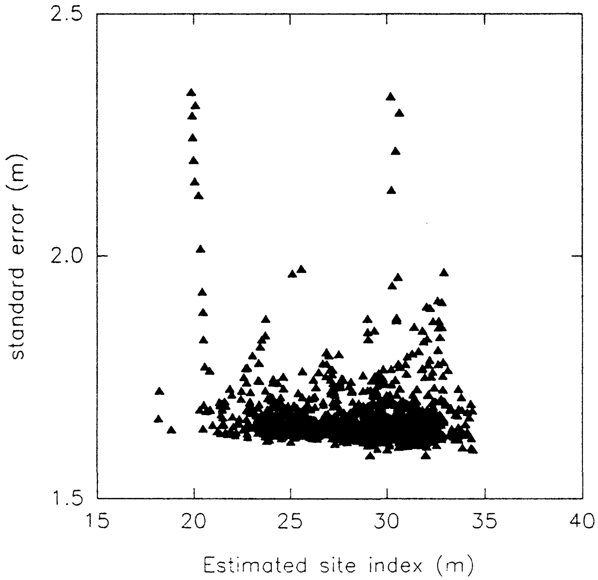

FIG. 7–Estimated values of site index produced by geostatistical analysis *versus* the standard errors associated with the estimates.

of the forest. The standard error increased with increasing distance from compartments of known site index in the main body of the forest. There were no measured values of site index on the spit which would have helped reduce the error.

The final stage of the exercise was to produce a full map of the forest, utilising compartment boundaries and both measured and estimated values for site index. This output from ARC/INFO is shown in Fig. 8. A colour spectrum from red (lowest site index), through orange, yellow, etc., to dark blue (highest site index) was used, with intervals of 2 m between categories. This output allows rapid visual assessments of growth patterns within the forest, and interrogation of the GIS database gives quick access to data of compartments of interest. The presentation of the map by either colour or gradational shading of compartments was preferred to contours of site index for three reasons. The colouring or shading of compartments was easier to see, contours bisected compartments in the map and this was confusing as data were based on compartment means, and thirdly, as there were values associated with all compartments on the GIS database, estimation of site index values from contours was no longer necessary.

CONCLUSIONS

Combining a GIS with statistical techniques allowed us to update the map of site index distribution over Kaingaroa Forest effectively. Geostatistical techniques were successfully used to produce estimates of site index for compartments without measured values, and these estimates were unbiased with minimum variance. Producing a colour-shaded map of site index for the forest, in conjunction with readily accessible data on the GIS database was an

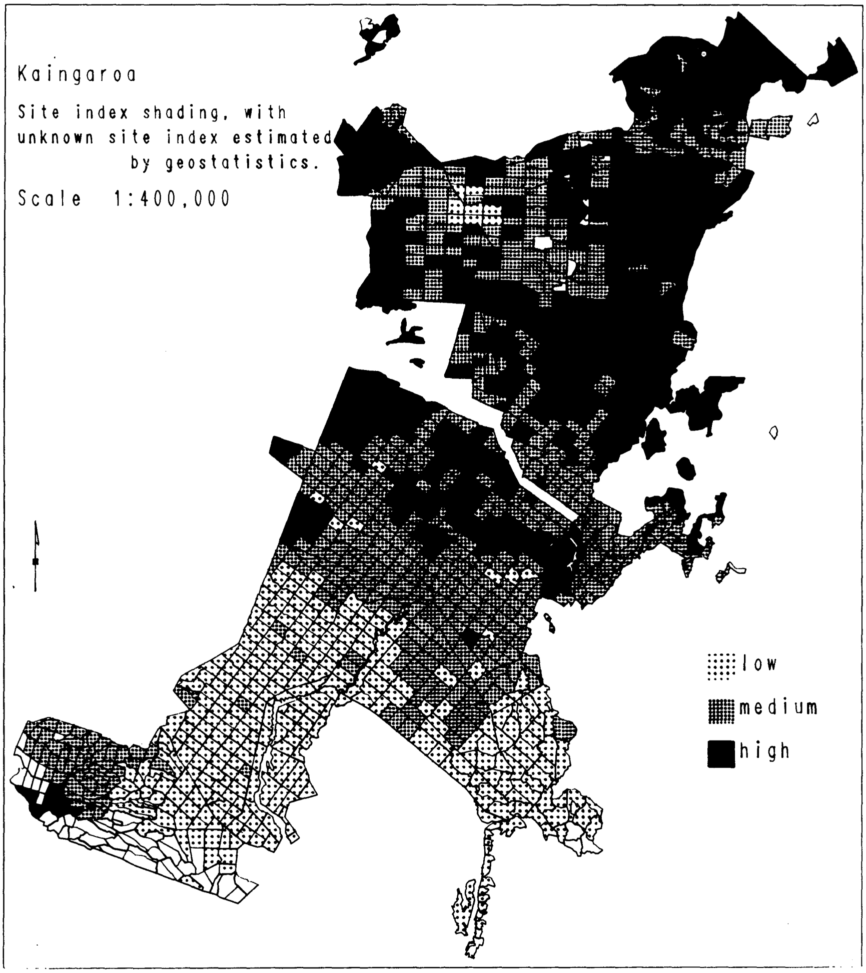

FIG. 8–Map of Kaingaroa Forest produced using both measured and estimated values.

effective way of presenting growth data for the forest. Additionally, digital storage of information will allow rapid—even automated—updating of the map in future as new data become available.

REFERENCES

ESRI 1989: "TIN Users Guide". Environmental Systems Research Institute Inc., Redlands, California.

KRIGE, D.G. 1966: Two-dimensional weighted moving average trend surfaces for ore evaluation. *Transactions of the South African Institute of Mining and Metallurgy 66*: 13–38.

LAM, N.S. 1983: Spatial interpolation methods: A review. *The American Cartographer 10(2)*: 129–49.

MATHERON, G. 1965: "Les Variables Régionalisées et Leur Estimation". Masson, Paris.

McBRATNEY, A.B.; WEBSTER, R. 1986: Choosing functions for semi-variograms and fitting them to sampling estimates. *Journal of Soil Science 37*: 617–39.

OLIVER, M.A.; WEBSTER, R. 1990: Kriging: A method of interpolation for geographical information systems. *International Journal of Geographical Information Systems 4(3)*: 313–32.

———1991: How geostatistics can help you. *Soil Use and Management 7(4)*: 206–17.

ROBERTSON, G.P. 1987: Geostatistics in ecology: Interpolating with known variance. *Ecology 68(3)*: 744–8.

SAMRA, J.S.; GILL, H.S.; BHATIA, V.K. 1989: Spatial stochastic modelling of growth and forest resource evaluation. *Forest Science 35(3)*: 663–76.

SAS INSTITUTE 1985: "SAS Users Guide: Statistics". Version 5 Edition. SAS Institute Inc., Cary, North Carolina.

WEBSTER, R.; OLIVER, M.A. 1990: "Statistical Methods in Soil and Land Resource Survey". Oxford University Press, Oxford.