

1 of 1

AM/ER/PP - 74856

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

RESCALED BOX COUNTING FOR THE TOPOLOGICAL ANALYSIS OF SPATIAL DATA

Craig Loehle
Environmental Research Division
Argonne National Laboratory ER/203
9700 South Cass Avenue
Argonne, IL 60439

RECEIVED
APR 18 1994
OSTI

ABSTRACT

Topological analysis of surfaces of natural objects can provide information about surface features (ridges, fragmentation, dendritic patterns) and surface roughness. Box counting is a general method useful for such surfaces, but it is currently limited to cases where the x, y, and z directions are all in the same metric. A method, rescaled box counting, is presented for overcoming these limitations. The local first omnidirectional semivariance (lag 1) is calculated for boxes of different sizes. If the semivariance is not small for small box sizes, then the z data need to be scaled up to allow detection of a difference between patches that are significantly different. This rescaling converts the z metric into a distance equivalent (z units are converted into distances based on the horizontal distance over which a significant change in z is found to occur). Once rescaling is done, box counting can be used to quantify surface topology.

The submitted manuscript has been authored by a contractor of the U. S. Government under contract No. W-31-109-ENG-38. Accordingly, the U. S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U. S. Government purposes.

MASTER

Se

In this paper, I propose methods for analysis of the topology of surfaces. Analysis of surface features has potential utility in climatology (surface roughness), geomorphology (terrain analysis), ecology (coral reef structure, vegetation structure, pollution deposition), and other fields. Of interest here are surface roughness and spatial patterning in three dimensions. Spatial patterning includes detection of linear features such as ridges, isolated peaks such as volcanoes, dendritic patterns such as river valleys, and isolated features such as chains of islands. Roughness, suitably quantified, gives information about exchange processes (water through a coral reef, air into a forest canopy), growth processes (mineral dendritic growth), and disturbance processes (blowdowns affecting forest canopy roughness). Many papers have dealt with spatial patterns of flat maps (Bertuzzi et al., 1990; Bradbury et al., 1984; Burrough, 1981; Chopard et al., 1991; De Cola, 1989; Gardner et al., 1987, 1989; Krummel et al., 1987; Lam, 1990; Milne, 1988; O'Neill et al., 1988 a, b; Palmer, 1988; Rex and Malanson, 1990; Wiens and Milne, 1989). However, only a few (e.g., Loehle, 1990, submitted; Steacy and Sammis, 1991) have yet addressed the 3-D case, which is more complex.

Consider the landscape in Figure 1. We would like to be able to describe its features at different scales, perhaps for a wildlife habitat analysis or for wind turbulence analysis. A method suitable for this purpose is box counting (Loehle, 1990; Voss, 1988). Although several methods are available for fractal calculation, box counting is the most general method that is easily applied to empirical data. The divider method is restricted to linear features. The perimeter/area ratio method is confined to well-defined patches on a flat map. Certain other methods either assume a single fractal dimension for the surface (e.g., semivariance analysis) or are difficult to compute for empirical data. Box counting has the advantage that it works equally well for line data, scattered point or patch data on a flat map, or 3-D surfaces. Many of the references cited above used box counting. Box counting is currently only applicable to restricted cases, however. Box counting fills or covers the object with cubes of different sizes (see below). It is necessary that the vertical (z) dimensional units of the data must be the same as the x and y dimensions before box counting may validly be used (Voss, 1988), as, for example, in the rock fracture analysis of Steacy and Sammis (1991). Thus a data set consisting of pollutant concentrations across an area can not be analyzed as a surface because the dimensions are incommensurate between ppm and meters. Available methods that allow us to analyze such data on a map are restricted in their utility. Spectral analysis is not very informative if no periodic patterns occur in the data. In contrast to time series data, meaningful periodicity will probably not

be found on a landscape. Calculation of a fractal dimension from a semivariogram involves a function of the slope of the rising part of the semivariance curve. In practice, this rising slope is rarely smooth (e.g., Isaaks and Srivastava, 1989), and calculation of a fractal dimension is problematic. Furthermore, the semivariogram for a surface with a single symmetric object in the middle such as a mountain peak is not interpretable at all fractally. Thus, since box counting does work for surfaces of all types if dimensions in all directions are commensurate, the problem of incommensurability bears reexamination. The objective of this paper is to propose a revision to box counting that removes this limitation.

Box counting quantifies the space-filling properties of objects at different scales. Changes in these properties with scale indicate whether the object being quantified is a solid object ($d=3$), a linear object ($d=1$), a flat object ($d=2$) or something intermediate (where d is the fractional dimension). For example, a series of balls in a line are of dimension $d=3$ at fine scales, but at coarser scales the linear feature of the line of balls dominates, yielding $d=1$. A series of d values is obtained at different scales, giving a fractional dimensional profile (as in Figures 3 and 6 of this paper). Note that in this approach one does not assume that d is the same at all scales (is self-similar) and d may even turn out to be integer-valued. This multiscale approach is similar in principle to that of Halsey et al. (1986), except that their approach is more suited to complex geometric objects such as cantor sets and chaotic attractor sets where box counting becomes inadequate. For the purpose of characterizing the topology of natural surfaces, a fractional dimensional profile is probably adequate.

Consider a square map in which every pixel is of type K. At the level of one division along the x-axis (the whole map is covered by one box), there is one square unit containing K. When the x-axis is subdivided into 2, there are 4 squares of type K. In general, the number of boxes equals t^d , where t is number of tick marks dividing up the axes. The fractional dimension d of the type K is the power of t . In this example $d = 2$, K fills the plane, and the map dimension is not fractional. Consider a second case, type P in which P is a narrow linear feature. In this case $d = 1$. If the map is a single point or empty, $d = 0$. Regions of the map that occupy space (have a height) but are fragmented will have $1 < d < 3$ as measured by cubes occupied at different scales. Those that are fragmented into tiny widely scattered fragments will have $0 < d < 1$. In these last two cases d is fractional. This single algorithm can find d for trail data (e.g., an animal trail or river), transect data, scattered point data, patch-type data, and elevation-type (3-D) data, thus obviating the need to use several different (possibly incommensurate) methods for different types of data.

Because we wish to avoid prejudging self-similarity, d is calculated at several scales. The dimension d_f at a given scale is defined by the slope $\Delta x/\Delta y$ (from the slope of a log-log plot of boxes as a function of scale), where Δx is $\log(B_{i-1}) - \log(B_i)$, \log is natural log, (B_i) is the number of boxes, i is the power of 2, and Δy is $\log(S_i) - \log(S_{i-1})$, where S_i is the number of steps along the map edge. The ratio $\Delta x/\Delta y$ gives a discrete estimate of d_f as one moves from one scale to another.

Rescaled Box Counting

For landscapes, it becomes evident that a modification to box counting is needed. Let

$$Z_{x,y}, \{x,y\} \in M$$

be the elevations at the points $\{x,y\}$, with $\{x,y\}$ contained in the set M , which defines the map extent (not necessarily square), and with z,x,y all in the same units. Consider a case where M is 2 by 2 miles of rugged terrain near the coast. Assume that an analysis of the roughness of the surface of this landscape gave a fractional dimension $d=2.3$. If an otherwise identical landscape were located at 5,000 ft elevation, the topological analysis of this object would become insensitive to the minor bumps on the 5,000 ft by 2 mile by 2 mile block on which they sit. A d value near 3.0 would be the result, which clearly does not describe what we wish to know about this surface.

For characterization of the surface properties of a soil, a coral reef, or a landscape, the method used should not exhibit the above sensitivity to absolute elevation. Correcting for this bias yields the new method of rescaled box counting. A simple way to correct for elevation differences is to subtract the lowest elevation over M from each data point. To keep all elevations positive, we add back a small number so that no point has zero elevation (which would cause holes in the map for box counting). This gives

$$Z'_{ij} = Z_{ij} - \text{Min}(Z_{ij}) + e$$

where e is equal to the width of the pixels used to characterize the data. Thus any point that has non-zero elevation is rescaled to be at least 1 pixel high. This rescaling of the elevation data allows one to proceed with box counting of the solid volume defined by Z'_{ij} . Using this rescaled box counting, a flat plane is quantified as flat ($d=2$) and a rugged one as rugged any where on the earth.

A second rescaling may be necessary if the z data are in units such as ppm of a pollutant at locations (x,y). It is not valid to use box counting when the dimensions (x,y,z) are not all in the same units. The following method was developed for rescaling. We may express the smoothness of a surface in terms of how far we must go to detect a significant change in z. One way of assessing the scale at which a surface is smooth is to compute the mean squared difference between the means in adjacent patches for patches of different sizes, using a moving window of size B to compute patch averages. For example, for a window size of 1 and a regular (x,y) grid of z values, we would sum $(Z_{11} - Z_{12})^2 + (Z_{11} - Z_{21})^2 + (Z_{12} - Z_{13})^2 + (Z_{12} - Z_{22})^2$ etc. The semivariance is then

$$\gamma = \frac{1}{N} \sum (Z_a - Z_b)^2$$

where N is the actual number of differences computed and a and b follow the noted right and down comparison scheme so that two points are only compared once (the usual correction of 1/2 need not be applied). For a window size of 2 x 2 (B=2),

$$\left(\frac{1}{4} \sum_{i=1}^2 \sum_{j=1}^2 Z_{ij} - \frac{1}{4} \sum_{i=1}^2 \sum_{j=3}^4 Z_{ij} \right)^2$$

would be the first squared difference of means,

$$\left(\frac{1}{4} \sum_{i=1}^2 \sum_{j=1}^2 Z_{ij} - \frac{1}{4} \sum_{i=3}^4 \sum_{j=1}^2 Z_{ij} \right)^2$$

would be the second, etc. A plot of γ vs. B gives a result like that in Figure 2. The value of γ is small for small B because we are comparing adjacent points at a very fine scale where they should be almost identical. At some scale B', γ begins to rise. This γ is designated $\bar{\gamma}$. This B' is in a sense the smallest average patch size in which z is uniform or very small. The γ value $\bar{\gamma}$ at which we find this B' corresponds to a Δz as

$$\gamma' = \frac{1}{N} \sum (\Delta z)^2$$

if we assume that a typical or mean difference between adjacent windows can be represented by Δz . Then

$$\gamma' = \frac{1}{N} N \Delta z^2$$

since the $N \Delta z$ are all the same, giving

$$\Delta z = \sqrt{\gamma'}$$

Thus the Δz corresponding to the average smallest detectably-different-from-its-neighbors patch size B' (in units of distance squared, not of pixel counts) is $\sqrt{\gamma'}$. We may use this value to scale all the z values.

A test was conducted with actual topographic data. For such data, the x , y , and z dimensions are all in the same units, so the result of rescaling should be that the original data are retained. The spearfish data set from GRASS was used. GRASS is the GIS package from the U.S. Army Corps of Engineers Construction Engineering Research Laboratory. The spearfish data set is a digital elevation model covering parts of two topographic 1:24,000 quads (Spearfish and Deadwood North) in western South Dakota (Figure 1). It includes a majority of the Black Hills National Forest. The data are on a regular 30 m grid. Because all data points are at a high elevation, the preliminary fractal calculation of the roughness of this topography gives a result of $d \approx 3$ at all scales (not shown). This value merely indicates that the site is at a high elevation and tells us nothing about the topology of the surface. The minimum elevation within the delimited area was thus subtracted from every point, as described above. After this, the local variances were computed (Figure 2). As expected, γ rises from a low level as B (box size) increases, and then levels off. At very large box sizes γ might decrease as the mean within each box approaches the overall mean elevation and the differences between adjacent boxes approach zero. The smallest patch size at which the data are uniform in this data set is 30x30 m (1x1 pixels). The very small patch-to-patch variance of 35.27 m corresponds to a Δz of 5.94 m over a 30 m distance. The gradient $\Delta z/\Delta L$ is thus $5.94/30 = .198$. This 20% gradient

between adjacent pixels at the finest scale is an adequate difference to detect, and thus we do not rescale the data. If this gradient were extremely small, then the grid resolution would be unnecessarily fine and could be coarsened to save computational effort. If the variance at the finest scales of B is large relative to the asymptote of γ , then the data grid is probably too coarse to accurately represent fine-scale features of the data. As a final step, the z values are converted from meters to pixels by dividing by 30. The x , y and z axes are all now expressed in 30 m increments with integer values. These data are fed into the box counting program, to yield Figure 3. At the finest scales the fractional dimensions range from $d_f = 2.5$ to 2.91, indicating rough topography as we would expect for this particular landscape. The 2.91 value at the 30x30 m to 60x60 m scale indicates that hills on the map are large and smooth relative to this size range. At scales from 1920 m to 3840 m, the topography is quantified as flat, indicating a lack of steep canyons or cliffs at these scales. Above these scales the map is quantified as less than 2-D because the map is rectangular.

A rescaling example helps to further illustrate the method. A wetland zone along Tinker Creek in southwestern South Carolina was classified by vegetation types. These types were arbitrarily ordinated along a dry-to-wet vegetation type "elevation" axis ranging from disturbed soil at $z=1$ to bottomland hardwood forest at $z=10$. Note that an ordination scaled from 0 to 1 would give a completely flat map with no vertical relief. The elevation map thus derived is shown in Figure 4 (with a great deal of vertical exaggeration). The objective of this elevational representation of habitat types was to compute a habitat diversity index (Loehle, submitted). The local elevational difference variances vs. block size were calculated as above (Figure 5) in order to calculate a rescaling of these arbitrary elevations in terms of the variance of types between vegetation patches at the smallest average patch size. The edges between the wetland zones and the flat (non-wetland) part of the map were not used in the variance calculation, because doing so artificially inflates the variance.

In contrast to the topographic data, the γ values for this data set begin high and then decline because the spatial grid on which cover types are defined is too coarse relative to the vertical scale difference between different types on the landscape. Thus this data set does need vertical exaggeration before fractal calculations or else all detail will be lost and the map will appear flat. Since the variability in the data appears to be confined to patch sizes below 4x4 pixels, we wish to scale up the data so that the smallest difference between types (1 arbitrary unit) is at least 3 pixels high. Thus the rescaling is 1 unit = 3 pixels. This

rescaling, as used in Loehle (submitted) gives the fractional dimensional profile shown in Figure 6. There is considerable topographic relief at fine scales and at coarse scales the shape of the delimited area dominates (see Loehle, submitted, for more details). Just as in the practice of ordination, different sites may be compared only if they are scaled to the same metric. The variance vs. block size analysis must be conducted using all maps to be compared to determine a scaling, and this scaling must be used for all of the maps. This procedure was followed for the wetland areas and both maps were scaled by the same 1 unit = 3 pixels scaling (Loehle, submitted).

In conclusion, this method offers hope for rescaling data on maps from any axis (ppm, densities, ordination axis distances) to equivalent elevations so that fractal measures of surface topology can be calculated. These measures have the advantage of being free of most assumptions (such as that periodicity exists) and are not affected by such features as isolated peaks which can distort variogram analyses. This method could therefore be very useful in the analysis of the spatial patterns of pollution, wildlife populations, ore deposits, and other spatial features.

ACKNOWLEDGEMENTS

This paper was prepared in connection with work performed under contract W-31-109-ENG-38 with the U.S. Department of Energy, Office of Energy Research, Office of Health and Environmental Research. Helpful reviews were provided by Barry Lesht and Carol Rosignolo. Programming assistance was provided by Carol Rosignolo. Topographic data were provided by Marc Levine. Remote sensing data collected and analyzed under contract DEAC-0976SR00-819 with the U.S. Department of Energy.

LITERATURE CITED

- Bertuzzi, P., G. Rauws, and D. Couralt. 1990. Testing roughness indices to estimate soil surface roughness changes due to simulated rainfall. *Soil and Tillage Res.* 17:87-89.
- Bradbury, R. H., R. E. Reichelt, and D. G. Green. 1984. Fractals in ecology: methods and interpretation. *Marine Ecology Progress Series* 14:295-296.
- Burrough, P. A. 1981. Fractal dimensions of landscapes and other environmental data. *Nature* 294:240-242.
- Chopard, B., H. J. Herrmann, and T. Vicsek. 1991. Structure and growth mechanism of mineral dendrites. *Nature* 353:409-412.
- DeCola, L. 1989. Fractal analysis of a classified landsat scene. *Photogrammetric Engineering and Remote Sensing* 55:601-610.
- Gardner, R. H., B. T. Milne, M. G. Turner, and R. V. O'Neill. 1987. Neutral models for the analysis of broad-scale landscape pattern. *Landscape Ecology* 1:19-28.
- Gardner, R. H., R. V. O'Neill, M. G. Turner, and V. H. Dale. 1989. Quantifying scale-dependent effects of animal movement with simple percolation models. *Landscape Ecology* 3:217-227.
- Halsey, T.C., M. H. Jensen, L.P. Kadanoff, I. Procaccia, and B. I. Shraiman. 1986. Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev. A* 33:1141-1151.
- Isaaks, E. H., and R. M. Srivastava. 1989. *Applied Geostatistics*. Oxford University Press, NY.
- Krummel, J. R., R. H. Gardner, G. Sugihara, R. V. O'Neill, and P. R. Coleman. 1987. Landscape patterns in a disturbed environment. *Oikos* 48:321-324.

Lam, N. S.-N. 1990. Description and measurement of Landsat TM images using fractals. *Photogrammetric Engineering and Remote Sensing* 56:187-195.

Loehle, C. 1990. Home range: A fractal approach. *Landscape Ecology* 5:39-52.

Loehle, C. Submitted. Landscape habitat diversity: An index. To: *Ecology*.

Milne, B. T. 1988. Measuring the fractal geometry of landscapes. *Applied Mathematics and Computation* 27:67-79.

O'Neill, R. V., B. T. Milne, M. G. Turner, and R. H. Gardner. 1988a. Resource utilization scales and landscape pattern. *Landscape Ecology* 2:63-69.

O'Neill, R. V., J. R. Krummel, R. H. Gardner, G. Sugihara, B. Jackson, D. L. DeAngelis, B. T. Milne, M. G. Turner, B. Zygmunt, S. W. Christensen, V. H. Dale, and R. L. Graham. 1988b. Indices of landscape pattern. *Landscape Ecology* 1:153-162.

Palmer, M. W. 1988. Fractal geometry: A tool for describing spatial patterns of plant communities. *Vegetatio* 75:91-102.

Rex, K. D., and G. P. Malanson. 1990. The fractal shape of riparian forest patches. *Landscape Ecology* 4:249-258.

Steacy, S. J., and C. G. Sammis. 1991. An automation for fractal patterns of fragmentation. *Nature* 353:250-252.

Voss, R. F., 1988. Fractals in nature: from characterization to simulation. pp. 21-70 in: H.O. Peitgen and D. Saupe (eds.). *The Science of Fractal Images*. Springer, N.Y.

Wiens, J. A., and B. T. Milne. 1989. Scaling of "landscapes" in landscape ecology, or landscape ecology from a beetle's perspective. *Landscape Ecology* 3:87-96.

Figure 1. Actual topography for the spearfish data set from GRASS (see text).

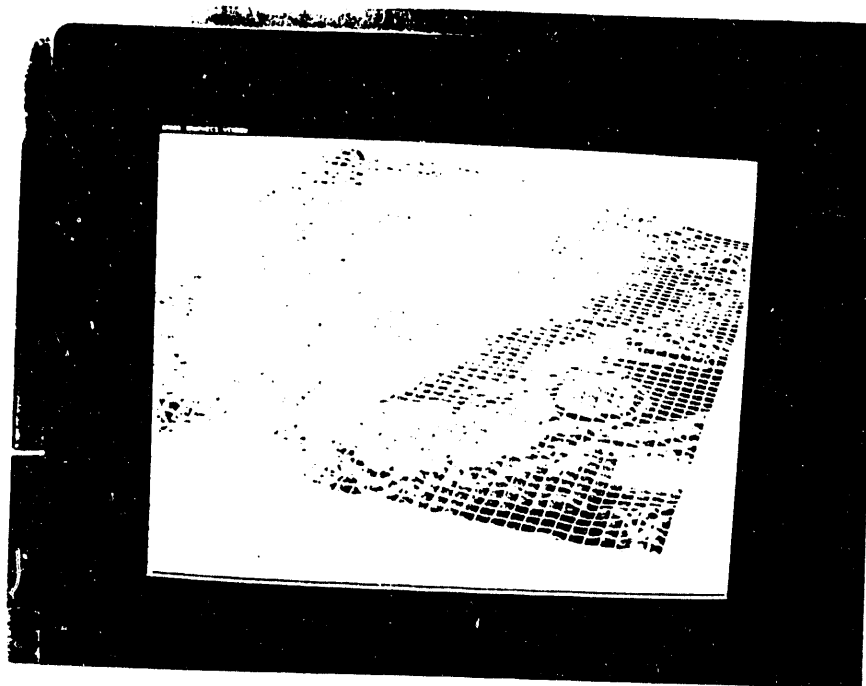


Figure 2. Local semivariance at different block sizes for the spearfish data.

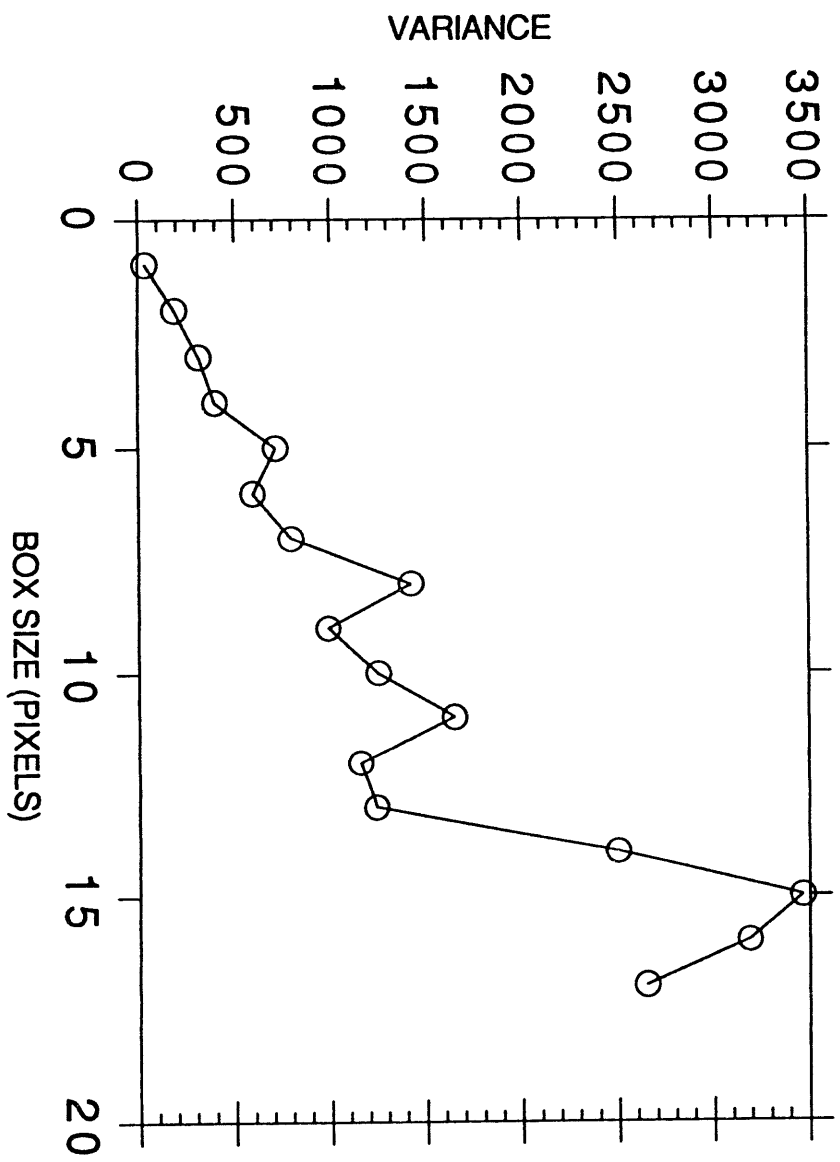


Figure 3. Fractional dimensional profile of the spearfish topographic data after rescaling to base elevation.

FRACTIONAL DIMENSION

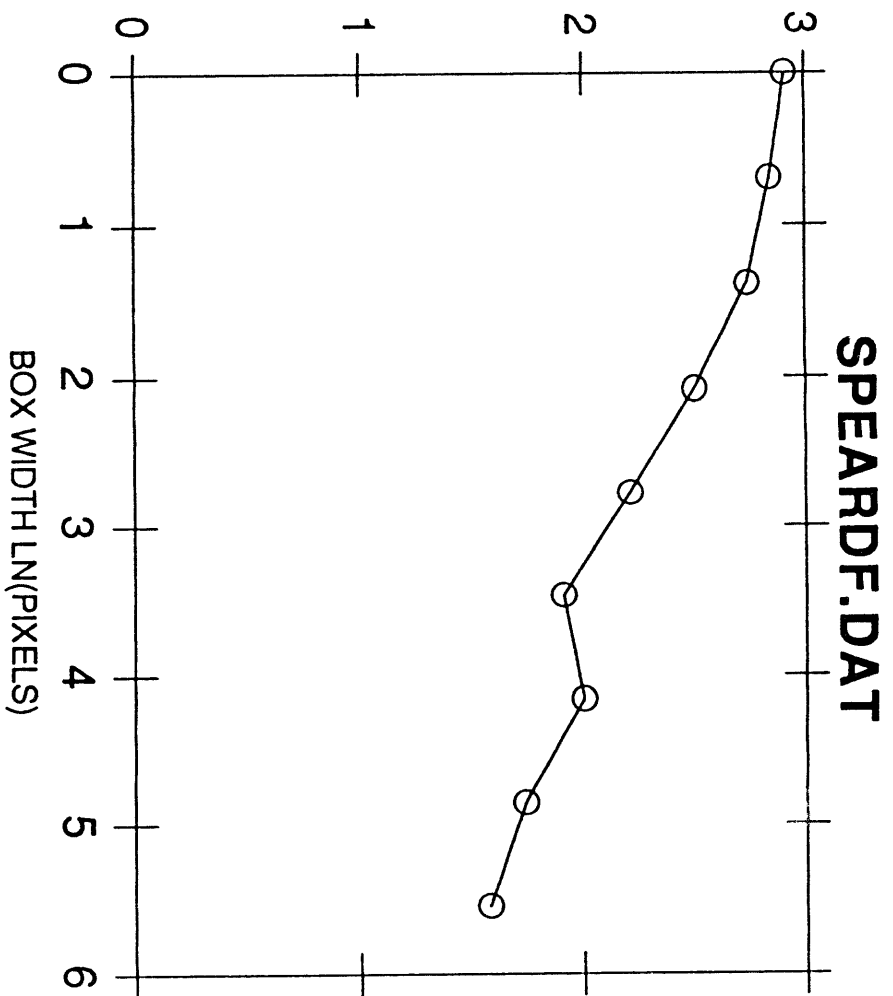


Figure 4. Artificial elevations based on an ordination of vegetation types in wetlands along Tinker Creek in South Carolina (vertical scale arbitrary).

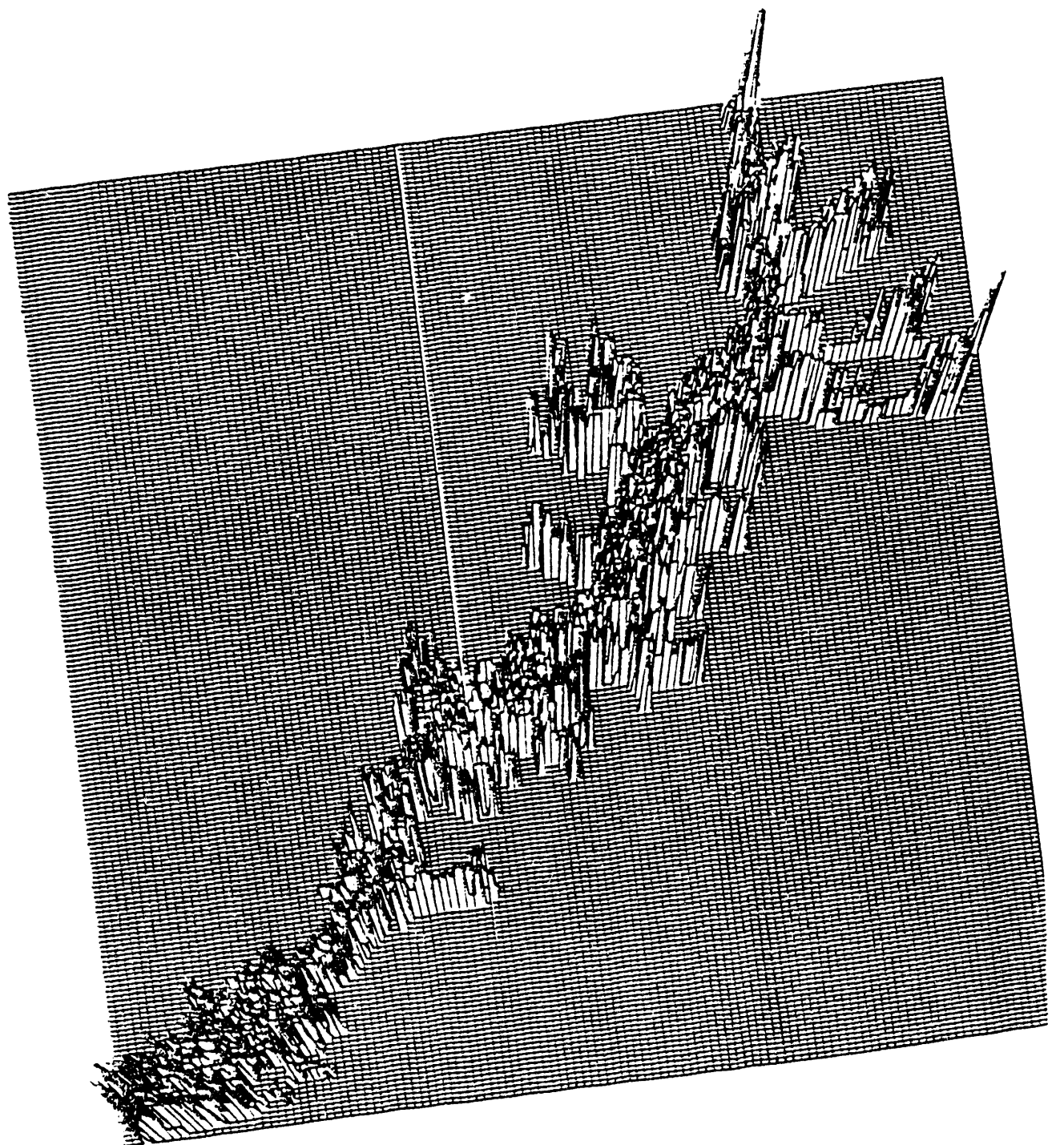


Figure 5. Local semivariances at different block sizes for Tinker Creek.

VARIANCE

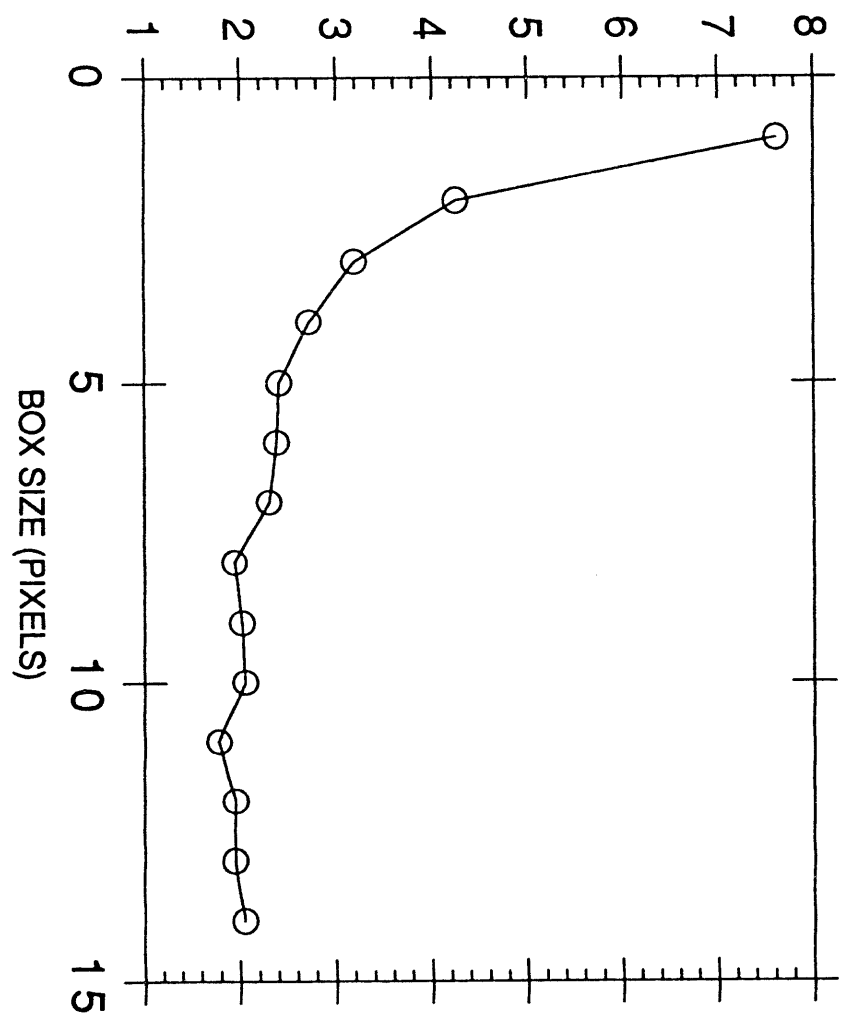
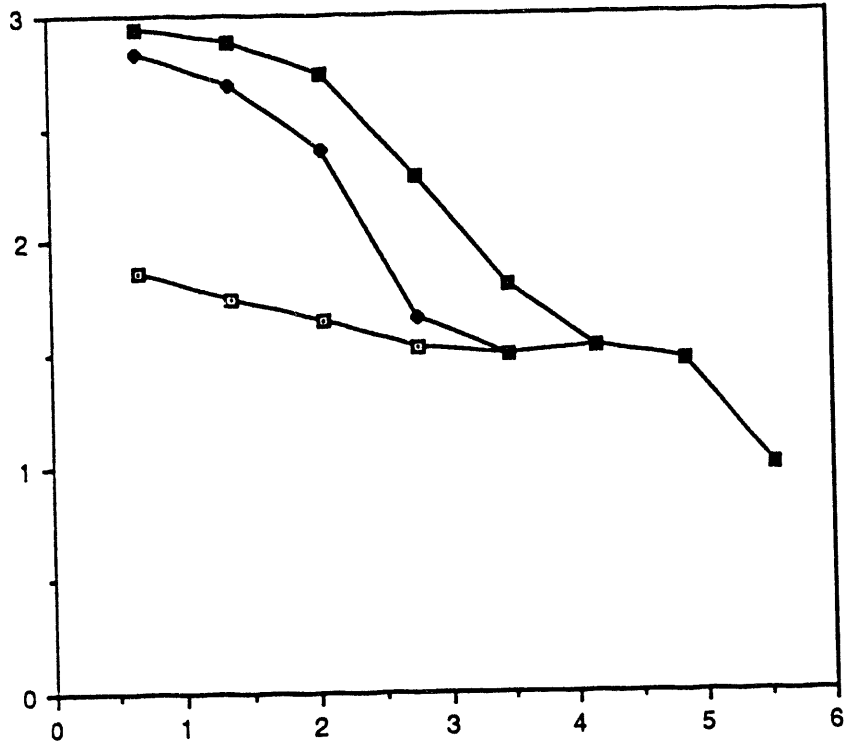


Figure 6. Fractional dimensional profiles before and after rescaling the Tinker Creek data. Two rescalings are shown.

FRACTIONAL DIMENSION



- tinker flat map
- tinker x 2
- tinker x 3

BOX LENGTH (LOG (# PIXELS))

DATE

FILMED

6/10/94

END

