

## EFFECTS OF DATA QUALITY ON ANALYSIS OF ECOLOGICAL PATTERN USING THE $\hat{K}(d)$ STATISTICAL FUNCTION

ELIZABETH A. FREEMAN<sup>1</sup> AND E. DAVID FORD<sup>2,3</sup>

<sup>1</sup>Interdisciplinary Graduate Program in Quantitative Ecology and Resource Management, Box 351720,  
University of Washington, Seattle, Washington 98195-1720 USA

<sup>2</sup>College of Forest Resources, Box 352100, and Interdisciplinary Graduate Program in Quantitative Ecology and Resource  
Management, University of Washington, Seattle, Washington 98195-2100 USA

**Abstract.** The  $\hat{K}(d)$  function is a summary statistic of all plant–plant distances in a mapped area. It offers the potential for detecting both different types and scales of patterns in a single map. Two types of errors occur in maps of individual plants. Data management errors, caused by transcription errors or other mishandling, are large errors and apply to small numbers of plants. Measurement errors, caused by the mapping techniques and equipment, are small errors that apply to all plants. Simulation of known spatial patterns combined with increasing levels of both types of error showed that: (1) data management errors cause the spatial patterns identified by the statistical function  $\hat{K}(d)$  to become less significant but do not cause a shift in scale of the identified patterns; and (2) measurement errors caused the spatial patterns identified by  $\hat{K}(d)$  to become less significant and to shift to larger scales. The effects of measurement errors are inversely proportional to the scale of interaction between plants on the map. Detection of inhibition between points is more sensitive to measurement error than detection of clustering; detection of small clusters is more sensitive than detection of large clusters; and measurement error tends to cause an overestimation of clumping size. For patterns with inhibition, estimating minimum establishment distance is more sensitive to error than the maximum distance at which inhibition affects survival probability.

Two examples of tree spatial distributions from the Wind River Canopy Crane Research Facility stem map data set were analyzed using the  $\hat{K}(d)$  function. Clusters of *Thuja plicata* were detected and were much larger than levels of mapping error identified in the data. Significant inhibition occurs between large (dbh  $\geq 20$  cm) trees of all species at a scale much greater than the level of mapping error. However, the minimum distance of significant inhibition (i.e., the distance within which neighbors are never found) was on the order of the mapping error. Accurate identification of inhibition may not be possible using  $\hat{K}(d)$ .

**Key words:** aggregation; clumping; data quality; error analysis; inhibition;  $K(d)$ ; point processes; regularity; spatial pattern; spatial statistics; stem map; *Thuja plicata*.

### INTRODUCTION

Spatial statistics calculated from measured distances between individual plants are increasingly used when studying interactions between plants and when investigating the structure of ecological communities. Single valued statistics reduce the entire spatial pattern to a single number to detect clumping or spatial inhibition. These statistics are based on nearest neighbor distances (Clark and Evans 1954, Pielou 1962, Donnelly 1978) or sample point to nearest neighbor distances (Hopkins 1954, Pielou 1959). From these statistics, it was proposed that single species forest stands (Laesele 1965) and desert vegetation (Phillips and MacMahon 1981) shift over time from clumped patterns of seedlings and young plants towards regularity due to competition.

However, single value statistics are limited. They

give no information on the spatial scales of patterns, nor do they provide information such as size of clusters or the distance of strongest inhibition. Most importantly they cannot define mixed patterns (Pielou 1962; e.g., inhibition at small scales within larger scale clumping). Increasingly the  $\hat{K}(d)$  function (Ripley 1977, Diggle 1983, Haase 1995) is being used for these purposes.  $\hat{K}(d)$  is a transformation of all interplant distances that is graphed against  $d$ . Software (e.g., MathSoft 1996) is available for calculating  $\hat{K}(d)$ . The basic assumptions of  $\hat{K}(d)$  analysis are that: (1) the individual plant is the elementary sampling unit and defines *grain size* (Legendre and Legendre 1998:708–709); and (2) the size of the mapped area defines the *extent* of the sample.  $\hat{K}(d)$  describes the distribution of *intervals* between the plants on the sample area and, unless otherwise mentioned, *scale* refers to the size of these intervals.

Ecologists using  $\hat{K}(d)$ , or its square root transformation,  $\hat{L}(d)$  (Besag 1977, Diggle 1983) have found more complex spatial patterns than simply clumped,

Manuscript received 7 October 1999; revised 10 November 2000; accepted 14 December 2000; final version received 9 March 2001.

<sup>3</sup> Address correspondence to this author.  
E-mail: edford@u.washington.edu.

random, or regular, and have used these results to make inferences about pattern forming processes. For example, Sterner et al. (1986) measured clumping in juveniles of three tropical tree species. The spatial pattern of adults was significantly more uniform than could be explained by random thinning of the juvenile clumps. Sterner et al. suggested that this pattern of mortality is consistent with the Connell–Janzen hypothesis for the maintenance of high species richness in tropical forests (Connell 1970, Janzen 1970). Kenkel (1988) used  $\hat{L}(d)$  analysis, along with a refined nearest neighbor analysis, to study self-thinning in pure stands of *Pinus banksiana*. Based on changes in spatial pattern and stand density with increasing stand age, he suggested that two distinct competitive phases occurred: an early scramble phase of two-sided competition for soil resources, and a later one-sided competition for light. In both the Sterner et al. (1986) and Kenkel (1988) analyses, the identification of regularity within the spatial pattern indicating competition was crucial to the ecological interpretations.

$\hat{L}(d)$  is also being used in ecological modeling to define the structure of the model and to estimate parameter values. See Moer (1993, 1997) for old-growth forest, Kenkel (1993) for the clonal herb *Aralia nudicaulis*, and Batista and Maguire (1998) for regenerating tropical forest. A crucial feature of each model is estimation of spatial inhibition.

For detailed analysis of spatial pattern, one needs to know the accuracy of  $\hat{L}(d)$ . This accuracy depends upon the accuracy of the mapped data. Several mapping techniques have been reported for stands of trees. To calculate tree positions Kenkel (1988) first established a rectangular grid and then measured the distribution of individuals within each grid segment to each of its corners. Kenkel (1993) used a similar procedure for mapping individuals of the herb *A. nudicaulis* in  $5 \times 5$  m plots. Moer (1993) cruised through sample plots by establishing individual reference points from which tree azimuths and distances were surveyed.

Research into developing field mapping techniques for individual plants has focussed on attaining a balance between the speed of a survey and the accuracy of the map obtained. Rohlf and Archie (1978) proposed a rapid technique based on progressive mapping of individuals relative to previously mapped trees, without establishing and surveying plot corners. Hall (1991) checked this procedure, using both a grid method and by simulation techniques, and concluded that propagation of errors was a serious problem and that locating errors can be nearly impossible. Boose et al. (1998) reported improvements on this method and found that errors in tree locations depended on how accurately the distances were measured between trees and on the number of trees on the plot.

In this paper we investigate how the type and magnitude of errors that might be acceptable depend upon both the type and scale of the pattern of interest. We

examine data management errors and measurement errors and calculate their effects on the identification of spatial patterns using  $\hat{L}(d)$  as well as their effects on detecting and defining cluster and inhibition patterns. We use both simulations and data examples from an old-growth *Pseudotsuga menziesii*–*Tsuga heterophylla* forest (Franklin and Dyrness 1988).

## METHODS

### *Strategy of investigation using simulation*

We simulated data points in order to specify both clustered and inhibition patterns. We modified individual points to simulate mapping errors. By comparing  $\hat{L}(d)$  from the modified data points to  $\hat{L}(d)$  from the original data points the effects of the simulated errors on the  $\hat{L}(d)$  transformation of  $\hat{K}(d)$  were examined. Both data management errors and measurement errors were found in the Wind River Canopy Crane Research Facility (WRCCRF) stem map data set (Freeman 1997). These were simulated, respectively, by replacement of points and by moving points specific distances according to the error. Results are given both as a summary of all the simulations, and by a single example of each of the four combinations of spatial patterns (clustered vs. inhibited) and error type (measurement error vs. data management error). The complete set of simulations of different types of pattern combined with different amounts of error is presented in Freeman (1997).

### *Simulated spatial patterns*

The simulated plot was a unit square of area  $A$ , with  $n = 100$  points. A toroidal wrap-around of the square boundaries was used so that edge effects were not a concern.

*Poisson cluster process.*—In a Poisson cluster process, parent events are distributed randomly each forming a cluster center assigned a random number of  $n$  offspring. Offspring are spatially distributed around their parent with a bivariate normal distribution with standard deviation  $\sigma_{\text{cluster}}$ . A range of cluster sizes was investigated:  $\sigma_{\text{cluster}} = 0.025, 0.05, 0.075, \text{ and } 0.10$ . On the  $200 \times 200$  m Wind River Canopy Crane Research Facility (WRCCRF) stem map this was equivalent to  $\sigma_{\text{cluster}} = 5 \text{ m}, 10 \text{ m}, 15 \text{ m}, \text{ and } 20 \text{ m}$ . Note that  $\sigma_{\text{cluster}}$  is the standard deviation of the distance between each offspring and its parent, so cluster size is  $\sim 2\sigma_{\text{cluster}}$ .

*Markov point inhibition process.*—A Markov point inhibition process assumes that only its potential neighbors affect the probability that a new point can become established. There is a cutoff distance,  $a$ , where, if any potential neighbor is closer than this distance then establishment is prohibited. There is a large distance,  $b$ , defining the outer limit of a plant's inhibitory effect. In this study, the probability of establishment, given a neighbor between the distances  $a$  and  $b$ , was chosen to be a linear function of distance. The minimum establishment distance was set at  $a = 0.005$  and the maxi-

mum distance of inhibition varied over  $b = 0.05, 0.10, 0.15,$  and  $0.20$ . On the WRCCRF stem map this was equivalent to  $a = 1$  m and  $b = 10$  m, 20 m, 30 m, and 40 m.

*Complete spatial randomness.*—Complete spatial randomness (CSR) was used, as an aid in the simulation procedures, as part of the error simulation, and as a null hypothesis in Monte Carlo tests. CSR assumes the mean number of points in any region  $A$  with area  $|A|$  follows a Poisson distribution with mean  $\mu = \lambda \times |A|$ , and given  $n$  points in  $A$ , the points are an independent random sample from the uniform distribution on  $A$  (Diggle 1983).

#### *The two types of errors*

*Data management errors.*—These include errors made in the field (e.g., measuring the wrong plant or misidentifying a plant) as well as errors in data handling (e.g., incorrectly transcribing data, where 314.2 degrees azimuth may be recorded as 312.4 degrees or as 134.2 degrees). Data management errors can be any size, but generally affect few data points.

*Measurement errors.*—These are inherent in mapping techniques and equipment. Their magnitude depends on the equipment used. A tree map made with a hand held compass and a measuring tape is likely to have greater absolute measurement errors than one made with a tripod based electronic theodolite and laser range finder. Mapping techniques also influence the magnitude of these errors. If a grid is established first, and individuals are mapped relative to the grid points, cumulative errors are minimized. If individuals are mapped sequentially, each relative to previously mapped individuals, large errors can accumulate (Hall 1991), although Boose et al. (1998) describe a procedure and calculation method for minimizing errors. Generally, measurement errors are small when a grid is used, but they can potentially affect all data points.

#### *A note on absolute vs. relative errors*

Error analysis of a mapped data set reveals absolute levels of error (e.g., the number of displaced points and the limits to mapping accuracy of correctly placed points). In this study, however, we consider relative error. For data management error, relative error is the number of displaced points relative to the total number of trees. This is straightforward to calculate, as it does not depend on the underlying, and possibly unknown, spatial pattern. For measurement error, relative error is the limit of mapping accuracy of correctly placed points, relative to the scale of the underlying spatial pattern. In simulations, the scale of the pattern is known, and thus the relative error can be calculated. For actual stem maps, the underlying scale is unknown; after all, discovering this scale may be the purpose of making the map. However, the measurement error can still be calculated relative to the scale of interest. For example, when investigating inhibition processes, the

interactions of interest are on a smaller scale than when examining cluster processes. Therefore, any mapping errors will be relatively larger, compared to small-scale patterns. Note that in distributions with more than one parameter, such as in Markov point processes, the error can be calculated relative to each parameter.

Examples of absolute error levels would be: “in the preliminary survey data management errors occurred in 11 out of 253 trees,” or “the measurement errors have a standard deviation of 0.43 m.” Relative error levels would be: “4.3% of the points are subject to data management errors,” or “the standard deviation of the measurement error is 67% of the predicted inhibition distance.”

#### *Simulated errors*

*Data management errors.*—Data management errors were simulated by creating large errors in location in a percentage of the data points. These points were randomly selected, and the true points were replaced with randomly located points to simulate poor data management. This is equivalent to a few plants of the wrong species being included in a map, or some of the mapping data being misrecorded or mistranscribed. The percentage of replacements was set to 0, 5, 10, and 20. This range is likely to be wider than most data management errors where the objects can be clearly identified and mapped (e.g., individual trees in a single species stand), but it was used to cover situations where data management can be difficult.

Primarily we used random points for replacement. There are many possible combinations of original distribution and replacement distribution, and examining all of them was prohibitive. We did, however, investigate two combinations: clustered distributions with replacement from an inhibited distribution with  $a = 0.005$  and  $b = 0.20$ ; and inhibited distributions with replacement from a clustered distribution with  $\sigma_{\text{cluster}} = 0.025$ . This is equivalent to mistakenly including a few trees of the wrong species, with a different spatial pattern, in a map.

*Measurement errors.*—Measurement errors were simulated as small errors in location applied to all data points. Each point was shifted a small distance based on a bivariate normal distribution centered on the point, with a standard deviation  $\sigma_{\text{error}} = 0, 0.025, 0.050,$  and  $0.100$ . Translating from the simulation to a square the size of the WRCCRF stem map (200 m on a side), this was equivalent to absolute errors of  $\sigma_{\text{error}} = 0$  m, 5 m, 10 m, and 20 m. These were several orders of magnitude larger than the measurement errors in the WRCCRF stem map. However, errors of several meters may be realistic for a tree stem map of a large area made with a tape measure and hand compass without using a fixed grid system. Also, in the investigation of grasslands, for example, absolute errors may be low due the small distances between individuals, while relative error may be high.

Simulations were made on the unit square with  $\sigma_{\text{error}}$ , the error relative to this unit square. The measurement error relative to the scale of the underlying pattern is the ratio of  $\sigma_{\text{error}}$  to the scale of the pattern. For clustered patterns, cluster size was  $2\sigma_{\text{cluster}}$ . This is because spatial analysis with  $\hat{K}(d)$  is related to the diameter of the clusters, and  $\sigma_{\text{cluster}}$  is related to the radius of the clusters. Inhibited patterns were generated with a Markov point process, with parameters  $a$  and  $b$ , and relative error was calculated for each parameter. We made 100 simulations for each combination of parameter value used in defining pattern with either replacement rate for data management errors or  $\sigma_{\text{error}}$  for measurement error.

#### *Spatial analysis of simulated data*

*The  $\hat{K}(d)$  statistic.*— $\hat{K}(d)$ , a second order parameter, considers all distances between all individual points, not just the distances between nearest neighbors. Consequently, it gives more information on the scale of the patterns than do first order tests such as  $G$  (the cumulative frequency distribution of nearest neighbor distances) or  $F$  (the cumulative frequency distribution of arbitrary points to nearest neighbors; see Diggle 1983).

$K(d)$  is the expected number of points within a given distance of an arbitrary point. This is calculated relative to the density of points:

$$K(d) = [E(\text{number of points within distance } d \text{ of an arbitrary point})]/(\text{the density of points})$$

where  $d$  = distance. Estimation is by

$$\hat{K}(d) = \frac{(\text{area of the plot})(\text{number of distances less than } d)}{(\text{total number of points})^2}$$

$L(d)$  is a square root transformation of  $K(d)$  suggested by Besag (1977):

$$L(d) = \left( \frac{K(d)}{\pi} \right)^{1/2} - d.$$

$L(d)$  linearizes  $K(d)$ , stabilizes the variance, and under complete spatial randomness the expected value of  $L(d)$  is approximately zero. It can be difficult to interpret  $K(d)$  visually. Using  $L(d)$  makes graphs easier to read. Examples for different types of pattern can be found in Diggle (1983) and practical application in Kenkel (1993) and Moer (1993).

$\hat{L}(d)$  is calculated for a sequence of distances. As  $d$  approaches the size of  $A$ , the results of  $\hat{L}(d)$  reflect edge effects rather than the spatial patterns of the data (Haase 1995). In this study,  $\hat{L}(d)$  was calculated at one hundred distances, evenly spaced between  $d = 0$  and  $d = 0.5$ .

For simulated inhibited patterns, analysis was carried

out on the distances between  $d = 0$  and  $d = 0.3$ . The effect of inhibition on  $\hat{L}(d)$  is primarily governed by interactions between nearby plants, while the effect of clustering on  $\hat{L}(d)$  is governed by size of the clusters. Thus inhibition is shown by  $\hat{L}(d)$  at smaller scales than is clustering (Fig. 1).

The results of  $\hat{L}(d)$  are then plotted against distance. Significance is usually evaluated by comparing the observed data with Monte Carlo envelopes from multiple simulations of the null distribution, usually complete spatial randomness (CSR). Each simulation generates an  $\hat{L}(d)$  function (Fig. 2a). The repeated series of simulations each generate a further  $\hat{L}(d)$  function and a probability envelope is calculated from their highest and lowest values (Fig. 2d). Typically an envelope is calculated from the  $\hat{L}(d)$  of 99 simulations and if the  $\hat{L}(d)$  of a pattern has some part outside of that envelope it is judged to be a significant departure from CSR.

Two points had to be considered in using Monte Carlo methods to assess the result of our simulations of the interaction between pattern and error. First, the Monte Carlo envelope typically has irregularities in the bounds it defines (Fig. 2d), and repeat constructions produce similar but not identical envelopes; therefore, the envelope provides a guide for assessment but not a strict test. Because of this it is inappropriate to generate a separate Monte Carlo  $\hat{L}(d)$  envelope for each simulation and expect to use each one as a rigorous arbiter of significant departure from CSR or not. As a general guide for the 100-point distributions, 99 simulations gave values of  $\hat{L}(d)$  within the limit of +0.01 to -0.01 (Fig. 2d).

Second, and particularly important for this work, errors may not only change the magnitude of the pattern, but also shift the scale. Assessment of error must be made through analysis of multiple simulations of the interaction between pattern and error, both of them stochastic processes. One hundred simulations were made of each combination. The mean point was calculated for these simulations at which  $\hat{L}(d)$  reached its extreme value (Mean Extreme Value or MEV). This MEV is a maximum for clustered patterns (Fig. 3a), and a minimum for inhibited patterns (Fig. 3b).

#### *Spatial analysis of forest data*

The WRCCRF was established in the Thornton T. Munger Research Natural Area, in southern Washington state near the Columbia River Gorge (45°49'13.76" N, 121°57'6.88" W) to give researchers access to the upper canopies of a low elevation old-growth Douglas fir–Western hemlock forest with canopy trees ~500 yr old. DeBell and Franklin (1987) and Franklin and DeBell (1988) provide background information on stand dynamics at this site. A 4-ha square with edges running north–south and east–west, and with the crane at its center point, contained 2168 trees with dbh  $\geq$  5 cm, including snags, divided among nine species. The most common species were Western hemlock *Tsuga*

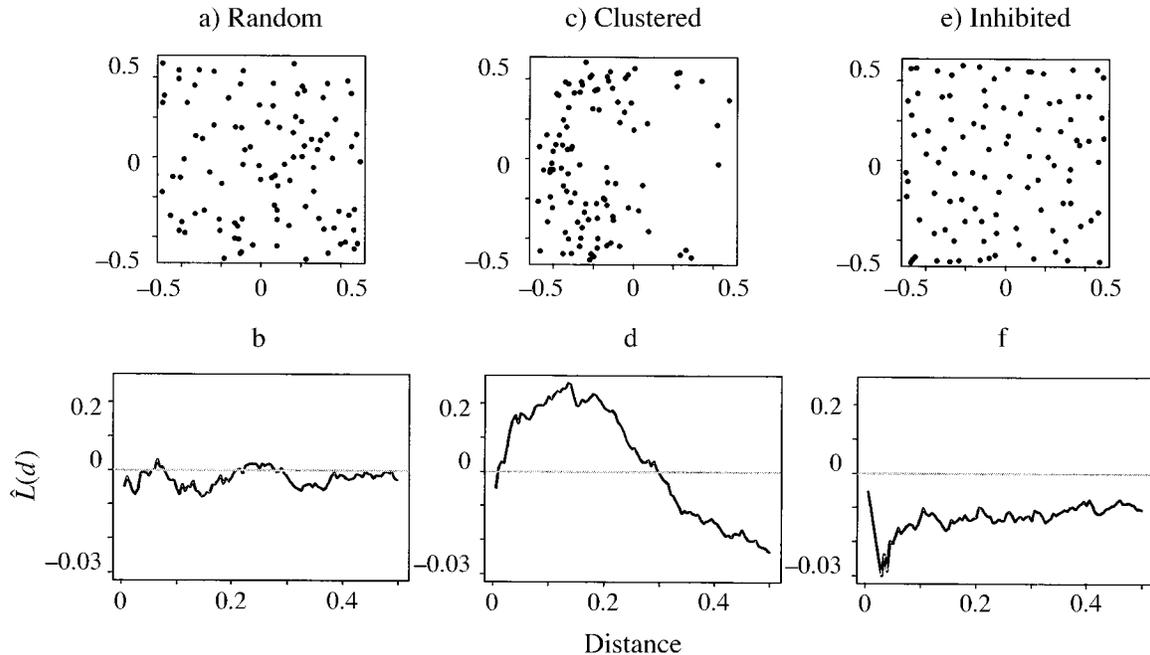


FIG. 1. Examples of simulated point patterns on the unit square and their corresponding calculated  $\hat{L}(d)$  functions: (a, b) random, (c, d) clustered, and (e, f) inhibited.

*heterophylla* ((Raf.) Sarg.), Pacific yew *Taxus brevifolia* (Nutt.), Douglas fir *Pseudotsuga menziesii* ((Mirb.) Franco), and Western red cedar *Thuja plicata* (Donn ex. D. Don).

Professional surveyors surveyed a 25-m grid on the 4-ha study site. All trees with dbh  $\geq 5$  cm were numbered and tagged, and species and dbh were recorded. Tree locations were mapped from surrounding grid points using a Criterion 400 Survey Laser (Laser Technologies, Englewood, Colorado, USA) to measure azimuth angle, horizontal distance, and inclination.

Eighty-five percent of the trees were measured from two grid points. In some cases only one measurement set was possible due to heavy underbrush or other trees. Measurements were converted to  $X$  and  $Y$  coordinates. To estimate the level of measurement errors, 37 trees located in two grid squares were each mapped from three grid points.

Measuring from two grid points increased the time involved in collecting the data. It did not make a great difference to the standard deviation of the measurement errors, but it did reveal data management errors. Errors

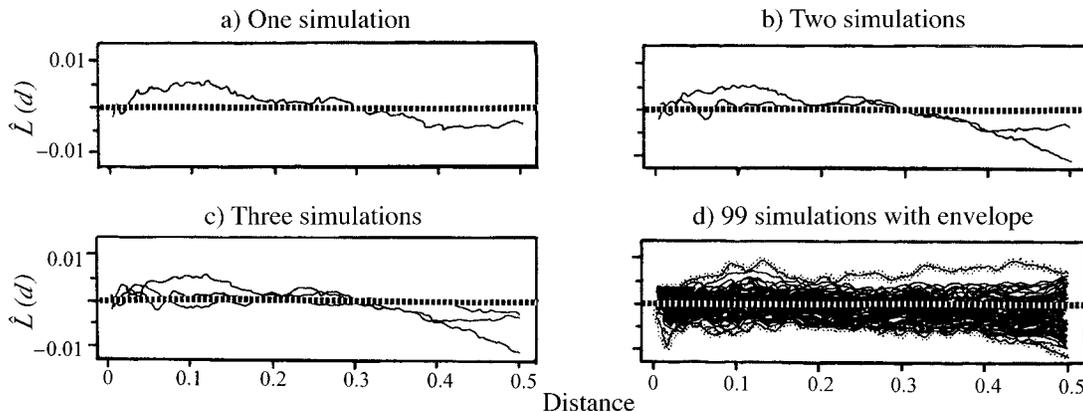


FIG. 2. Procedure for generating a Monte Carlo envelope for  $\hat{L}(d)$ . (a) One simulation of complete spatial randomness (CSR) for  $n$  points. (b) A second simulation of CSR for  $n$  points added to the first simulation. (c) A third simulation of CSR for  $n$  points added. (d) Ninety-nine simulations of CSR and the resulting Monte Carlo envelope (stippled line). Typically a test is constructed by considering data as the 100th case and examining whether it exceeds the envelope at any scale.

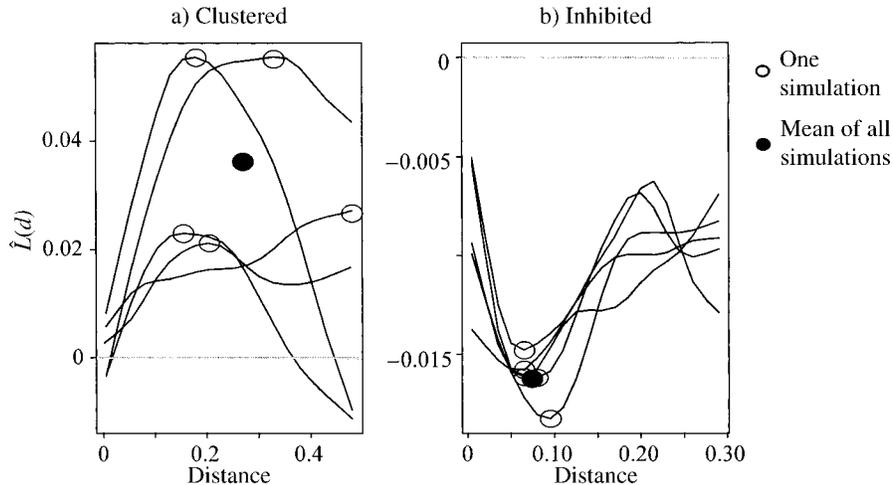


FIG. 3. Calculation of Mean Extreme Value for sets of clustered or inhibited patterns generated by the same process. (a) For clustered patterns, the maximum values of each simulation (open circles) are averaged to find the mean of the maximum values (solid circle). (b) For inhibited patterns, the minimum values of each simulation (open circles) are averaged to find the mean of the minimum values (solid circle).

for 98 trees (4.5%) were caught in this manner when locations that were calculated from two surveyed positions differed substantially.

By assuming that measurement error followed a bivariate normal distribution, with no correlation between position of the point and size of the error, the within tree standard deviation was  $\sigma = 0.123$  m when trees were mapped from at least two grid points. Points mapped from the mean of two measurements would follow a bivariate normal distribution with  $\sigma_{\text{error}} = 0.087$  m.

One theory of competition states that the distribution of plants becomes more spatially even over time, since plants growing close together are subject to increased competition, and thus to increased mortality (Ford 1975, Sterner et al. 1986, Kenkel 1988, 1993, Kenkel et al. 1997). In this forest, the older, and thus larger, trees are assumed to be survivors of such competition, and thus would be more likely to show the effects of inhibition as found by Moer (1993) in an old-growth coniferous forest. The WRCCRF has a policy of strictly limiting destructive sampling, and cores to age trees could not be taken.

## RESULTS

### *The effects of errors on the mean extreme values*

**Data management error.**—Generally the detection of both clustering and inhibition using  $\hat{L}(d)$  is robust to data management error. The addition of simulated data management errors, drawn from a complete spatial randomness (CSR) distribution, to clustered patterns reduced the amplitude of the maximum values of  $\hat{L}(d)$  but did not shift the maximum to a larger scale (Fig. 4). Clustering was still detected even when 20% of the points had been replaced from CSR (i.e.,  $\hat{L}(d)$  at the

distance of 0.10 was much greater than 0.01). Similarly, simulated data management errors added to inhibited patterns made minimum values of  $\hat{L}(d)$  less extreme, though they remained much less than  $-0.01$ , and shifted them slightly to larger scales (Fig. 5). Clustered patterns with inhibited replacement showed little difference from CSR replacement at replacement rates of up to 20%. The inhibited patterns with clustered replacement were affected more strongly than with CSR replacement, but only after replacement rate reached 20%.

The Mean Extreme Value (MEV) of  $\hat{L}(d)$  was calculated for 100 simulations of each of four clustered and four inhibited patterns, each combined with four levels of simulated data management errors from a CSR distribution. The change in MEV with increasing simulated error for both clustered (Fig. 6a) and inhibited (Fig. 6b) patterns was generally small relative to that found for measurement error (Fig. 6c, d). Clustered patterns decreased in the absolute value of the MEV, but the value was always substantially  $>0.01$ , and no consistent change in the estimated scale of pattern was shown. For inhibited patterns the absolute value of the MEV increased but did not reach  $-0.01$ . With inhibited patterns, the MEV did show a slight shift to larger scales, but not nearly to the extent that they shifted with measurement errors.

**Measurement error.**—The addition of measurement errors over the range of simulated patterns had substantially greater effects than data management errors. The apparent size of clusters increased (Fig. 7a, c, e, and g) and maximum values of  $\hat{L}(d)$  decreased, becoming closer to CSR, and shifting to larger scales (Fig. 7b, d, f, and h). The shift in scale was present in all four clustered patterns (Fig. 6). It was more obvious,

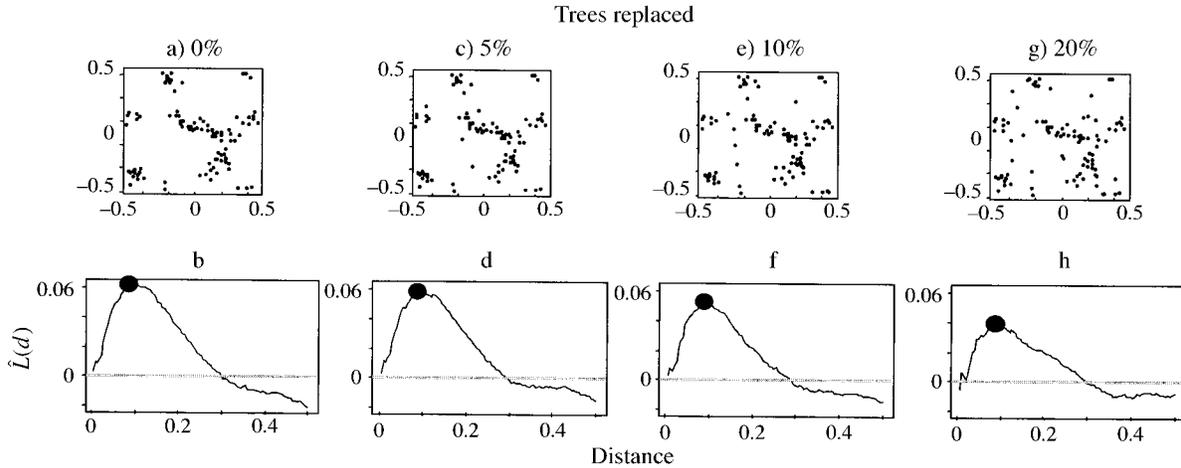


FIG. 4. Data management error applied to a clustered pattern ( $\sigma_{\text{cluster}} = 0.10$ ), illustrating how increasing data management errors cause the corresponding maximum value of  $\hat{L}(d)$  to lose amplitude, but not to move to larger distances. Solid circles = maximum  $\hat{L}(d)$ ; panel (a) corresponds with (b), (c) with (d), and so forth. Calculations are based on the unit square.

and showed up at smaller levels of error, in the distributions with smaller clusters,  $\sigma_{\text{cluster}} = 0.025, 0.05$ , as the level of error relative to the cluster sizes was larger in these distributions. The shift was most pronounced when  $\sigma_{\text{error}}$ , the standard deviation of the measurement error, was  $\geq 2\sigma_{\text{cluster}}$ , the scale of the clusters (i.e., when the error relative to the cluster size was  $\geq 100\%$ ; see Fig. 6c).

For addition of measurement error to an inhibited distribution the minimum values of  $\hat{L}(d)$  increased, to become closer to CSR, and shifted to larger scales (Fig. 8). Note that when looking only at the MEV, it would appear that measurement error has a greater effect when  $b$  (the maximum distance at which establishment is affected by neighbors) is large (i.e., when inhibition is occurring at greater distances). However this is partly

an artifact of the summarizing process. MEV provides a useful summary but is not appropriate when the level of error completely obscures the underlying pattern (i.e., once the apparent distribution resembles CSR). When the distribution approaches CSR, MEV lies at the midpoint of the distances examined. Thus in this case (where  $\hat{L}(d)$  was calculated from  $d = 0-0.3$ ), MEV for distributions resembling CSR tended to be located near  $d = 0.15$  (Fig. 6d). As a result, for the smallest radius of inhibition influence tested (i.e.,  $b = 0.05$ ), as measurement errors increased to a scale larger than  $b$  (i.e., from  $\sigma_{\text{error}} = 0.05$  to  $\sigma_{\text{error}} = 0.10$ ), MEV actually reversed its trend, and shifted slightly towards smaller scales. Thus when examining the effect of measurement errors on inhibited distributions, the most sensitive combination of error and distribution types, it was nec-

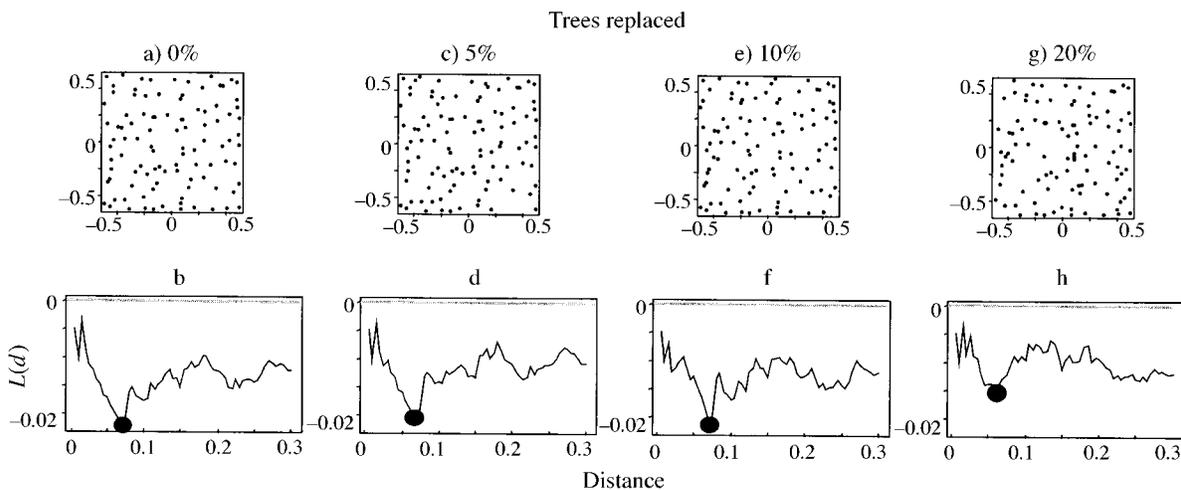


FIG. 5. Data management error applied to an inhibited pattern generated as a Markov point inhibition process with model parameters  $a = 0.005, b = 0.20$ . Data management errors cause the minimum value of  $\hat{L}(d)$  to lose amplitude, but do not move to larger distances. Solid circles = minimum  $\hat{L}(d)$ ; panel (a) corresponds with (b), (c) with (d), and so forth. Calculations are based on the unit square.

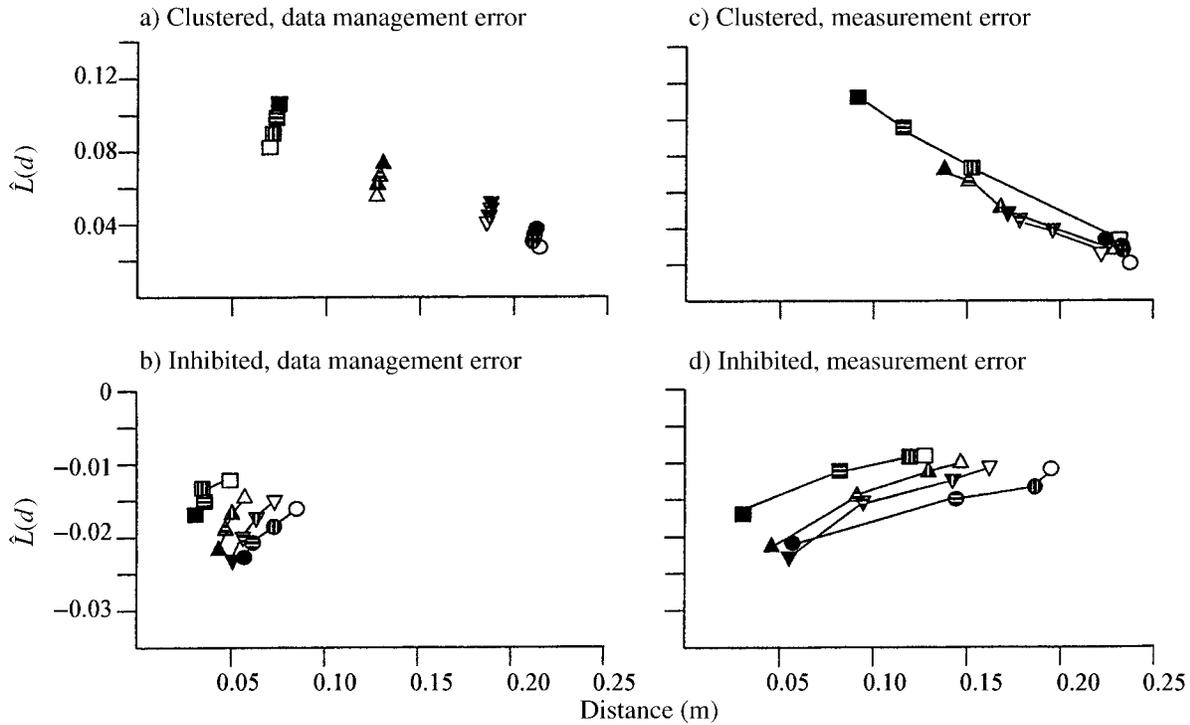


FIG. 6. Effects on Mean Extreme Value of  $\hat{L}(d)$  calculated from 100 simulations of addition of (a, b) data management error or (c, d) measurement error, on (a, c) clustered and (b, d) inhibited patterns. For clustered patterns,  $\sigma_{\text{cluster}}$  was varied: 0.05 (■), 0.10 (▲), 0.15 (▼), or 0.20 (●). For inhibited patterns the parameter  $a$  of the Markov point process was held constant at 0.005, while parameter  $b$  varied: 0.05 (■), 0.10 (▲), 0.15 (▼), or 0.20 (●). Data management error was simulated by replacing a percentage of points using complete spatial randomness (CSR): solid symbols, no change; horizontal hatching, 5% change; vertical hatching, 15% change; open symbols, 20% change. Measurement error was simulated by addition of error from a bivariate normal distribution with  $\sigma_{\text{error}}$  to all points: solid symbols, no change; horizontal hatching, 0.025; vertical hatching, 0.05; open symbols, 0.10.

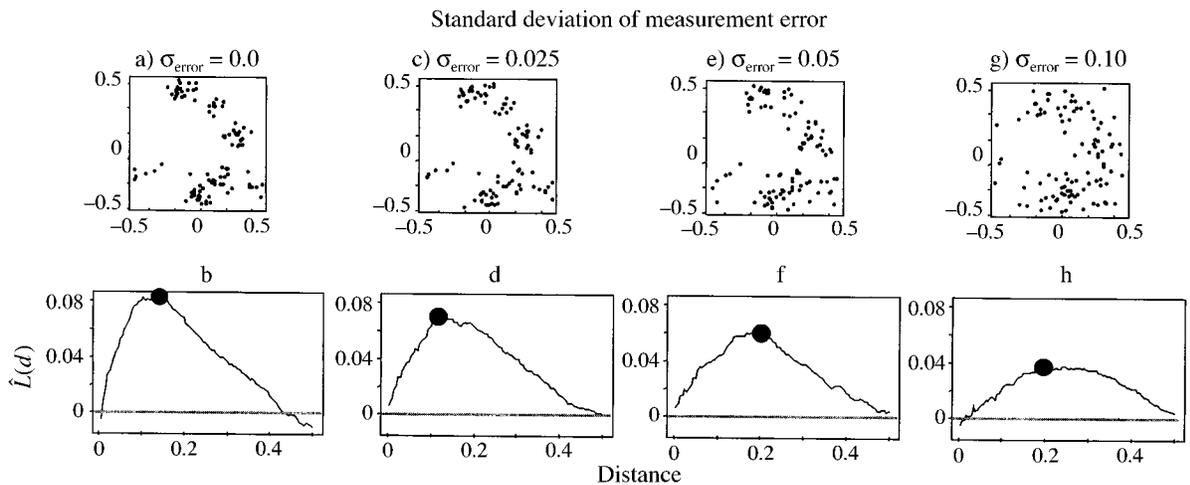


FIG. 7. Measurement error applied to a clustered pattern ( $\sigma_{\text{cluster}} = 0.10$ ), illustrating how increasing data measurement errors cause the corresponding maximum value of  $\hat{L}(d)$  to lose amplitude and move to larger distances. Solid circles = maximum  $\hat{L}(d)$ . Calculations are based on the unit square.

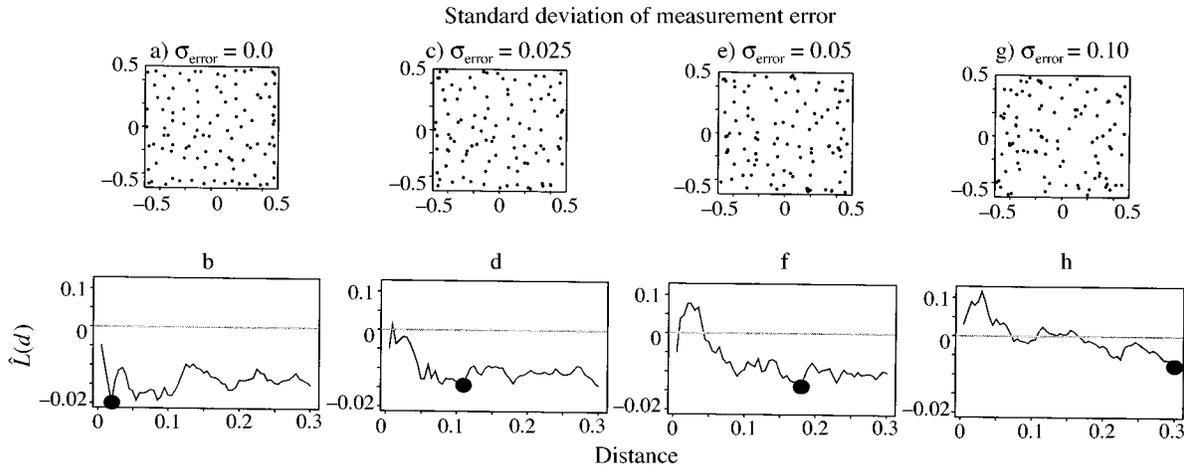


FIG. 8. Measurement error applied to an inhibited pattern generated as a Markov point inhibition process with model parameters  $a = 0.005$ ,  $b = 0.20$ . Measurement errors cause the minimum value of  $\hat{L}(d)$  to lose amplitude and move to larger distances. Solid circles = minimum  $\hat{L}(d)$ . Calculations are based on the unit square.

essary to look at the unsummarized data, an example of which is shown in Fig. 8.

When  $b$  is small, measurement errors are larger relative to the scale of the inhibition, and thus have a greater effect on  $\hat{L}(d)$ . Once  $\sigma_{\text{error}}$  (the standard deviation of the measurement error) approaches  $b$  (i.e., the measurement error relative to  $b$  approaches 100%) it is impossible to distinguish between the mapped pattern and CSR. Even at levels of  $\sigma_{\text{error}}$  as small as 20% of  $b$ , the regular pattern resulting from an inhibition process is more sensitive to the effects of error than clustered patterns. Often  $\hat{L}(d)$  will no longer indicate inhibition at small distances, while still showing inhibition at larger scales. When looking at small distances, the relative error is larger, and thus errors tend to affect small distances first.

Generally detection, and particularly determination of scale, of both clustered and inhibited processes using  $\hat{L}(d)$  is less robust to measurement error than to data management error. Inhibition patterns and small scale clustering are more sensitive than large scale clustering.

#### Standard deviation of Mean Extreme Value (MEV)

The type of simulated error had little effect on the variance of the extreme value of  $\hat{L}(d)$ , or of the variance of the distance at which this extreme value occurred. The standard deviation of the MEV ranged from 0.001 to 0.002 for clustered distributions, and from 0.0002 to 0.0004 for inhibited distributions. These values are much smaller than the differences caused by either type of error. The standard deviation of the mean distance at which MEV occurred ranged from 0.002 to 0.008 for clustered distributions, and from 0.002 to 0.011 for inhibited distributions. These values are much smaller than the differences caused by measurement error, but on the order of the differences caused by data management error.

#### $\hat{K}(d)$ applied to the WRCCRF stem map

Spatial analysis of all nine canopy tree species at the WRCCRF was conducted (Freeman 1997). Here we choose two examples; live *Thuja plicata* for investigating clustered distributions, and all live trees  $>20$  cm dbh, for investigating inhibition patterns.

Clustering is the dominant spatial pattern of individuals of single species (Freeman 1997). Clustering may be the result of a parent-child process with small seed spread distances. Also some microclimate variations may favor a single species and produce clusters. We chose *T. plicata* for illustration of the potential effects of errors because of its reasonable sample size ( $n = 121$ ), and its lack of a visually obvious pattern in its tree map.

*Thuja plicata* shows significant clustering at medium to large scales. Living trees are clustered from  $\sim 10$  m to 60 m, with marginal clustering at smaller scales, and random distributions at larger scales (Fig. 9a). The cluster size is considerably greater than the level of measurement error (within tree standard deviation  $\sigma = 0.123$  m for trees located using one survey point,  $\sigma = 0.087$  m for trees located with two). Thus measurement error has little effect on detection or estimate of cluster size even if trees were mapped from only one grid point.

Live trees with dbh  $\geq 20$  cm show inhibition between 0.4 m and 13.2 m (Fig. 9b). The minimum intertree distance was 0.32 m. Levels of measurement error for this data have little effect on detecting the upper bound of the inhibition. However, in the simulations, detection of inhibition began to disappear with levels of relative measurement error as small as 20% and was entirely lost when the relative error was 100%. For the data of live trees of all species with dbh  $\geq 20$  cm inhibition occurring at distances  $< 0.087$ – $0.435$  m (1–5 times the measurement error) could be obscured by measurement

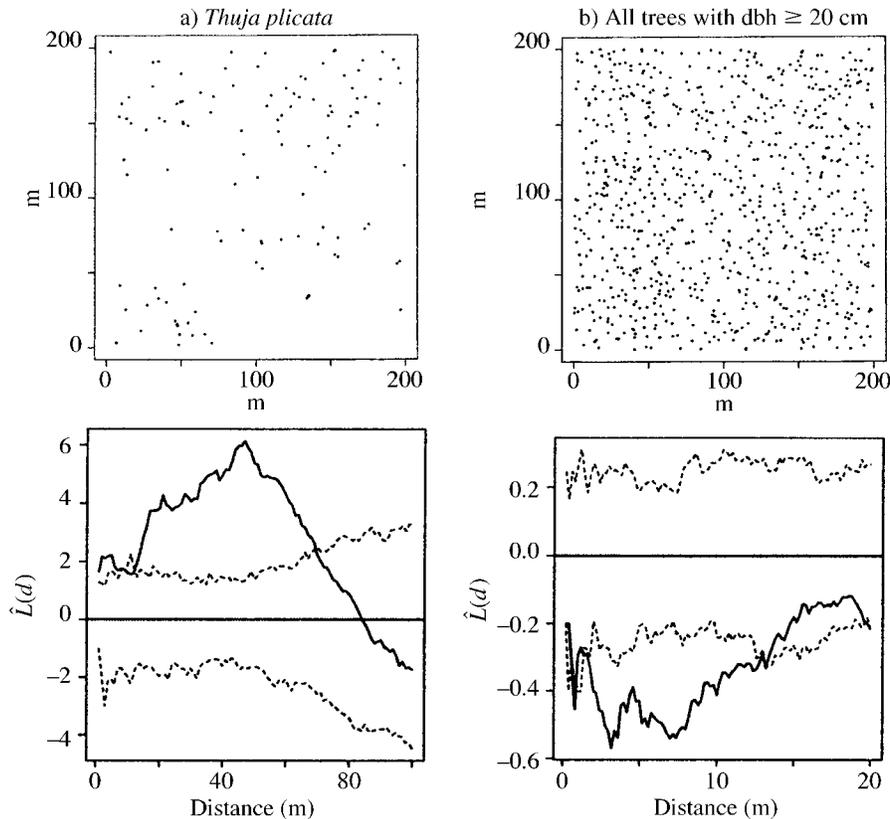


FIG. 9.  $\hat{L}(d)$  for subsets of the WRCCRF stem map: (a) Western red cedar, (b) all trees  $>20$  cm dbh. The subset by species shows significant clustering at medium to large scales (10–60 m), while the subset by size shows significant inhibition at small scales (0.4–13.2 m).

error. If a single grid point had been used for mapping tree locations, inhibition at distances  $<0.123$ – $0.615$  m would not have been detected. For this map,  $\hat{K}(d)$  can be used effectively to detect the outer range of competitive influence, *b*. But the minimum distance between trees with dbh  $\geq 20$  cm was 0.32 m. Thus the levels of error found reduced the accuracy of estimations of the minimum establishment distance. Furthermore, data of this quality would be unsuitable for detecting the onset of competition process among small trees such as clumps of young natural regeneration that grow at smaller neighbor distances.

#### DISCUSSION

*Data management error.*—Data management errors reduced the strength of identification of patterns by the  $\hat{L}(d)$  transformation of  $\hat{K}(d)$  but this type of error did not shift the identification of the scale of the patterns, even at relatively large error levels, such as 20% replacement. Data management error was simulated with a two-stage process: first, points were removed from the pattern; second, they were replaced with random locations. Clustered distributions are characterized by large numbers of small interpoint distances. Removing points from these patterns also removed some of these

small distances, causing clusters to become less dense, but the estimated size of the clusters was not affected. Random replacement added a few new large distances, but their contribution to the perceived scale of the clusters was negligible when compared to that of the original clusters. Inhibited distributions are characterized by the lack of small interpoint distances, and the removal of points has little effect on this characteristic. Random replacement added a few new small interpoint distances, and thus reduced the perceived strength of the inhibition, but again the scale did not shift.

Nonrandom replacement at rates  $<20\%$  showed little difference from random replacement. With replacement rates  $>20\%$ , pattern identification using  $\hat{K}(d)$  became less clear. With these higher replacement rates, inhibited distributions were more affected than clustered distributions. However if actual field data has this amount of data management error it is probably time to rethink the study. We only looked at two combinations of spatial pattern and error distributions. It is possible that other combinations, such as replacing points with the same pattern but at a different scale, may have greater effects. This should be kept in mind if this type of error may be significant in a study.

*Measurement error.*—Addition of measurement er-

rors reduced the amplitude of  $\hat{L}(d)$  and the MEV peaks to larger scales. This effect was strongest when the scale of errors approached the scale of the underlying pattern. Because of this (1) inhibition is more sensitive than clustering, (2) small clusters are more sensitive than large clusters, and (3) in inhibition, the minimum establishment distance was more sensitive than the maximum inhibition.

Both clustered and inhibited patterns showed an apparent shift to larger distances with the addition of measurement error. Inhibited patterns became indistinguishable from randomness at lower error levels than clustered patterns. In this study the inhibited patterns were defined by two parameters;  $a$ , the distance below which no nearest neighbor distances are found, and  $b$ , the maximum distance at which an inhibition effect occurs, with  $a$  smaller than  $b$ . While  $b$  was of the same scale as  $2\sigma_{\text{cluster}}$ ,  $a$  was much smaller. Thus error relative to  $a$  was much larger than error relative to  $b$  and  $2\sigma_{\text{cluster}}$ .

Also while measurement error caused peak maximum deviation of  $\hat{L}(d)$  from zero for both clustered and inhibited patterns to shift to larger scales, the causes of this shift were different for the two distributions. In clustered patterns, measurement error actually caused the perceived size of clusters to increase (Fig. 7). In inhibited patterns, however, the addition of measurement error did not increase the perceived inhibition distance. Instead, it obscured small-scale inhibition while leaving visible larger scale inhibition already present in the data (Fig. 8).

To understand this effect, the characteristics of clustering and inhibition must be considered, and how these are reflected in  $\hat{L}(d)$ . For clustered patterns, the shape of  $\hat{L}(d)$  is determined by the increased number of small interpoint distances relative to randomness. For example, if there are clusters 20 m in diameter, then there are an increased number of interpoint distances between 0 m and 20 m. If measurement errors are added to all the points, even measurement errors as large as  $\sigma_{\text{error}} = 10$  m or  $\sigma_{\text{error}} = 20$  m, there are still clusters. They just look larger and less concentrated. Some of the clusters may blur into each other, but this will just form still larger clusters. When  $\hat{L}(d)$  is calculated from this data, the maximum value is less, due to the lower density of trees within the clusters, and the peak shifts to larger scales, due to the larger size of the clusters. However  $\hat{L}(d)$  still shows clustering.

On the other hand,  $\hat{L}(d)$  for inhibited patterns is characterized by the *absence* of small interpoint distances relative to complete spatial randomness (CSR). If, for example, there is gradually increasing inhibition between 20 m and 0 m, then even a level of measurement error as small as  $\sigma_{\text{error}} = 2$  m or  $\sigma_{\text{error}} = 1$  m will move points sufficiently to increase the number of interpoint distances markedly. Then  $\hat{L}(d)$  will not show the smallest scale inhibition. If the maximum inhibition distance is greater than the scale of measurement error, there will still be some evidence of inhibition at the larger

distances; but once the level of measurement error reaches the scale of the inhibition, nothing will be left of the pattern.

*Wind River Canopy Crane Research Facility stem map.*—In the WRCCRF stem map, mapping each tree from two grid points made it possible to catch data management errors by comparing the two tree locations for discrepancies. This technique reduced data management errors by 4.5%. But both clustered and inhibited patterns are relatively insensitive to data management errors <20%, and these errors would have had little effect on the results of  $\hat{L}(d)$ . However measurement errors play a bigger role. At levels of relative measurement errors on the order of those found in the WRCCRF stem map, the increase in measurement accuracy from the second grid point may be important. This is particularly so if the researcher is interested in estimating the minimum establishment distance for an inhibition process, or in establishing the onset of possible inhibition among small trees.

#### RECOMMENDATIONS

(1) Use at least two measurements to locate each plant. In this study the second measurements on each tree made at the WRCCRF minimized the fieldwork required to correct data management errors once after they were identified. With two or more measurements to locate each plant it is simply necessary to determine which measurement is incorrect, and then map the plant from the remaining measurements. If plants are surveyed only from one measurement, and it is flawed, the false location may not be discovered; and if it is discovered, then it is necessary to resurvey the plant in the field.

(2) Define absolute measurement error for mapped points under field conditions. It is important not to assume there is no measurement error. The type and size of errors depend upon the survey procedure and instruments used. Two techniques can be valuable; (a) a more comprehensive survey of a subplot, and/or one using more accurate equipment, to assess error when reduced measurements are used, and (b) repeated survey of a subplot. However these techniques may not enable estimation of cumulative errors over a large survey area.

(3) Define relative error for the scale of pattern of interest. Necessary map accuracy is a function of the question being asked. A map of individual plant locations cannot be used to investigate processes occurring at scales that approach the accuracy of the measurements. If a map is only going to be used to investigate large-scale processes, such as clustering, then it does not need to be as accurate as a map that will be used to investigate small-scale processes such as inhibition. However use of  $\hat{K}(d)$  in ecology has frequently been concerned with detecting inhibition and defining its scale. If an individual plant location map is used for a general investigation of all spatial processes oc-

curing in the community, it should be kept in mind that any spatial processes approaching the scale of the mapping accuracy might not be revealed by  $\hat{K}(d)$ . When errors are considered,  $\hat{K}(d)$  is not equally powerful for all values of  $d$ . Relative error is greater for small values of  $d$ , so there is a greater chance of accepting the null hypothesis that  $\hat{K}(d)$  does not differ from complete spatial randomness, when in fact it does (Type II error).

#### ACKNOWLEDGMENTS

We are grateful to Dr. David Shaw and his colleagues at the WRCCRF for having facilitated the mapping of trees, and to Dr. Peter Guttorp for helpful suggestions during analysis. Research was performed in whole or in part at the Wind River Canopy Crane Research Facility located in the Thornton T. Munger Research Natural Area in Washington State, USA. The facility is a cooperative scientific venture among the University of Washington, the U.S. Forest Service Pacific Northwest Research Station, and the Gifford Pinchot National Forest.

Although the research described in this article has been funded in part by the U.S. Environmental Protection Agency through agreement CR825173-01-0 to the University of Washington, it has not been subjected to the Agency's required peer and policy review, and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

#### LITERATURE CITED

- Batista, J. L. F., and D. A. Maguire. 1998. Modeling the spatial structure of tropical forests. *Forest Ecology and Management* **110**:293–314.
- Basag, J. 1977. Contribution to the discussion of Dr. Ripley's paper. *Journal of the Royal Statistical Society B* **39**:193–195.
- Boose, E. R., E. F. Boose, and A. L. Lezberg. 1998. A practical method for mapping trees using distance measurements. *Ecology* **79**:819–827.
- Clark, P. J., and F. C. Evans. 1954. Distance to nearest neighbor as a measure of spatial relationships in a population. *Ecology* **35**:445–453.
- Connell, J. H. 1970. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest tree diversity. Pages 298–312 in P. J. den Boer and G. R. Gradwell, editors. *Dynamics of numbers in populations*. Proceedings of the Advanced Study Institute, Centre for Agricultural Publication and Documentation, Wageningen, The Netherlands.
- DeBell, D. S., and J. F. Franklin. 1987. Old-growth Douglas-fir and Western hemlock: a 36 year record of growth and mortality. *Western Journal of Applied Forestry* **2**(4):111–114.
- Diggle, P. J. 1983. *Statistical analysis of spatial point patterns*. Academic Press, London, UK.
- Donnelly, K. 1978. Simulations to determine the variance and edge-effect of total nearest neighbor distance. Pages 91–95 in I. Hodder, editor. *Simulation methods in archeology*. Cambridge University Press, London, UK.
- Ford, E. D. 1975. Competition and stand structure in some even-aged plant monocultures. *Journal of Ecology* **63**:311–333.
- Franklin, J. F., and D. S. DeBell. 1988. Thirty-six years of tree population change in an old-growth *Pseudotsuga-Tsuga* forest. *Canadian Journal of Forest Research* **18**:633–639.
- Franklin, J. F., and C. T. Dyrness. 1988. *Natural vegetation of Oregon and Washington*. Oregon State University Press, Corvallis, Oregon, USA.
- Freeman, E. A. 1997. *The effects of data quality on spatial statistics*. Thesis. University of Washington, Seattle, Washington, USA.
- Haase, P. 1995. Spatial pattern analysis in ecology based on Ripley's  $K$ -function: introduction and methods of edge correction. *Journal of Vegetation Science* **6**:575–582.
- Hall, R. B. W. 1991. A re-examination of the use of interpoint distances and least squares in mapping forest trees. *Ecology* **72**:2286–2289.
- Hopkins, B. 1954. A new method for determining the type of distribution of plant individuals. *Annals of Botany* **18**:213–227.
- Janzen, D. H. 1970. Herbivores and the number of tree species in tropical forests. *American Naturalist* **104**:501–528.
- Kenkel, N. C. 1988. Pattern of self-thinning in jack pine: testing the random mortality hypothesis. *Ecology* **69**:1017–1024.
- Kenkel, N. C. 1993. Modeling Markovian dependence in populations of *Aralia nudicaulis*. *Ecology* **74**:1700–1706.
- Kenkel, N. C., M. L. Hendrie, and I. F. Bella. 1997. A long-term study of *Pinus banksiana* population dynamics. *Journal of Vegetation Science* **8**:241–254.
- Laesele, A. M. 1965. Spacing and competition in natural stands of sand pine. *Ecology* **46**:65–72.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Second English edition. Elsevier Science, Amsterdam, The Netherlands.
- MathSoft. 1996. *S+SPATIALSTATS User's manual, Version 1.0*. MathSoft, Seattle, Washington, USA.
- Moeur, M. 1993. Characterizing spatial patterns of trees using stem-mapped data. *Forest Science* **39**:756–775.
- Moeur, M. 1997. Spatial models of competition and gap dynamics in old-growth *Tsuga heterophylla-Thuja plicata*. *Forest Ecology and Management* **94**:175–186.
- Phillips, D. L., and J. A. MacMahon. 1981. Competition and spacing patterns in desert shrubs. *Journal of Ecology* **69**:97–115.
- Pielou, E. C. 1959. The use of point-plant distances in the study of the pattern of plant populations. *Journal of Ecology* **47**:607–613.
- Pielou, E. C. 1962. The use of plant-to-neighbour distances for the detection of competition. *Journal of Ecology* **50**:357–367.
- Ripley, B. D. 1977. Modeling spatial patterns (with discussion). *Journal of the Royal Statistical Society. Series B* **39**:172–212.
- Rohlf, F. J., and J. W. Archie. 1978. Least-squares mapping using interpoint distances. *Ecology* **59**:126–132.
- Sterner, R. W., C. A. Ribic, and G. E. Schatz. 1986. Testing for life history changes in spatial patterns of four tropical tree species. *Journal of Ecology* **74**:621–633.