Application of fractal geometry in determining optimal quadrat size for vegetation sampling

Anubhav Chaturvedi\textsuperscript{1,*} and P. Rama Chandra Prasad\textsuperscript{2}

\textsuperscript{1}Centre for Computational, Natural Sciences and Bioinformatics, and \textsuperscript{2}Lab for Spatial Informatics, International Institute of Information Technology, Gachibowli, Hyderabad 500 032, India

Geometry in ecological patterns of landscape and vegetation is not truly fractal, and varies across a range of scales, whereas fractal geometry provides tools for predicting and describing ecological patterns. In this study, fractal analysis is used to assess presence of pseudo random quadrats or spatial dependence which hamper generality and performance of classical inferential statistics. Fractal dimension (FD) as a function of scale is used to determine quadrat size which eliminates spatial dependence. The semivariograms are plotted with fractograms to correlate structures of spatial dependence of the properties studied. The use of FD as a degree of spatial dependence of variables is the basis of applications of fractals.

Keywords: Ecological patterns, fractal geometry, quadrat size, spatial dependence, vegetation sampling.

A significant challenge encountered in plant ecological studies is vegetation sampling\textsuperscript{1,2}. Researchers worldwide have analysed ecological attributes (species diversity, richness, dominance, etc.) of vegetation using random or stratified random sampling or by laying transects across some gradient\textsuperscript{3–6}. Ecologists have used larger contiguous area and researchers have also designed certain plots as ‘long-term ecological plots’ or ‘permanent dynamic plots’ to monitor variability in species characteristics in spatiotemporal domain\textsuperscript{7–15}. Whichever method is adopted, sampling is always a time-consuming process. Also, the size of the plot or quadrat that is used as the basic unit of sampling, varies depending on the type of vegetation and area covered. Though significant variation exists in the ecological patterns captured by random method (usually high and diverse) compared with large-area contiguous plots, these are basically used to understand the behavioural patterns of the species in contiguous scale\textsuperscript{11,16–19}.

Studying the large-area plots (which range between 1 and 50 ha), researchers have subdivided the entire plot into smaller units for better and quick sampling. The size of the smaller units are 1 m × 1 m, 10 m × 10 m, 30 m × 30 m or sometimes circular plots with varied dimensions\textsuperscript{20–22}. Within sub-units, ecologists study characteristic features of a species and its population or general diversity patterns, and compare the changing attributes across the quadrats conceptualizing the pattern at higher scale\textsuperscript{23}. But when comparisons are made between neighbouring or adjacent quadrats, probability of variation is low as it lies in the same homogenous conditions – may be precipitation, edaphic, sometimes topography. This indicates greater similarity in two closely spaced quadrats compared to those that are separated by larger distances. These samples may be referred to as pseudo replicates, violating the most important assumption of classical inferential statistics that the samples are spatially independent\textsuperscript{24}.

A homogeneous distribution is one that remains similar on repeated sub-division\textsuperscript{25}. The arrangement or ordering of data as a function of location is called spatial autocorrelation of the function and the range of spatial scales in which spatial autocorrelation exists is called spatial dependence\textsuperscript{26}. Avoiding spatially dependent quadrats (pseudo replicates), that do not contribute significant changes in any ecological property is necessary, to improve the performance of classical inferential statistics, as the existence of spatial dependence hampers the generality of results and overall performance of classical inferential statistics\textsuperscript{27}.

The concepts of fractal geometry can suggest a better statistically rectified sampling scheme, which eliminates the problem of spatial dependence among the quadrats\textsuperscript{25}. Optimal quadrat sizes for homogeneous or spatially independent distribution can be determined by using the methods of fractal analysis on the data. The quadrats of suggested sizes will be spatially independent and thus independent of the distances by which they are separated.
Fractals\textsuperscript{28-41} essentially have a non-integer dimension, unlike the corresponding Euclidean form, and can be described by fractal dimension (FD), which can be used to compare and categorize fractals\textsuperscript{40,42-44}. A non-integer dimension implies that the fractal has a dimension different from the space in which it resides\textsuperscript{45}. The FD has also been widely used as a measure of the space-filling ability of a pattern\textsuperscript{42,46}.

A perfectly random distribution will be homogeneous and have a FD of two\textsuperscript{25,45}. So by dividing the sample space into quadrats of size at which the FD is two, we can have a distribution which will be homogeneous.

In the perspective of these concepts, an attempt was made to suggest appropriate quadrat sizes for homogeneous distribution best suited for the methods of classical statistics, using FD analysis and to suggest appropriate spacing between the quadrats, restricting the number of quadrats to be sampled to a lower value. The study further aims, as a secondary objective, at drawing inference regarding the variation and correlation among various properties studied in the contiguous plot.

We have used three contiguous plot data, each of size 3 ha sampled in three different forest types, viz. evergreen (EG), semi-evergreen (SEG) and moist deciduous (MD) forests of Northern Andaman Islands. Intact, isolated plots were selected and each 3 ha plot was divided into 30 quadrats of 0.1 ha size (32 m × 32 m). In each quadrat primary attributes of tree data such as their identification, number, height and girth at breast height (> 30 cm) were measured. The data were used to derive mean height, basal area, density, species richness (number of species in each quadrat) and diversity (Shannon index) in each plot. These five attributes are considered important in characterizing the vegetation patterns in contiguous plots.

Semivariogram is the basic unit of geostatistics which summarizes the variance observed in a dependent variable as a function of scale and is defined as

\[
y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z(i) - z(i + h))^2,
\]

where \( y \) is the semivariance at scale \( h \); \( z \) the dependent variable, at any point \( i \), has the value \( z(i) \); \( z(i + h) \) the value of the dependent variable at a point separated from the point \( i \) by a distance \( h \) and \( N(h) \) is the number of points separated by distance \( h \).

The properties with similar semivariograms have similar structure of spatial dependence\textsuperscript{26}. The slope \( m \) of the double logarithmic plot of the semivariogram gives the FD of the distribution \( D \). The slope of the double logarithmic semivariogram for every distance \( h \) is calculated as\textsuperscript{25}

\[
m = \frac{\log \{ y(2h) \} - \log \{ y(h) \}}{\log(2h) - \log(h)}.
\]

This simplifies to,

\[
m = \frac{\log \{ y(2h) \} - \log \{ y(h) \}}{\log(2h) - \log(h)}.
\]

This formula can be considered as the change in variance of a dependent variable as one doubles the scale. The FD is calculated by\textsuperscript{47,48}

\[
D = \frac{4 - m}{2},
\]

and plotted as a function of scale that can be used to interpret the scale at which the distribution can be considered as homogeneous. The degree of spatial dependence across a range of sample scales is indicated by FD, whereas semivariograms indicate the structure of scale dependence\textsuperscript{27}.

According to eq. (1), if \( z \) is a linear function of \( h \), the semivariogram will be a parabola, as \( z(i) - z(i + h) \) is a linear function of \( h \). A linear function squared is a parabola, the slope of the double logarithmic plot of a parabola is 2, which corresponds to a FD of 1. If the values of \( z \) in two near samples are no more or less different from two distant samples, in other words, if the distribution is perfectly homogeneous, the slope of the semivariogram will be zero, corresponding to a FD of 2. The double logarithmic semivariogram may not be linear, hence both \( m \) and \( D \) are not necessarily constant functions of scale\textsuperscript{29}.

Further, fractal geometry, in particular the FD is not restricted between 1 and 2, i.e. just a function of position on a line, but can be readily extended to between 2 and 3, as a function of points on the plane\textsuperscript{22}.

At the first instance, one cannot check for homogeneity on scales less than that of the length of a side of the quadrat, which is 30.62 m, because of the restriction imposed by the unavailability of data within a quadrat. Further the formula applied, being fairly accurate, calculates FD by taking the difference between a scale value and the double value. This restricts the upper bound of the scale in the fractogram to half its value used to plot the semivariogram. This in turn ensures that the size of a quadrat does increase more than three times the original, which is half the maximum scale in the semivariogram.

As is evident from Figure 1a, the curve of semivariance versus scale is a parabola, the slope of double logarithmic plot of the parabola will be two, and the corresponding FD will be one, that is, strictly linear spatial dependence. FD is extremely close to one at all the scale points (Figure 1b). The slight error is because input was generated only to five decimal points, as \( \sqrt{2} \) is an irrational number, so more the decimal points, more accurate and close to one will be the result.

The difference in the values for a pair of closely spaced quadrats (lower scale) is more or less the same as compared to the difference in the values for a pair of quadrats...
placed at a larger distance (larger scale) (Figure 1 c). This implies that the slope of double logarithmic plot will be 0 and FD of the distribution will be 2 or close to 2 (random function of python is in effect pseudo random). In Figure 1 d, the FD is close to 2 at lower scale values and suffers a deviation due to pseudo random function of the python. A perfectly random distribution will result in a constant FD equal to 2, thus verifying the code and methodology applied.

Graphs like those of FD of height in EG and SEG forests, basal area in SEG forest (Figure 2) can be helpful in suggesting an ideal sampling scheme by choosing appropriate quadrat sizes, such that the quadrat delineates homogeneous vegetation, and quadrat spacing which allow spatial dependence. The U-shaped section of the fractogram, with two limbs of U lying on \( d = 2 \), is necessary for suggesting such a scheme. We have suggested an ideal sampling scheme wherever we obtained such FD versus scale graphs. The quadrat side should be of size corresponding to where the left limb of \( U \) reaches \( d = 2 \). The quadrats should be separated by a minimum distance corresponding to where the right limb crosses \( d = 2 \). If the graph increases further and starts decreasing sharply, the corresponding distance should be the maximum distance between any two neighbouring quadrats.

The results are in effect invariant to white/coloured noise (of a constant amplitude) addition to the value of the function, shown below

\[
z'(i) = z(i) + \text{noise}(0, 1) \times \text{constant}. \tag{5}
\]

The function noise(0, 1) is independent of \( i \) or the transect variable and follows a uniform random/Gaussian probability distribution depending upon the type of noise added. So,

\[
z'(i) - z'(i + h) = z(i) - z(i + h) + \delta, \tag{6}
\]

where

\[
\delta = (\text{noise'}(0, 1) - \text{noise'}+h(0, 1)) \times \text{constant}. \tag{7}
\]

This study attempts to identify reliable plot size for field sampling, taken ideally for intact undisturbed forests and for such samples the value of the constant is small and hence the value of delta tends to zero. The value of the constant depends on sampling precision and is low for repeated and rigorous sampling, which is assumed. Addition of noise in general moves the FD towards that of a
Figure 2. Semivariance versus scale/quadrant size/distance.
Figure 3. Fractal dimension versus scale/quadrant size/distance.
homogenous distribution. We suggest quadrant size for homogenous distribution which will not be affected by added random noise. Further, all results are based on FD of the distribution, which is dependent on the slope of semivariance (see below), and for small values of the constant the slope is not affected. So in general the suggested quadrant size will not be affected by environmental disturbance.

These values can be used to study specific properties of specific forest types with sufficient generality.

Figures 2 and 3 show the semivariance versus scale (quadrat size) and FD versus scale results respectively, for different properties for three forest types, viz. EG, SEG and MD.

In general, FD obtained was more close to 2 than 1, which suggests that vegetation has low spatial dependence. Nevertheless, the spatial dependence, even if it is low, hampers the performance of classical inferential statistics and results in loss of generality and needs to be dealt with.

The values of scale (quadrat size), where the graph of FD versus scale crosses FD = 2, with respect to the length of the original quadrat size are given in Table 1 (1 unit ~ 30.62 m). Figure 2 shows that the FD cannot be found for the values of scale less than that of the original quadrat size and greater than half of the highest scale used for plotting the semivariogram graphs, leading to lower and upper bounds to the change in quadrat size.

Plots for which ideal sampling scheme can be suggested are listed in Table 2, where a is the size of side of the quadrat, b is the minimum distance between two quadrats, and c is the maximum distance between any two neighbouring quadrats. The value of c, which is where the graph further increases after the right limb of the U shape and then starts decreasing sharply, has not been obtained for all the following plots due to constraints on scale.

Figure 3 shows that the plots obtained for species diversity and species richness for a particular forest type are similar, i.e. they have peaks at the same value of scale. This implies that these two properties for a particular forest type essentially have similar dependence on the scale (distance) transect. Thus FD versus scale plots are also similar for species diversity and species richness for a particular forest type. The semivariograms (Figure 2) of the plots of basal area for all three forests are similar especially basal area plots for SEG and MD forests, inferring that the dependence of basal area on the scale transect does not vary much with the change in forest type. We are not referring to the value of the property ‘basal area’, but to the spatial dependence of the property. The FD versus scale plots show similarities as well; one notable evidence is that all the three basal area plots cross FD = 2, at two units of scale, which suggest that these values are homogenous quadrats. Semivariograms (Figure 2) of density and height for a particular forest type suggest that both properties vary similarly with the scale transects. This similarity is less when compared to the above-stated cases, especially at higher values of scale.

The suggested quadrant sizes for height and basal area are different implying that even if the particular forest type is defined by these characteristics, they do not vary in the same way with change scale. In a mathematical approach, one can use average values of quadrat sizes suggested for basal area and height.

The fact that FD can be used as a measure of the extent to which a variable is dependent on another, is key for studying variation in properties of vegetation samples along with the fact that FD is not a constant function of scale. One of the primary assumptions of methods of classical inferential statistics is that the samples should be randomly placed or have a homogenous distribution. Selection of quadrat size so that the distribution is homogenous within the quadrat is important to eliminate the pseudo random samples, in order to improve the generality and accuracy of results. FD of 2, which occurs at the suggested value of scale or quadrat size, has a homogenous distribution. By observing the variation of FD along the scale for different properties one can deduce correlations between the spatial dependence of these properties.

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>Appropriate quadrant size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basal area</td>
</tr>
<tr>
<td>Moist-deciduous</td>
<td>2.05</td>
</tr>
<tr>
<td>Semi-evergreen</td>
<td>2.05</td>
</tr>
<tr>
<td>Evergreen</td>
<td>2.00</td>
</tr>
<tr>
<td>Average for properties</td>
<td>2.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2.</th>
<th>Ideal sampling scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot</td>
<td>A</td>
</tr>
<tr>
<td>Evergreen → Height</td>
<td>1.42</td>
</tr>
<tr>
<td>Semi-evergreen → Basal area</td>
<td>1.4</td>
</tr>
<tr>
<td>Semi-evergreen → Height</td>
<td>1.08</td>
</tr>
<tr>
<td>Moist-deciduous → Basal area</td>
<td>1.55</td>
</tr>
<tr>
<td>Moist-deciduous → Density</td>
<td>1.02</td>
</tr>
</tbody>
</table>


ACKNOWLEDGEMENTS. We have used field inventory data of Andaman vegetation collected under the project ‘Biodiversity characterization of Andaman and Nicobar islands using GIS and remote sensing’. We thank Dr Sudhakar Reddy, Scientist, National Remote Sensing Centre, Hyderabad, for help during field data collection. We also thank Dr Anupam Chaturvedi, without whose efforts this work would not have been completed in time.

Received 12 April 2013; revised accepted 12 September 2013

CURRENT SCIENCE, VOL. 105, NO. 9, 10 NOVEMBER 2013 1281