P. F. Crapper
Division of Water and Land Resources
CSIRO
Canberra, A.C.T. 2601, Australia

An Estimate of the Number of Boundary Cells in a Mapped Landscape Coded to Grid Cells

The number of boundary cells has been determined given the size of the overlay grid, the shape factor of the regionalization, and the mean and variance of the areal distribution of regions.

Introduction

For many applications a grid-based data structure is preferable to a hierarchical polygon based data structure, because it is simpler to implement, update, and use. For the case of computerized geographic information systems containing biophysical attributes, grid-based data structures have become widespread (see, for example, Antenucci (1979), Cross (1979), Schlesinger *et al.* (1979), Svatos (1979), or Wilson (1979)). The major advantage of grid-base data structures is the ease of finding data items for a particular location because the location is keyed directly to the position in the array. The accuracy of representing a given pattern can be improved to any required level by decreasing the size of the grid cell. Alternatively, it may be desirable to use two or more sizes of grid cells if the vegetation and/or topographic complexity varies significantly. In a grid-based system the boundaries between adjacent regions appear stepped. The boundaries for a polygonal data base appear smooth and well behaved, which can lead to erroneous conclusions about their spatial accuracy.

The grid-based data structure can be referenced either to an image based information system (sometimes called image raster) or to a pre-existing map grid system. Halsey and Smith (1978) have referred to the former approach as using an implicit referencing system and the later approach as using an explicit referencing system. In the implicit approach an automated raster scanning device scans an area of land, divides the land into regular grid cells, and records one or more spectral signals for each grid cell. The segment of land could be the Earth's surface, in which case the scanner might be mounted on an aeroplane or satellite (the Landsat situation), or it could be a representation of the Earth's surface (e.g., a map or photograph), in which case the scanner might be a microdensitometer type instrument. The image is then generally rectified, registered, and resampled so that each grid cell is a convenient size (e.g., 1 ha, 25 ha, etc). Additional or collateral data are then sampled to this grid cell size and overlaid on the existing resampled image (see, for example, Bryant and Zobrist (1978), Zobrist (1982), or Smith and Blackwell (1980)). The basic premise is that geocoded data sets can be referenced

Abstract: *One of the central functions of geographic information systems is to display, manipulate, or analyze thematic data. Often such systems are used to interrelate spatial thematic data derived from varying sources. The interrelationships between a grid cell data base and a polygon based data base are examined in this paper. Uncertainty in assigning a theme to a grid cell can occur when the grid cell lies near or on the boundary of a region. Accordingly the thematic map accuracy is related to the number of boundary cells. Unfortunately, there are many more boundary cells than one would intuitively expect. A formula has been developed for determining the number of boundary cells and the standardized error of area estimates for a regionalized map. One example of a regionalized map is a Landsat classified image.*

Photogrammetric Engineering and Remote Sensing,
Vol. 50, No. 10, October 1984, pp. 1497-1503.

0099-1112/84/5010-1497$02.25/0

to a raster scan that is equivalent to an ultra-fine mesh grid cell data set and that images taken of the Earth's surface or representations of the Earth's surface can be converted to the same raster scan. In the Landsat case the collateral data set could include elevation, slope, aspect, etc., and then specific questions could be answered by analyzing or classifying some or all of the channels in the multivariate data set.

In the explicit approach a grid is overlaid on an existing map or photograph and one or more data items are read for each grid cell. The grid is generally chosen to coincide with a pre-existing grid system. Within Australia the Australian Map Grid (AMG) is overlaid on most maps and subdivisions; AMG cells (1 km by 1 km) or aggregations provide the most convenient definition of a grid cell (see, for example, Crapper and Spate (1982)). One potential problem which can be encountered when using AMG (or equivalent zone referencing systems) is the presence of a zone boundary. In other countries, latitude/longitude subdivisions may be more appropriate. Laut and Paine (1982) have shown that 4 km^2 grid cells can be used in conjunction with numerical taxonomic procedures to provide a classification of a regional landscape. In general, the explicit approach is used with significantly larger grid cells than the implicit. However, Maggio *et al.* (1983) in a study of the distribution of pecan nuts in southeast Texas overlaid 1 ha grid cells on 1:10 000 color infrared positive transparencies. They referenced the grid cells to a UTM grid in spite of latitude/longitude subdivisions being shown on the 1:24 000-scale quadrangle sheets.

One of the major applications of geographic information systems is to present thematic data. In both the implicit and explicit approach, new thematic data are introduced into the geographic information system by overlaying a grid cell structure on the theme to be acquired. For the 'pure' grid cells there should be little difficulty in determining which theme to associate them with; however, for the boundary cells (often called mixels or mixed grid cells or 'impure' cells) there can be considerably difficulty in determining with which theme to associate them. The fundamental problem is that the grid cells must be included or excluded in their entirety as there is no spatial information at the sub grid cell level.

The classification process, in which cells are allocated to classes, can react to boundary cells in three ways. First, treat the boundary cells separately and allocate them in an unbiased fashion. Thus, approximately 50 percent of the boundary cells will be allocated to each adjacent region and the areas of the adjacent regions will be true estimates. Second, treat the boundary cells in a biased fashion so that substantially more than 50 percent will be allocated to one region and substantially less than 50 percent will be allocated to the adjacent region. In this case the area estimates will not be true estimates but biased estimates, with one region area overestimated and the other region area underestimated. Third, treat the boundary cells as a separate class and do not associate them with either of the adjacent regions. In this case the area of both adjacent regions will be underestimated. Accordingly, the accuracy of region area estimates is related to the number and treatment of boundary cells.

As to which of the above three possibilities occurs and whether it occurs globally depends on the classification process used and the image itself, and may vary within the image. Hence, it is not possible to make any general theoretical statements about classification accuracy. However, it is possible to derive theoretical relationships giving the number of boundary cells, and this number can be used as a measure of the classification accuracy. Unfortunately, there are many more boundary cells than one would intuitively expect. Jupp *et al.* (1979) found for a mapping exercise in the Batemans Bay area of New South Wales that, for the case where the ratio of the average region area to the grid cell area was 16, 54.9 percent of the total cells were boundary cells.

In this paper a relationship is derived giving the number of boundary cells in a mapped landscape. Crapper (1980) has examined the case where the boundary cells were treated in an unbiased fashion and found a relationship between the number of boundary cells and the region area for a single bounded region. In the present work a relationship is assumed for the size distribution of region areas and a formula is developed for the total number of boundary cells in a regionalized map and the standardized error of the area estimates. This result can be used for answering two separate questions. First, given a regionalization with known (or assumed) region area mean and variance, what is the maximum grid cell size for a given percentage of boundary cells? Second (which is more common), given a grid cell size, what is the average region area for a given percentage of boundary cells? One example of a regionalized map is a Landsat classified image.

Determination of Boundary Cells for Single Region

For any set of similar (that is, same shape but different size) polygons, there is a unique relationship between the perimeter length (or any other representative length scale) and the area. This relationship has been called a shape factor, and for different applications different shape factors have been proposed (Haggett and Chorley, 1969). Shape factors are generally defined to be non-dimensional so that for a family of similar polygons the shape factor will be constant. Most shape factors involve manipulation of the basic parameters of area and perimeter or comparisons with other regular shapes. The shape factor which appears to be most extensively used in a geographical context is that

defined by Goodchild and Moy (1976) or Frolov and Maling (1969). The shape factor k_{1j} is defined to be

$$k_{1j} = \frac{L_j}{2\sqrt{\pi A_j}} \quad (1)$$

where L_j is the perimeter length and A_j the area of region j.

The derivation of the number of boundary cells or mixels N_j^B in region j is given in Crapper (1980), and for convenience is briefly outlined below.

If a grid with regular cells in overlaid on the region, then

$$L_j = \sum_{i=1}^{i=N_j^B} \ell^{ij}$$

where ℓ_{ij} is the length of the boundary in the i^{th} cell for region j. Thus, the average length of the perimeter per boundary cell is

$$\bar{\ell}_j = \frac{L_j}{N_j^B} = \frac{1}{N_j^B} \sum_{i=1}^{i=N_j^B} k_{2ij}\,\bar{L} = \bar{k}_{2j}\,\bar{L}$$

where $\bar{L}$ is the average length of a random straight line laid across a grid cell and k_{2ij} $(= \ell_{ij}/\bar{L})$ is defined to be within-cell contortion parameter. In general, k_{2ij} will be different for each boundary cell. If the section of the boundary within cell i were equal in length to the random straight line then $k_{2ij} = 1$, however, k_{2ij} could vary from slightly greater than zero to greater than one. Thus, N_j^B is given by

$$N_j^B = \frac{L_j}{\bar{k}_{2j}\,\bar{L}} = \frac{2k_{1j}\sqrt{\pi A_j}}{\bar{k}_{2j}\,\bar{L}} \quad (2)$$

where

$$\bar{k}_{2j} = \frac{1}{N_j^B} \sum_{i=1}^{i=N_j^B} k_{2ij}.$$

Hence, for region j the number of boundary cells has been related to a shape factor k_{1j}, a within pixel contortion parameter $\bar{k}_{2j}$, and the region area A_j.

In general, the evaluation of k_{1j} and $\bar{k}_{2j}$ is difficult as they depend on boundary definition. The shape of the boundary of most biophysical features cannot be defined exactly because in most cases there is not a sharp border between two adjacent land attributes. The border region generally consists of a gradual transition from one attribute to the other. Thus, as the shape cannot be defined exactly, it is not surprising that the length of the boundary cannot be defined uniquely.

The measured length of a geographic line will vary according to the fidelity of the map and the accuracy of the measurement, and the more accurately the curve is measured, the greater will be the length. Mandelbrot (1977) has called the class of curves which do not approach a limit a fractal. Fractals have the characteristic that their irregularly extends with no breaks and no relative attenuation down to infinitesimal scales. Galloway and Bahr (1979) discuss in detail the concept of 'self-similarity' where each portion of a curve can be considered a reduced-scale image of the whole, and whether an upper limit exists for the Australian coastline, and give several references. They conclude that before estimating the length it is necessary to consider the reason for measurement. For different land uses there will be different lengths, and the available total length will vary with the use envisaged. Thus, for a given region it is not possible to give unique values for k_{1j} or $\bar{k}_{2j}$ as they both vary with scale. However, if the scale of the mapping is appropriate for the exercise, then the k_{1j} and $\bar{k}_{2j}$ values determined will also be appropriate.

In order to evaluate k_{1j}, it is necessary to have spatial information about the region boundary. If the boundary were stored in analog form (e.g., a map), then the perimeter could be measured with an opisometer and the area with a planimeter, thus allowing an evaluation of k_{1j}. In general, this evaluation for more than a small number of regions would be very tedious and quite error prone unless the regions were of a regular shape. If the boundary information were stored in digital form in a computer compatible format, then k_{1j} could be evaluated much more easily. Crapper (1981a) determined the shape factor for 1605 biophysical regions defined by a polygonal data base on the south coast of New South Wales (subsequently referred to as the South Coast Data Base). The regions were defined on the basis of relative homogeneity in the spatial pattern of the biophysical properties at a scale of 1:25 000. It is this selection of mapping scale which determines boundary definition and hence k_{1j}. This regionalization was produced by an experienced team of professional scientists (see Austin and Cocks, 1978) and, although in detail it may differ from another biophysical regionalization of the same area at the same scale, its geometric properties would most probably be representative. The regions covered a total area of approximately 6000 km^2 (average area 3.8 km^2) and included significant variations in landform, vegetation, and soils. The areal distribution of the regions is shown in Figure 1 and is discussed in greater detail later. Crapper (1981a) found that the shape factor varied slightly with landform and area. Ignoring the variation with region area and averaging for different landforms, $\bar{k}_{1j} = K_1 = 1.82$. Galloway and Bahr (1979) have observed a similar relationships between L and $\sqrt{A}$ for islands adjacent to the Australian mainland coastline.

The evaluation of $\bar{k}_{2j}$ is more difficult as it requires the spatial information about the region boundary to be overlaid with a regular array of grid cells. Again, k_{2ij} could be evaluated, if the boundary were in analog form, by measuring the boundary within each boundary cell and dividing it by the average

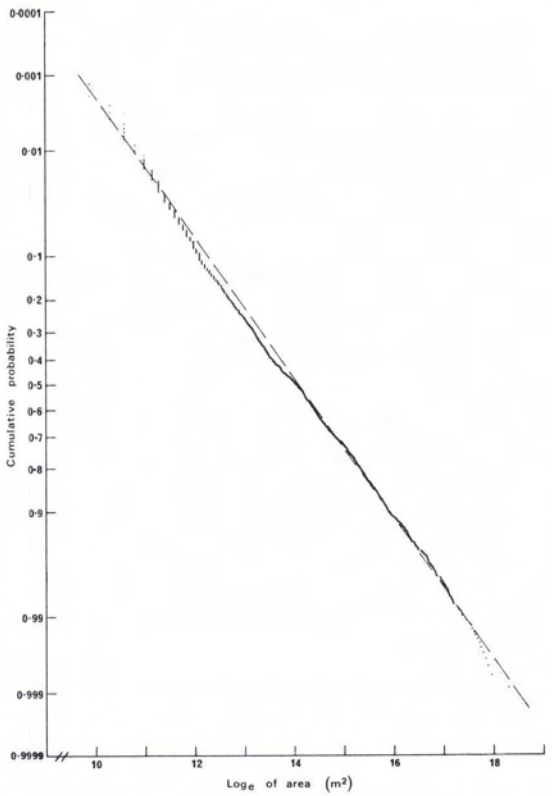

FIG. 1. The cumulative probability is plotted against the region area (m^2) using probability and natural logarithmic axes, respectively. The straight line was fitted by eye. The regions were defined on the basis of relative homogeneity in the spatial pattern of the biophysical properties.

length of a random straight line laid across the cell. This calculation for even one region would be tedious and error prone. If the boundary data were in digital format, then there would still be considerable problems as the polygons would have to be overlaid with the regular grid cells, points of intersection determined, and the length of boundary within each cell calculated. Goodchild and Moy (1976) discuss further the problems in evaluating $\bar{k}_{2j}$.

The value of $\bar{k}_{2j}$ for one of the worst cases one could reasonably expect was evaluated by Crapper (1981b). The region boundary was defined by the coastline of a small (36 ha) highly indented island. The overlay grid used was a Landsat pixel grid. It was found that $\bar{k}_{2j} \cong 1.3$. The value may be considered to be an upper bound as the region-to-pixel ratio was comparatively small (80) and the boundary was highly contorted. The value of $\bar{k}_{2j}$ for more representative regions and region-to-pixel ratios would be much smaller, and it seems reasonable to assume that $\bar{k}_{2j} = K_2 = 1$. This is equivalent to assuming that the section of boundary within each boundary cell can be approximated by a straight line.

Thus, with these assumptions Equation 2 becomes

$$N_j^B = \frac{2K_1\sqrt{\pi A_j}}{\bar{L}} \tag{3}$$

Determination of Boundary Pixels for an Ensemble of Regions

For an ensemble of regions the number of boundary cells N^B will be given by

$$N^B = {}^1/_2 \sum_{j=1}^{j=M} N_j^B = \sum_{j=1}^{j=M} \frac{K_1\sqrt{\pi A_j}}{\bar{L}}$$

as each boundary cell will occur in two adjacent regions, ignoring outer boundary effects. If the number of regions is comparatively large, as in the South Coast Data Base, then this assumption is justified. (For situations where the number of regions is small, an underestimation of the number of boundary cells will occur. If the fraction of boundary segments in the outer boundary is known, a correction could be applied. Thus,

$$N^B = \frac{K_1}{\bar{L}}\sqrt{\pi}\, M\overline{\sqrt{A}} \tag{4}$$

where

$$\overline{\sqrt{A}} = \frac{1}{M}\Sigma\sqrt{A_j}\,.$$

The total number of cells N^T is given by

$$N^T = \frac{M\bar{A}}{A_C}$$

where $\bar{A} = (1/M)\,\Sigma\, A_j$ and A_c is the area of a cell. Thus, the fraction of boundary cells F is given by

$$F = \frac{N^B}{N^T} = K_1\sqrt{\pi}\left(\frac{A_c}{\bar{L}}\right)\left(\frac{\overline{\sqrt{A}}}{\bar{A}}\right) \tag{5}$$

For a square cell of side a, Crapper (1980) found that $\bar{L} = 0.7935a$, $A_c = a^2$, and $K_1 = 1.82$; thus,

$$F = 4.065\, a\left(\frac{\overline{\sqrt{A}}}{\bar{A}}\right). \tag{6}$$

For a Landsat pixel, Crapper (1980) found that $\bar{L} = 53.04\ m$; thus,

$$F = 274.56\left(\frac{\overline{\sqrt{A}}}{\bar{A}}\right). \tag{7}$$

In order to simplify Equation 5, an assumption must now be made about the distribution of the A_j's. As previously mentioned, the distribution of A_j has been examined by Crapper (1981a) for 1605 regions. It was found that the lognormal distribution provided a very good model for the size distribution of region areas. The region areas are displayed on a lognormal plot on Figure 1. Goodchild and Moy

(1976) also found that the lognormal distribution provided a good fit to maps of land use. In general, the areal distribution of regions is not tabulated in the scientific literature; however, an exception is Ackerson and Fish (1980). The areal distribution of their landscape units (their Table 3) was plotted and can be quite satisfactorily modeled by a lognormal distribution.

The lognormal distribution has been treated in considerable detail by Aitchison and Brown (1957). Many examples of lognormal distributions are noted in fields ranging from sedimentary petrology to literary style to economic data. Speight (1971) has observed lognormal distributions of landform slope in a number of independent studies. The distribution provides a satisfactory model when the processes underlying change or growth combine in a multiplicative rather than an additive fashion (as required by the normal distribution).

Now assuming that the distribution of region areas can be adequately modeled by a lognormal distribution, then $E(\sqrt{A})$ can be evaluated (see Appendix A).

Thus, substituting Equation A-2 into Equation 5 gives

$$F = K_1 \sqrt{\pi} \left(\frac{A_c}{\overline{L}}\right) \exp\left\{ - \left(\frac{\mu}{2} + \frac{3\sigma^2}{8}\right)\right\} . \quad (8)$$

Evaluation of the Number of Boundary Cells and Standardized Error of Area Estimates for the South Coast Data Base when Overlaid by Landsat Grid

The number of boundary cells can now be determined for the previously mentioned geographic data base. From the ln a data, displayed on Figure 1, the $\mu = 14.0584$ where a is given in m^2 and $\sigma^2 = 2.1831$. Hence, from Equation A-2, $E(\sqrt{A}) = 1483$ m. The mean and variance can be determined directly from Figure 1 using the quantile relationships given in Aitchison and Brown (1957). The cumulative probability § is given by

$$\begin{aligned} \S_{0.16} &\cong \mu - \sigma \\ \S_{0.50} &= \mu \\ \S_{0.84} &\cong \mu + \sigma \end{aligned}$$

Thus, $\mu \cong 14$ and $\sigma \cong 1.5$, which agree reasonably well with the above results.

For this case $E(\sqrt{A})$ could be independently calculated from the raw area data and was found to be 1484 m, which was an excellent agreement with the above result.

Now, collecting the previously quoted values $K_1 = 1.82$, $\overline{L} = 53.04$ m, $A_c = 4514$ m^2, $\mu = 14.0584$, and $\sigma^2 = 2.1831$ and substituting into Equation 8, one finds that 10.7 percent of the total pixels are mixels.

The error variance s_j^2 of the area estimate for one bounded region j has previously been determined by Crapper (1980), i.e.,

$$s_j^2 = N_j^B \, \overline{a^2} \quad (9)$$

where $\overline{a^2}$ is the mean square of the smaller area obtained when a random straight line is laid across a grid cell. Thus, the average error variance $\overline{s^2}$ of the area estimate is given by substituting Equation 4 into Equation 9 as follows:

$$\overline{s^2} = \frac{1}{M} \Sigma \, s_j^2 = \frac{\overline{a^2}}{M} \Sigma \, N_j^B$$

and hence

$$\overline{s^2} = \frac{\overline{a^2}}{\overline{L}} \cdot 2K_1 \sqrt{\pi} \, \sqrt{\overline{A}}.$$

The square root of the average error variance (this is an approximation to the average standard deviation) can be normalized by the average area to produce a standardized error estimate (it is to be noted that this an approximation to, but not the same as, the more commonly used relative error $(\overline{s/A})$ for which it was not possible to develop an analytical expression) as follows:

$$\frac{\sqrt{\overline{s^2}}}{\overline{A}} = \sqrt{\frac{\overline{a^2}}{\overline{L}} \cdot 2K_1 \sqrt{\pi}} \cdot \sqrt{\frac{\sqrt{\overline{A}}}{\overline{A}}}$$

$$\frac{\sqrt{\overline{s^2}}}{\overline{A}} = \sqrt{\frac{\overline{a^2}}{\overline{L}} \cdot 2K_1 \sqrt{\pi} \cdot \exp\left\{ - \left(\frac{3\mu}{4} + \frac{7\sigma^2}{16}\right)\right\}} . \quad (10)$$

For the case where the overlaying grid is a Landsat grid, Crapper (1980) found that $\overline{a^2} = 1.269 \times 10^6$ m^4. Substituting the parameters relevant to the South Coast Data Base into Equation 10 gives the standardized error $(\sqrt{\overline{s^2}}/\overline{A})$ equal to 0.4 percent. This result seems entirely reasonable as the average region area to pixel ratio is a comparatively large 842.

Conclusion

In this paper an expression has been derived for the number of grid cells which contain a boundary segment when a polygon type map is coded to grid cells. The implicit assumption is that the regions have definite boundaries. This assumption is appropriate if one were using a microdensitometer type instrument to scan an existing regionalized map. However, if instead of a regionalized map one were scanning the Earth's surface (the Landsat situation) or an aerial photograph, the assumption of an existing regionalization may not be so appropriate. The assumption of definite boundaries would be satisfied if the Earth's surface contained sharp transitions, e.g., intensive horticulture, agriculture, pasture, or silviculture. In general, however, the boundaries are imposed on the scene by the interpreter for his own reasons, and it may not even be possible to find them in nature. This situation arises whenever there is a gradual transition from one type of land cover to another, for example, grassland-woodland-forest. However, in the Landsat situation,

independent of whether the 'boundaries' are produced by sharp transitions or gradual transitions, the classification procedure (given appropriate instructions) will produce an objective and repeatable regionalization. The areal statistics of this regionalization can then be used to calculate the number of boundary pixels.

Imprecision in boundary definition may limit the resolution of satellite (and other raster scanned data) to an extent not immediately obvious from the pixel size. In this paper the number of boundary cells has been determined given the size of the overlay grid, the shape factor of the regionalization, and the mean and variance of the areal distribution of regions. This result can be used for answering two separate questions. First, given a regionalization with known (or assumed) mean and variance, what is the maximum grid cell size for a given percentage of boundary cells? Second, given a grid cell size (which is more common), what is the average region area for a given percentage of boundary cells?

Acknowledgments

The author would like to express his thanks to Dr. D. L. B. Jupp for many helpful conversations and checking the mathematics and to Mr. P. A. Walker for helpful discussions.

References

Abramowitz, M., and I. A. Stegun (Editors), 1965. *Handbook of Mathematical Functions*, Appl. Math. Ser., 55. National Bureau of Standards, Washington, D.C., 1046 p.

Ackerson, V. B., and E. B. Fish, 1980. An Evaluation of Landscape units. *Photogrammetic Engineering and Remote Sensing*, V. 46, pp. 347-358.

Agterberg, F. P., 1974. *Geomathematics, Mathematical Background and Geo-science Applications*, Elsevier, Amsterdam.

Aitchison, J., and J. A. C. Brown, 1957. *The Lognormal Distribution*, C.U.P., Cambridge.

Antenucci, J. C., 1979. Maryland Automated Geographic Information System: Expansions and Applications in State Government. Boston, Massachusetts: Harvard Library of Computer Graphics. 1979 mapping collection, Volume 3, *Urban, regional and state applications:* plus a special section on cadastral systems, pp. 5-21.

Austin, M. P., and K. D. Cocks, 1978. *Land Use on the South Coast of New South Wales*, 4 volumes, CSIRO, Melbourne.

Bryant, N. A., and A. L. Zobrist, 1978. An Image Based Information System: Architecture for Correlating Satellite and Topological Data Bases. *Proceedings of First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems*, Vol. IV, Cambridge, Mass., Harvard. Laboratory for Computer Graphics and Spatial Analysis.

Crapper, P. F., 1980. Errors Incurred in Estimating an Area of Uniform Land Cover Using Landsat, *Photogrammetric Engineering and Remote Sensing*, V. 46, pp. 1295-1301.

———, 1981a. Geometric Properties of Regions with Homogeneous Biophysical Characteristics, *Australian Geographical Studies*, V. 19, pp. 117-124.

———, 1981b. The Relationships Between Region Shape and Error Variance. *Proceedings of Landsat 81 Conference*. Canberra 1981.

Crapper, P. F., and A. P. Spate, 1982. Rational Land Use Planning: Puckapunyal Army Training Area. *Journal of Environmental Management*, V. 15, pp. 351-361.

Cross, R. H., 1979. A Case Study of Computer-aided Transmission Line Siting. Boston, Massachusetts: Harvard Library of Computer Graphics. 1979 mapping collection, Volume 4, *Computer mapping in natural resources and the environment: including applications of satellite derived data*, pp. 13-33.

Frolov, Y. S., and H. D. Maling, 1969. The Accuracy of Area Measurement by Point Counting Techniques. *The Cartographic Journal*, V. 6, pp. 21-35.

Galloway, R. W., and M. E. Bahr, 1979. What is the Length of the Australian coast? *Australian Geographer*, V. 14, pp. 244-247.

Goodchild, M. F., and W. S. Moy, 1976. Estimation from grid data: the map as a stochastic process. *Proceedings of the Commission on Geographical Data Sensing and Processing*. Moscow, pp. 67-81.

Haggett, P., and R. J. Chorley, 1969. *Network Analyses in Geography*, Edward Arnold, London.

Halsey, G., and G. B. Smith, 1978. Geographic Information Systems—The Reality. *Proceedings of the Eight Australian Computer Conference*. Canberra 1978.

Jupp, D. L., E. M. Adomeit, M. P. Austin, P. Furlonger, and K. K. Mayo, 1979. The Separability of Land Cover Classes on the South Coast of N.S.W. *Proceedings of Landsat 79 Conference*, Sydney, May 1979.

Laut, P., and T. A. Paine, 1982. A Step Towards an Objective Procedure for Land Classification and Mapping. *Applied Geography*, V. 2, pp. 109-126.

Maggio, R. C., R. D. Baker, and M. K. Harris, 1983. A Geographic Data Base for Texas Pecan, *Photogrammetric Engineering and Remote Sensing*, V. 49, pp. 47-52.

Mandelbrot, B. B., 1977. *Fractals: Form, Chance and Dimension*, W. H. Freeman, San Francisco.

Peuquet, D. J., 1982. Vector/Raster Options for Digital Cartographic Data. *Proceedings of US—Australian seminar on design and implementation of computer-based Geographic Information Systems*, Hawaii.

Schlesinger, J., B. Ripple, and T. R. Loveland, 1979. Land Capability Studies of the South Dakota Automated Geographic Information System (AGIS). Boston, Massachusetts: Harvard Library of Computer Graphics. 1979 mapping collection, Volume 4, *Computer mapping in natural resources and the environment: including applications of satellite derived data*, pp. 105-114.

Smith, A. Y., and R. J. Blackwell, 1980. Development of an Information Data Base for Watershed Monitoring, *Photogrammetric Engineering and Remote Sensing*, V. 46, No. 8, pp. 1027-1038.

Speight, J. G., 1971. Log-normality of Slope Distributions, *Z. Geomorph. N.F.*, V. 15, pp. 290-311.

Svatos, V. C., 1979. A Case Study of the Analysis of the Capability of Undeveloped Lands to Contribute to

Water Quality Degradation. Boston, Massachusetts: Harvard Library of Computer Graphics. 1979 mapping collection, Volume 4, *Computer mapping in natural resources and the environment: including applications of satellite derived data*, pp. 115-133.

Wehde, M., 1982. Grid Cell Size in Relation to Errors in Maps and Inventories Produced by Computerized Map Processing, *Photogrammetric Engineering and Remote Sensing*, V. 48, No. 8, pp. 1289-1298.

Wilson, P. M., 1979. BASIS: The Bay Area Spatial Information System. Boston, Massachusetts: Harvard Library of Computer Graphics. 1979 mapping collection, Volume 3, *Urban, regional and state applications plus a special section on cadastral systems*, pp. 151-156.

Zobrist, A. L., 1982. Integration of Landsat Data with Geographic Data Bases. *Proceedings of US—Australia seminar on Design and Implementation of Computer-Based Geographic Information Systems*, Hawaii.

(Received 12 December 1983; accepted 22 March 1984; revised 17 June 1984)

Appendix A

Following Agterberg (1974), p. 195, or Aitchison and Brown (1957), the probability density function for log normally distributed area a is given by

$$f(a) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{\ell n\, a - \mu}{\sigma}\right)^2\right\} \quad \text{(A-1)}$$

where μ and σ are the mean and variance of the normally distributed variate $\ell n\, a$. Aitchison and Brown (1957) present formulae for the mean and the variance of A when $\ell n\, A$ is normally distributed but is not possible to say anything about the mean of $\sqrt{A}$ as required by Equation 5.

If A is treated as a continuous positive definite random variable, then

$$E(A) = \int_0^\infty a f(a) da$$

Further, if g(a) is a real single valued function of a, then the expectation of g(a)

$$E(g(A)) = \int_0^\infty g(a) f(a) da.$$

Hence, if

$$g(A) = \sqrt{A},$$

then

$$E(\sqrt{A}) = \int_0^\infty \sqrt{a}\, f(a) da$$

and substituting in Equation A-1, then

$$E(\sqrt{A}) = \int_0^\infty \frac{a^{-1/2}}{\sigma\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}\left(\frac{\ell n a - \mu}{\sigma}\right)^2\right\} da$$

Now let

$$t = \frac{1}{\sqrt{2}} \frac{\ell n a - \mu}{\sigma}.$$

Then

$$a = \exp\{\sqrt{2}\sigma t + \mu\}$$

and

$$\frac{da}{dt} = \sqrt{2}\sigma \exp\{\sqrt{2}\sigma t + \mu\}.$$

Hence,

$$E(\sqrt{A}) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp - \left\{t^2 - \frac{\sigma}{\sqrt{2}} t - \frac{\mu}{2}\right\} dt$$

$$= \frac{1}{\sqrt{\pi}} \left(\int_0^\infty \exp - \left\{t^2 - \frac{\sigma}{\sqrt{2}} t - \frac{\mu}{2}\right\} dt + \int_0^\infty \exp - \left\{t^2 + \frac{\sigma}{\sqrt{2}} t - \frac{\mu}{2}\right\} dt \right).$$

Abramowitz and Stegan (1970) §7.4.2, p. 302, give

$$\int_0^\infty \exp -\{at^2 + 2bt + c\} dt = \tfrac{1}{2}\frac{\sqrt{\pi}}{\sqrt{a}} \exp\left\{\frac{b^2 - ac}{a}\right\} erfc \frac{b}{\sqrt{a}} \text{ for R } (a) > 0.$$

Hence,

$$E(\sqrt{A}) = \tfrac{1}{2} \exp\left\{\frac{\sigma^2}{8} + \frac{\mu}{2}\right\} \left(erfc\left(- \frac{\sigma}{2\sqrt{2}}\right) + erfc\left(\frac{\sigma}{2\sqrt{2}}\right)\right) = \exp\left\{\frac{\sigma^2}{8} + \frac{\mu}{2}\right\} \quad \text{(A-2)}$$