

TESTING FOR HOMOGENEITY OF TWO-DIMENSIONAL SURFACES

JAMES J. FILLIBEN AND KAREN KAFADAR

Statistical Engineering Division
National Bureau of Standards
Washington, DC 20234, USA

DOUGLAS R. SHIER

Department of Mathematical Sciences
Clemson University
Clemson, South Carolina 29631, USA

Abstract—This paper presents and develops three approaches for evaluating the degree of “homogeneity” or “randomness” present in a two-dimensional material. Such problems of detecting homogeneity arise in several technological, biological and geological applications. The first approach defines the concept of a “cluster” and develops some test statistics based on the number of clusters and their size distribution. The second approach divides the material into blocks and then applies a nonparametric analysis of variance to these blocks. The third approach employs a graph-theoretic model, in which the length of a certain minimum spanning tree is used as a measure of homogeneity. These approaches are compared using simulation experiments, and are then applied to a problem from materials science.

1. INTRODUCTION

Materials which are developed for use as calibration standards for scientific measurement are frequently required to be checked for homogeneity. An example arises in connection with the preparation of standard reference material for the calibration of X-ray diffraction equipment. At the National Bureau of Standards, small metal austenite/ferrite disks containing a fixed percentage of austenite are made available for use by the general scientific community. Ideally, so as to provide unbiased calibration readings in the field, the austenite should be evenly distributed throughout the disk. The characterization of this ideal state of a homogeneous material and the subsequent testing for it are the subject of this paper.

The actual process which produces these disks of, say, 5% or 10% (by weight) austenite involves a blending procedure. The metal contains retained austenite which, upon blending, preparing, and compaction, transforms into martensite. Homogeneity is affected by physiogeometric properties of the austenite particles such as slope, size, and density. The homogeneity or “mix” of austenite and ferrite is strongly affected by blending time—typical blending times are 8 to 15 min. Too short a time is unsatisfactory due to insufficient merging; too long a time is also unsatisfactory due to segregation resulting from particle fields being set up by the tumbling action. There is, therefore, the question as to what constitutes an

Address all correspondence to Dr. D. R. Shier, Department of Mathematical Sciences, 0-104 Martin Hall, Clemson University, Clemson, SC 29631.

optimum blending time. If homogeneity can be characterized quantitatively, then a test could be performed to determine if the material has been sufficiently blended. Unfortunately, the data available cannot be collected during the blending process but can be taken only after a quantity of the mix has been extended from the batch, pressed into disk form, and surface processed. Such surface processing includes an etching so as to provide optical discrimination of austenite versus ferrite (austenite will appear black and ferrite white). These disks are then examined in detail (about 600 subfields per disk surface) via an optical scanner to generate the raw data for the disk, viz., the austenite/ferrite proportion for each subfield along with the coordinates of the subfield. Such data are then statistically analyzed to determine *a posteriori* homogeneity.

This two-dimensional homogeneity issue arises in other scientific fields as well. In quality assurance programs, blocks of material may be checked for surface smoothness. In biological investigations, the density of plant or animal tissue may be recorded at equally spaced locations, for purposes of detecting regions of high/low density. In numerical analysis, the output for various 2-(or higher-) dimensional random number generators may be checked for randomness.

In all these situations, the underlying question remains the same: Given a two-dimensional grid of measurements, can we assure the lack of a gradient, structure, or trend in the values? Notice that this formulation of the question bypasses any concept of a formal statistical distribution: regardless of the underlying distribution, the focus is on the relative values of nearby observations and on their local predictability. The question is not how to probabilistically characterize the observations generally; rather it is to characterize how observations relate (or covary) with another locally. This concept of homogeneity is identical to the concept of randomness. At first glance, the association appears inappropriate, for a uniformly dense material (constant percentages throughout) is as far from random as one can imagine. In the austenite example, however, one must keep in mind that the reported values are percentages, which are accumulations from (hopefully) many randomly-scattered particles of the element throughout the subarea. If, on the contrary, the particles are clumped together, then whole patches of high percentages are likely to appear, suggesting inhomogeneity. Furthermore, the ideal case of a constant surface in such a situation is unobtainable. Comparing the values of local subareas in relation to the variability in the entire material is a more reasonable method of assessing homogeneity, which brings us back to the randomness concept.

This notion of homogeneity is closely related to the notion of stationarity. In higher dimensions, such stationary random processes are often called homogeneous random fields [1] and are typically studied in connection with turbulence [2]. There has, however, been little attention devoted to whether a given process is in fact a homogeneous random field. Miller and Kahn [3] suggest a two-way analysis of variance (rows and columns of the grid) to test whether a horizontal or vertical trend exists in geological mapping data. However, trends may exist in other directions, and may not extend across the entire region. Variograms and correlograms are suggested in [4] for the purpose of model specification; presumably one may use these tools to detect nonstationarity, though no attempt seems to have thus far been considered. The book by Ripley [5] addresses the problem of homogeneity (particularly Chaps. 5 and 6) and also references related work in other fields in its excellent bibliography, but again a model for the underlying distribution of the data is assumed. For the problem at hand, a formal test should be distribution-free; it should not matter whether the data are uniformly, or normally, or Poisson distributed. A general approach towards discerning local patterns is therefore needed.

Three approaches to the detection of inhomogeneity are explored in this paper. The first approach groups the data into clusters, where a few number of clusters indicates inhomogeneity but a large number suggests random placement of high and low values (i.e.,

homogeneity). This method has the advantage that it can be applied even with the most discrete data. The second approach subdivides the material and applies a nonparametric analysis of variance to the subareas. The third approach generates a minimum spanning tree from the differences in response values of adjacent neighbors.

In Sec. 2, we explore and formally define the concept of homogeneity. Section 3 develops the clustering approach, Sec. 4 investigates the suitability of Kruskal-Wallis tests on subareas, and Sec. 5 treats the minimum spanning tree approach. The various approaches are compared on various inhomogeneous alternatives in Sec. 6 and are then applied to data from a sample austenite standard reference material in Sec. 7. Extensions and additional considerations are discussed in Sec. 8.

2. CHARACTERIZATION OF HOMOGENEITY

One starting point is the dictionary definition of homogeneity: "of the same or a similar nature; of uniform structure or composition throughout." Since we are measuring the percent density of an element in a material, ideally we would like constant percentages throughout. In cases where the response is fixed throughout, we adhere to the dictionary definition and the surface is described as "homogeneous."

More commonly, however, the response is variable over the surface. Loosely speaking, a surface in which the percentages are highly variable within small areas may be considered "homogeneous," if such highly-varying subareas occur consistently throughout the material. In this regard, one may more easily characterize the lack of homogeneity as distinct patches of consistently small/large values. In trying to achieve homogeneity by mixing, a few number of large sparsely or densely populated patches would suggest insufficient mixing, whereas a large number of small patches would indicate that homogeneity had been more or less achieved.

A more rigorous definition of homogeneity is, however, needed. There are many synonyms (both statistical and colloquial) which are used to define homogeneity; for example, random, independent, uniform, uncorrelated, unpredictable, structureless, patternless, etc. Some of these terms (e.g., independence) have precisely-defined statistical definitions having to do with the decomposability of probability functions; others do not. All are related to the homogeneity concept.

In this paper, homogeneity is taken to be statistically equivalent to 2-dimensional randomness. If response values are "randomly" distributed across the various subfields of a surface, then such a surface will be called homogeneous. This begs the question to some extent because we have replaced one concept (homogeneity) by another equally-slippery concept (randomness).

In practice, the issue is resolved by defining homogeneity or randomness in terms of nonpredictability. The essence of homogeneity is related to our ability to make predictions based on *a priori* data. To be precise, suppose that we know what a response value is at some arbitrary (but fixed) point on a surface. How well are we able to predict the response value at adjacent points on the surface? If we can predict the adjacent readings "quite well" then such values are said to covary and we do not have homogeneity. On the other hand, if we can predict response values at locally adjacent points with no better chance than at some distant point across the 2-dimensional surface, then the surface will be said to be "homogeneous."

Having defined what homogeneity is, what is it not? To answer this question, consider the following four key concepts which characterize a general data set:

- (1) Typical Values/Location—Where is the distribution of the data centered? What is a single good typical value to summarize the data?
- (2) Variation/Scale—How spread out is the distribution of the data?

(3) Distribution—What formal statistical distribution (Gaussian, Poisson, uniform, Weibull, etc.) describes the shape of the density function of the data?

(4) Randomness.

Homogeneity clearly relates to concept 4 and has no relation to concepts 1, 2, or 3. We shall use the words “homogeneous” and “random” interchangeably because of this conceptual equivalence. The homogeneity of a surface has nothing to do with the location, scale, or statistical distribution of the data. The term “homogeneity” is a distribution-free concept.

To amplify on this, consider the following: Suppose that we have 200 response value drawings from some fixed distribution (say, Gaussian) with some fixed location (say, $\mu = 100$) and some fixed dispersion (say, $\sigma = 5$). Suppose these 200 values are by construction randomly scattered across the subfields of some 2-dimensional surface. We have therefore, by construction, a homogeneous surface—given a reading at some particular subfield of the surface, we can predict a reading at neighboring points no better than predicting a reading at some distant point.

Suppose further the 200 raw readings are transformed (e.g., each new value is the logarithm of the original value). The resulting distribution of the new response value is now no longer Gaussian [since if y is distributed as a Gaussian, then $\log(y)$ cannot be]. Is the surface still homogeneous? The answer is affirmative because, given a reading at some point on the surface, predicting a reading at a neighboring point still cannot be done any better than predicting a reading at some distant point. The statistical distribution of the response values over the surface has nothing to do with the homogeneity of that surface.

A major source of confusion is that the words “homogeneous,” “uniform,” and “uniform distribution” (i.e., having a flat density function) are often used interchangeably. This is incorrect because (as seen from the previous paragraph) the randomness of a surface has nothing to do (in general) with the statistical distribution of the surface, and (in particular) has nothing to do with whether or not a surface has a uniform statistical distribution. Further, in most physical applications, the statistical distribution of data on a surface is rarely uniform per se. Much more common are distributions such as Poisson (as in the distribution of fiber in a material) or Gaussian (as in the mix of austenite particles in a calibration standard). The approaches described in this paper are therefore aimed at the assessment of homogeneity without reference to the underlying statistical distribution of the data.

3. APPROACH 1: CLUSTERS

3.1. *Definition*

A natural method of assessing homogeneity involves contouring. This may be done at various levels of complexity. For simplicity, we consider only two possible codes (leading to contours of like coded values): those subregions for which the observed values are above the median value of the response, indicated by a “+,” and those which are below, indicated by a “−.” We shall concentrate on clusters of one type only: we define a “cluster” in a grid of subareas as a group of “+”’s which can be connected by horizontal or vertical lines. A neighbor is thus considered to be in one of at most four possible directions.

This simple definition may be easily extended to more complex contouring and inclusion of additional neighbors in the cluster definition. Notice also that this definition leaves the shape of the material or subregion unspecified; only the concept of “neighbor” is required. The exclusion of diagonals in this definition is for reasons of simplicity. (The approach of Sec. 5 does include such diagonal neighbors, however.) A substantial amount of literature exists in physics on the identification of arbitrarily defined clusters [6, 7, 8, 9], some of which utilizes concepts of graph theory [10, 11, 12].

3.2. Associated statistics

In analogy with one-dimensional runs-tests, we will count the number of clusters of a given size, as well as the total number. For simplicity, let us assume that our grid is square, of size $N = n \times n$, and let

N_k = number of clusters of size k (i.e., consisting of exactly k "+"s),

$$N_T = \text{total number of clusters} = \sum_{k=1}^{[N/2]} N_k,$$

$S_{\max}$ = size of largest cluster = $\max\{k: N_k > 0\}$.

By definition, we will have equal numbers of "+"s and "-"s. (When N is odd, we will assume that the median has been replaced by a "-.")

It is clear that the maximum number of clusters is $[N/2]$, obtained by a pattern of alternating "+" and "-" signs. In addition, the maximum number of clusters of size k does not exceed $N/2k$, obtained by ignoring boundary effects and alternating "+" clusters, interspersed with only as many "-"s as needed to isolate them.

Ideally, we search for the probability of obtaining m clusters of size k , in order to determine $E(N_k)$ and hence $E(N_T)$. Together with a variance, we might formulate a test statistic based on N_T , and similarly for $S_{\max}$. The 3×3 (126 possible patterns) and 4×4 (12,870 possible patterns) cases were generated via computer, and the corresponding probabilities and expected values are given in Table 1. The distribution of N_k (one of which is shown in Fig. 1) and of $S_{\max}$ (Fig. 2) are all very skewed. The distribution of N_T for the 3×3 case is more symmetrical (Fig. 3), making it the more likely candidate for a test statistic (by virtue of the central limit theorem).

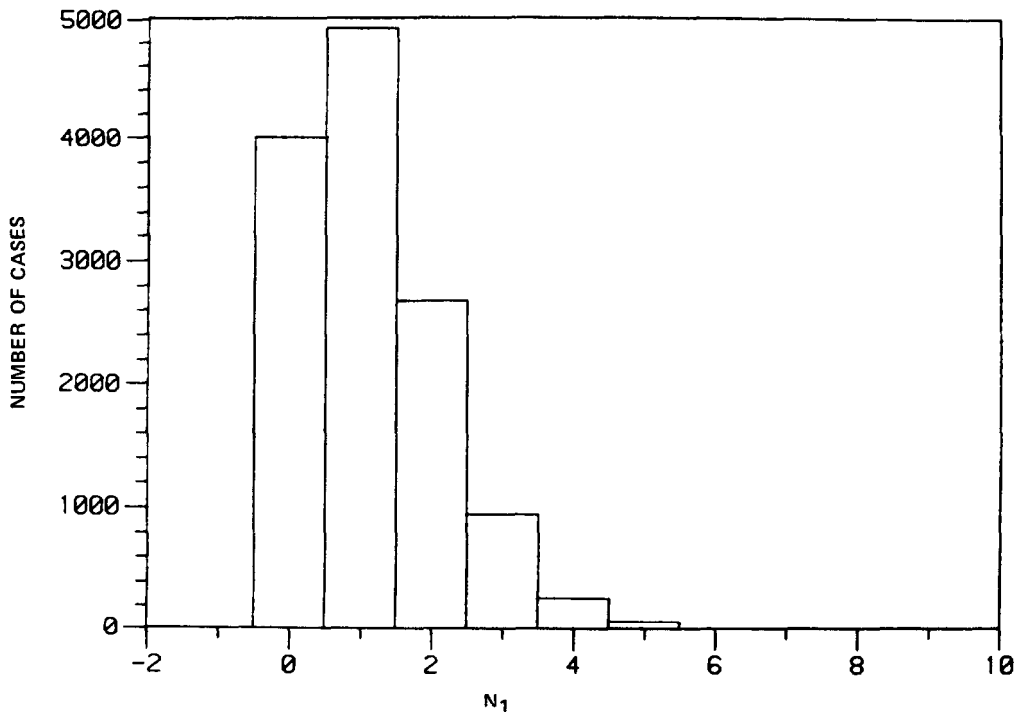


Fig. 1. Histogram for statistic N_1 , 4×4 grid (total number of cases = 12,870).

Table 1. Probabilities and expected values for exhaustive 3×3 and 4×4 grids.

(A) 3×3 case: $P\{N_k = m\} \cdot T$ (T = total number of cases = 126)											
k/m	0	1	2	3	4	$E(N_k)$		$SD(N_k)$			
1	48	44	28	—	6	0.984		1.016			
2	86	28	12	—	—	0.413		0.658			
3	82	44	—	—	—	0.349		0.477			
4	90	36	—	—	—	0.286		0.452			
$E(N_T) = 2.032$					$SD(N_T) = 0.835$						
$E(S_{\max}) = 2.873$					$SD(S_{\max}) = 0.882$						
(B) 4×4 case: $P\{N_k = m\} \cdot T$ (T = total number of cases = 12,870)											
k/m	0	1	2	3	4	5	6	7	8	$E(N_k)$	$SD(N_k)$
1	3997	4920	2678	944	257	64	8	—	2	1.128	1.038
2	8435	3624	752	44	15	—	—	—	—	0.413	0.628
3	9644	2516	710	—	—	—	—	—	—	0.306	0.568
4	10,165	2417	288	—	—	—	—	—	—	0.233	0.472
5	10,310	2560	—	—	—	—	—	—	—	0.199	0.399
6	10,490	2380	—	—	—	—	—	—	—	0.185	0.388
7	10,706	2164	—	—	—	—	—	—	—	0.168	0.374
8	11,396	1474	—	—	—	—	—	—	—	0.115	0.318
$E(N_T) = 2.746$					$SD(N_T) = 1.071$						
$E(S_{\max}) = 5.398$					$SD(S_{\max}) = 1.575$						

N.B. Dashes are zeroes.

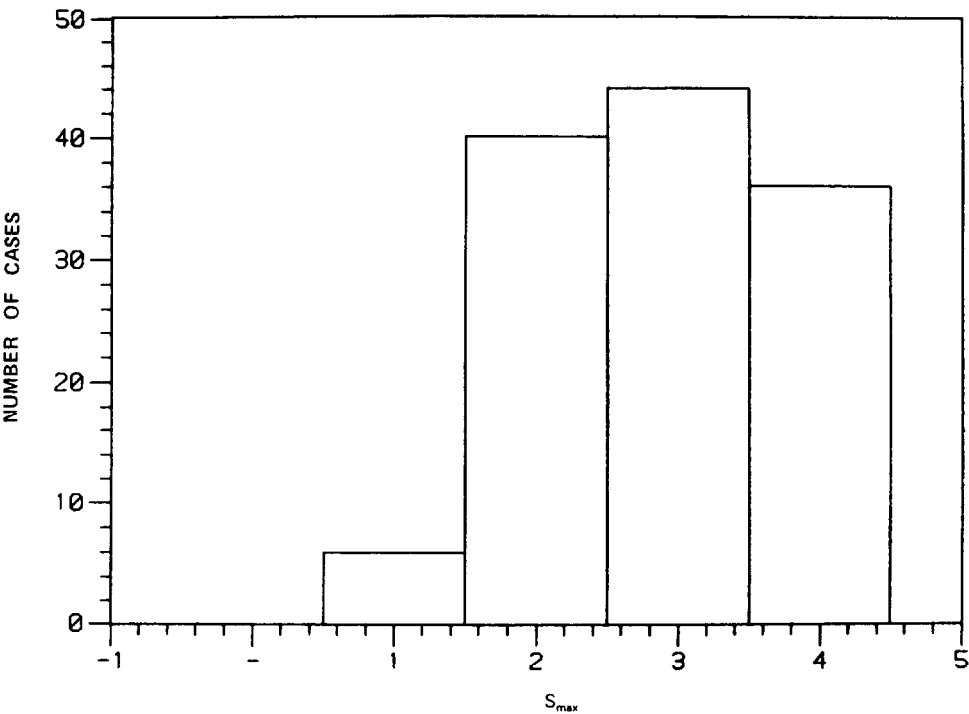


Fig. 2. Histogram for statistic $S_{\max}$, 3×3 grid (total number of cases = 126).

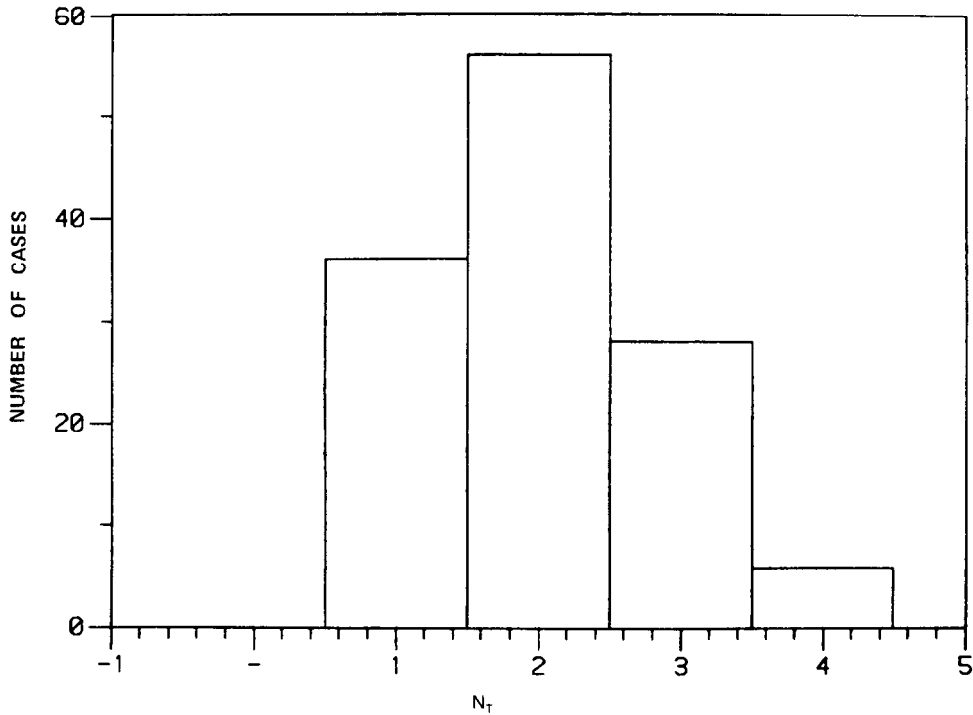


Fig. 3. Histogram for statistic N_T , 3×3 grid (total number of cases = 126).

3.3. Simulation results and approximations

For larger values of N , only a sample (usually 1000) of the $\binom{N}{\lfloor N/2 \rfloor}$ cases was generated, and the expected values and standard deviations of $S_{\max}$, N_T , and N_1 were calculated (Table 2). In addition, N_T and $S_{\max}$ are negatively correlated (the larger the size of the largest cluster, the fewer clusters there will be); for example, the correlation in the 3×3 case is found to be -0.94 . As a function of grid size, $E(S_{\max})$ appears linear and $E(N_T)$ quadratic (Figs. 4 and

Table 2. Expected values, standard deviations, and relative standard deviations for $S_{\max}$, N_T , and N_1 (based on a subset of the number of possible cases).

	Grid Size (No. of Samples)					
	5 (1000)	10 (1000)	15 (822)	20 (975)	25 (1000)	30 (1000)
$S_{\max}$	7.03 (2.24) 0.32	21.64 (7.71) 0.36	37.00 (14.59) 0.39	53.65 (21.29) 0.40	68.66 (25.68) 0.37	83.57 (32.65) 0.39
N_T	3.83 (1.34) 0.35	10.24 (2.46) 0.24	20.17 (3.59) 0.18	33.74 (5.16) 0.15	50.76 (5.82) 0.11	70.10 (7.01) 0.10
N_1	1.62 (1.27) 0.78	4.31 (2.09) 0.48	8.75 (3.07) 0.35	15.10 (4.05) 0.27	23.22 (4.94) 0.21	31.80 (5.77) 0.18

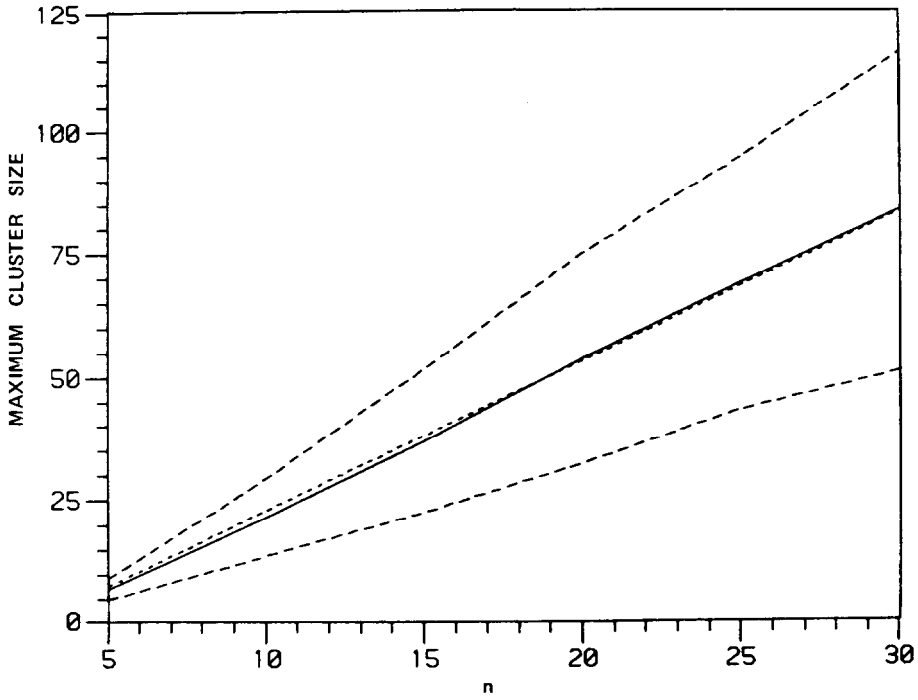


Fig. 4. $E(S_{\max})$ as a function of grid size n , shown with limits of one standard deviation. Solid line = actual, dotted line = predicted.

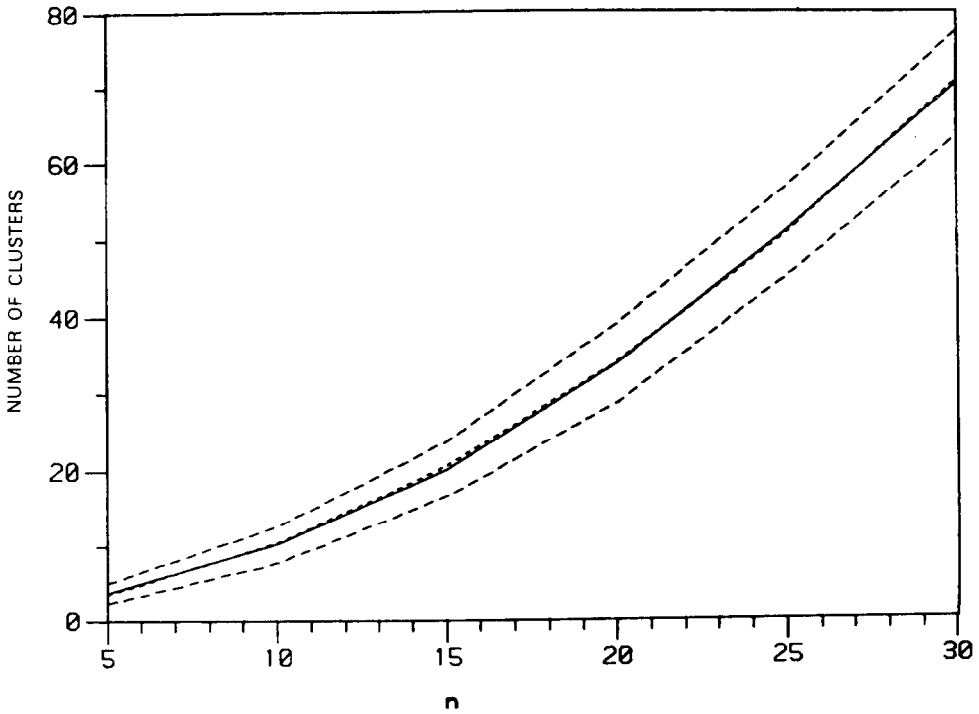


Fig. 5. $E(N_T)$ as a function of grid size n , shown with limits of one standard deviation. Solid line = actual, dotted line = predicted.

5, shown with limits of one estimated standard deviation of this average); however, $S_{\max}$ has a larger standard deviation and a slowly decreasing relative standard deviation. Least squares fits as a function of grid size n are shown in Figs. 4 and 5 (solid line) and are given by

$$\begin{aligned} E(N_T) &\cong 0.377 + 0.344n + 0.0666n^2 \\ \text{SD}(N_T) &\cong 0.180 + 0.231n \end{aligned} \quad (1)$$

$$\begin{aligned} E(S_{\max}) &\cong -7.403 + 3.027n \\ \text{SD}(S_{\max}) &\cong -3.327 + 1.190n. \end{aligned} \quad (2)$$

Using the central limit theorem, a (two-sided) test for homogeneity might be as follows: Declare inhomogeneous if

$$\begin{aligned} \text{(a) total number of clusters} &< (0.377 + 0.344n + 0.067n^2) - 1.96(0.180 + 0.231n) \\ &\text{or} > (0.377 + 0.344n + 0.067n^2) + 1.96(0.180 + 0.231n) \end{aligned} \quad (3)$$

(i.e., $N_T < 0.024 - 0.109n + 0.067n^2$ or $N_T > 0.730 + 0.797n + 0.067n^2$); or

$$\begin{aligned} \text{(b) size of largest cluster} &> -7.403 + 3.027n + 1.96(-3.327 + 1.190n) \\ &\text{or} < -7.403 + 3.027n - 1.96(-3.327 + 1.190n) \end{aligned} \quad (4)$$

(i.e., $S_{\max} > -13.924 + 5.359n$ or $S_{\max} < -0.882 + 0.695n$). The use of (4) is *very* tentative, as the value of n for which the central limit theorem is appropriate has not been determined. Such tests should have approximately 5% error rate.

4. APPROACH 2: KRUSKAL-WALLIS ON SUBAREAS

A second approach considers differences in the values of neighboring blocks of various sizes. Consider a partition of the data set into n_b blocks, each of size b , containing the observations $\{X_{ij}; i = 1 \dots b; j = 1 \dots n_b\}$. The concept of homogeneity suggests that the patterns of response values ought to be independent of the level of detail at which we look. In terms of our blocks, this says

$$\begin{aligned} E(X_{ij}) &= E(\bar{X}_{.j}) = \mu_j(b) \\ \text{var}(X_{ij}) &= b \cdot \text{var}(\bar{X}_{.j}) = \sigma_j^2(b) \end{aligned} \quad j = 1 \dots n_b \quad (5)$$

In fact, these means and variances should be independent not only of the block (of a given size), but also of the block size itself. We may therefore consider tests for differences in $\{\mu_j(b), j = 1 \dots n_b\}$ as well as differences for different values of b . Since we wish to keep any statistical tests free of underlying distributional assumptions, we suggest the Kruskal-Wallis statistic for each case of a given block size [13]:

$$K(b) = \frac{12}{N(N+1)} \sum_{j=1}^{n_b} b(\bar{R}_j - \bar{R}_{..})^2$$

where $\bar{R}_j$ = average of the ranks in j th block and $\bar{R}_{..}$ = average of all ranks $= (N+1)/2$. However, blocks of different geometric shapes will provide different answers and possibly even different conclusions. Several tests using different block sizes will hopefully attenuate this degree of dependence.

Each block size will result in a test statistic $K(b)$, which has an asymptotically chi-squared distribution on $n_b - 1$ degrees of freedom. In order to make it as difficult as possible to reject the null hypothesis of homogeneity, we suggest reporting the least significant statistic; i.e., let T_K be defined by the following equation:

$$P(\chi_{(n_b-1)}^2 \leq T_K) = \min P\{\chi_{(n_b-1)}^2 \leq K(b)\}. \quad (6)$$

The values of b over which this minimum is taken will of course depend upon the number of data points; surely one would want to limit b to less than 20% of the data, to permit at least 5 blocks for comparison. Since T_K is defined as a minimum, its distribution is not strictly chi-squared, but reporting the probability levels for the least significant statistic will be conservative. A less robust version of this approach involving a comparison of mean squares is suggested in [5].

Having compared the differences in block averages for various block sizes, we now turn to estimates of spread of the observations in the blocks. In general, Eq. (5) for the variance of the average $\bar{X}_{j,}$ is

$$b \cdot \text{var}(\bar{X}_{j,}) = \sigma_b^2 + b^{-1}[b(b-1) \text{covariance terms}].$$

If the material is homogeneous, then a plot of σ_b^2 versus b should be relatively constant. If inhomogeneity is present, then the effect of the covariance terms should become more noticeable with increasing values of b , up until the point at which the internal heterogeneity of the observations within the larger blocks dominates that between the blocks. Thus, the effect of inhomogeneity may be seen by plotting an estimate of $b \cdot \text{var}(\bar{X}_{j,})$ versus block size b (or $\sqrt{b} \cdot \text{SE}(\bar{X}_{j,})$ versus $\sqrt{b}$, where SE denotes standard error). For our purposes, the median, $M_{j,}$ provides a better estimate of central block tendency than does the average, for, regardless of the nature of the mechanism generating the underlying distribution, we are more interested in that value which informs us that half the data fall on either side of it. In this study, we took

$$\sqrt{b} \cdot \text{SE}(M_{j,}|b) = (1.25 \cdot \text{hinge spread})/(1.35 \cdot \text{cf}), \quad (7)$$

where the hinge spread is uniquely defined in [14]. The factors 1.25 and 1.35 are appropriate for the Gaussian-based approximation (see [15] for the suitability of this approximation to other distributions). The correction factor cf is given in [14, p. 632], and it adjusts for the fact that

$$\phi(\text{upper hinge}) - \phi(\text{lower hinge})$$

is not exactly 0.50 for a Gaussian sample of finite size n (ϕ denotes the standard cumulative Gaussian distribution function). A comparison box plot of the estimates (7) obtained on the blocks of different sizes may indicate patterns which contradict the assumption of homogeneity. A more formal test based on this exploratory procedure needs to be developed. An example of this approach on a data set will be presented in Sec. 7.

5. APPROACH 3: MINIMUM SPANNING TREE

Inasmuch as several types of problems in clustering and pattern recognition can also be studied using minimum spanning trees [16, 17, 18], this concept also seems appropriate for the current study. Accordingly, the present section will investigate a statistic based on the minimum spanning tree for analyzing homogeneity or randomness.

5.1. Definition

Informally, a minimum spanning tree connects together nodes of some network at minimum total “cost.” For example, in planning the construction of a distributed computer system, the *nodes* i correspond to computer sites, the undirected *edges* (i, j) join two sites that can be directly connected by a high-speed communication link, and the *cost* of each edge is simply the cost of constructing that particular communication link. In order to connect up all sites so that any two sites can communicate (either directly or indirectly) with one another, a *spanning tree* is required: a tree that meets, or “spans,” every node. Moreover, if the total cost of a spanning tree is defined as the sum of its constituent edge costs, then a minimum cost spanning tree is sought.

In the present context, the nodes of the network are associated with subareas of the given grid; a response value or density ρ_i is defined at each node i . Every node i is joined by an edge to those nodes adjacent to i by a single horizontal, vertical or diagonal displacement. Thus, for this approach every “interior” node of the grid now has eight neighbors (see Fig. 6), while the “boundary” nodes have either three or five neighbors. The example in Fig. 6 displays the 20 nodes and 55 edges of a 4×5 grid network. More generally, an $m \times n$ grid network contains $N = mn$ nodes and $E = 4mn - 3n - 3m + 2$ edges.

We define the *cost* of edge (i, j) to be $|\rho_i - \rho_j|$; that is, the unsigned difference in response values between the two nodes joined by the edge. Thus, nodes of similar densities will have a low edge cost, while nodes of fairly different densities will have a large edge cost. The minimum spanning tree (MST) provides a way of aggregating these edge costs into a single quantity that can be used in evaluating the homogeneity or randomness of the grid densities.

Because the minimum spanning tree connects all nodes at smallest possible cost, any nodes forming a “cluster” of spatially contiguous and similar densities will be connected together by the MST at a small aggregate cost. Accordingly, if there are large clusters (indicating nonrandomness) then the total cost of the MST will tend to be small. On the other hand, if the densities are randomly distributed on the grid, then we should expect to incur a number of large edge costs—whence the total MST cost should be large. It is hoped, then, that the total cost of the MST might serve as an indicator of nonrandomness, with larger values for the situations deemed “more random.”

Several basic procedures have been proposed for constructing minimum spanning trees [19, 20] and numerous variants of these procedures have also been developed. Because the

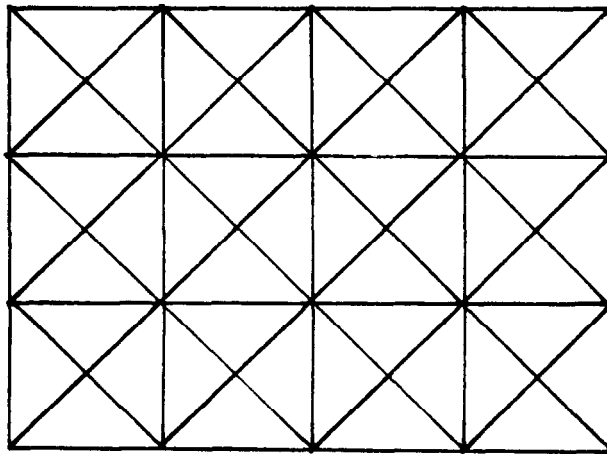


Fig. 6. A sample 4×5 grid network.

applications of interest may have grid sizes of 35×35 or larger—the former involving more than 1200 nodes and 4600 edges—there is definitely a need for an efficient solution algorithm that has modest storage requirements as well. Therefore, a quite efficient computer implementation of a Kruskal-type MST algorithm [21] has been employed in the simulation experiments conducted here.

5.2. Simulating the random (homogeneous) case

A first series of simulation experiments modeled the situation of “random grid densities.” Namely, densities ρ_i for each node i of an $n \times n$ grid network (with $N = n^2$ nodes and $E = 4n^2 - 6n + 2$ edges) were drawn randomly from a uniform distribution over $(0, 1)$. Edge costs c_{ij} were calculated using

$$c_{ij} = |\rho_i - \rho_j|,$$

and the total cost (TC) of a MST with respect to these edge costs was calculated by the MST algorithm:

$$TC = \sum_{(i,j) \in \text{MST}} c_{ij} = \sum_{(i,j) \in \text{MST}} |\rho_i - \rho_j|.$$

The minimum spanning tree is clearly unaffected by positive scaling ($\rho_i \rightarrow k\rho_i$) or translation ($\rho_i \rightarrow \rho_i + k$), but the quantity TC is not scale invariant. In order to remedy this situation and obtain a “dimensionless” test statistic, we also computed the normalized quantities

$$TN_1 = \frac{\sum_{(i,j) \in \text{MST}} |\rho_i - \rho_j|}{\sum_{i=1}^N |\rho_i - \bar{\rho}|}, \quad (8)$$

$$TN_2 = \frac{\sum_{(i,j) \in \text{MST}} (\rho_i - \rho_j)^2}{\sum_{i=1}^N (\rho_i - \bar{\rho})^2}, \quad (9)$$

where

$$\bar{\rho} = \frac{1}{N} \sum_{i=1}^N \rho_i$$

is the sample mean of the node densities ρ_i . Statistic TN_2 generalizes to two dimensions the well-known von Neumann ratio used in testing for the randomness of univariate time series [22].

The above statistics TC, TN_1 , and TN_2 were calculated for each of 100 repetitions (using a different sample of the ρ_i 's) on $n \times n$ grids. In this study, a range of grid sizes was studied, from 5×5 grids (25 nodes) to 35×35 grids (1225 nodes). The mean and standard deviation of each such test statistic are shown in Table 3, averaged over the 100 repetitions. As that table indicates, the relative standard deviation of each statistic decreases as the grid size n increases. This desirable property means that the distribution of the statistic is becoming more and more sharply localized as the grid size becomes larger. Also, the relative standard deviations for TC and TN_1 are comparable and decidedly smaller than that for TN_2 .

In order to test the sensitivity of these results to the assumption of a uniform distribution for the ρ_i 's, a second set of simulations was conducted using normally distributed (and

Table 3. Means and standard deviations for three test statistics, based on uniformly distributed node densities over $n \times n$ grids (100 repetitions).

Grid Size	TC		TN ₁		TN ₂	
	Mean	SD	Mean	SD	Mean	SD
5 × 5	3.12	0.482	0.518	0.0802	0.360	0.107
10 × 10	10.4	0.937	0.417	0.0419	0.224	0.0396
15 × 15	22.2	1.28	0.396	0.0235	0.199	0.0222
20 × 20	38.4	1.35	0.385	0.0173	0.187	0.0169
25 × 25	58.5	1.89	0.376	0.0138	0.176	0.0121
30 × 30	83.7	2.46	0.373	0.0119	0.172	0.00991
35 × 35	112.4	2.74	0.368	0.0101	0.167	0.00847

Table 4. Means and standard deviations for three test statistics, based on normally distributed node densities over $n \times n$ grids (100 repetitions).

Grid Size	TC		TN ₁		TN ₂	
	Mean	SD	Mean	SD	Mean	SD
10 × 10	39.2	4.29	0.490	0.0361	0.290	0.0465
20 × 20	145.6	7.18	0.457	0.0163	0.254	0.0207

independent) node densities, with mean 10 and standard deviation 1. Results for 10×10 and 20×20 grids are given in Table 4. It is seen that the assumed distribution for the ρ_i 's does indeed have an effect on the mean and standard deviation of the test statistics. Ideally, one would prefer a test statistic that is relatively insensitive to the distributional form postulated for the node densities.

5.3. A distribution-free statistic

In an attempt to purge the test statistic of possible distribution dependence, we can first replace each node density ρ_i by its *rank* R_i . Thus, the smallest of the N node densities receives rank 1, the second smallest receives rank 2, and so forth. A new test statistic can then be based on the total cost of a minimum spanning tree with respect to the node ranks R_i :

$$TR = \sum_{(i,j) \in MST} |R_i - R_j|. \quad (10)$$

Because the MST is scale and translation invariant, then so is TR. More generally, TR is unaffected by any monotone transformation of the original node densities.

Another series of simulation experiments was performed using both uniformly and normally distributed node densities, independently generated at each node of an $n \times n$ grid. The resulting means and standard deviations for TR, based on 100 repetitions, are displayed in Table 5. It is seen that the relative standard deviation of TR decreases monotonically as the grid size increases, indicating a more and more sharply defined test statistic. Moreover, at the selected 5×5 , 10×10 , 20×20 , and 30×30 grid sizes, there is reasonably close agreement between the values of the test statistic for uniform $U(0, 1)$ and normal $N(0, 1)$ node densities.

The distribution of the 100 values for TR at any grid size appears to be fairly symmetric and approximately normal. As an illustration, Fig. 7 displays a histogram for the TR values obtained using a 25×25 uniform grid. In addition, Fig. 8 shows a normal probability plot

Table 5. Means and standard deviations for TR based on uniform and normal node densities over $n \times n$ grids (100 repetitions).

Grid Size	Uniform		Normal	
	Mean	SD	Mean	SD
5 × 5	83.2	11.0	82.3	9.21
6 × 6	159.3	13.9		
8 × 8	464.2	35.6		
10 × 10	1066.8	78.7	1063.8	68.5
12 × 12	2156.7	121.7		
15 × 15	5049.7	247.7		
17 × 17	8202.8	377.8		
20 × 20	15,473.1	535.5	15,469.5	536.4
22 × 22	22,501.9	897.7		
25 × 25	36,799.0	1142.6		
27 × 27	49,942.7	1534.9		
30 × 30	75,494.0	1998.4	75,692.5	2351.5
32 × 32	97,324.9	2237.2		
35 × 35	137,914.3	3146.2		

for these same data; such a plot indicates approximate normality of data by the approximate linearity of the plotted points. While Fig. 8 suggests deviations from normality in the tail areas, the majority of the data appears to approximate that from a normal distribution.

As is evident from Table 5, the mean and standard deviation of TR clearly increase as the grid size n increases. In order to provide a calibration curve for the dependence of these

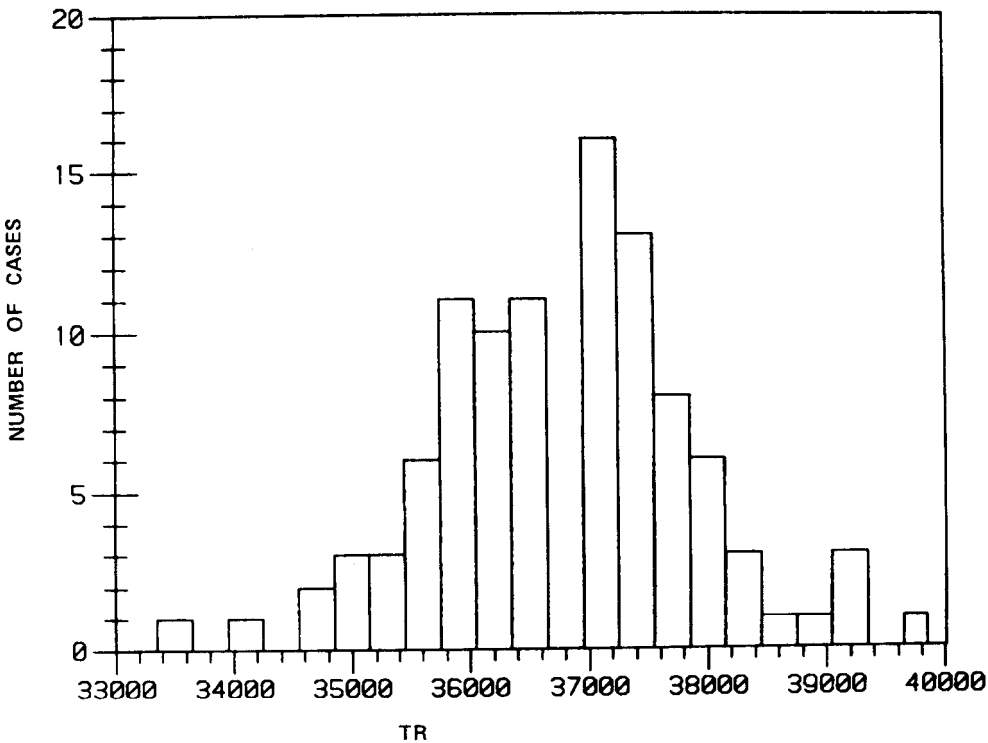


Fig. 7. Histogram for TR, 25 × 25 grid networks, using uniformly generated node densities.

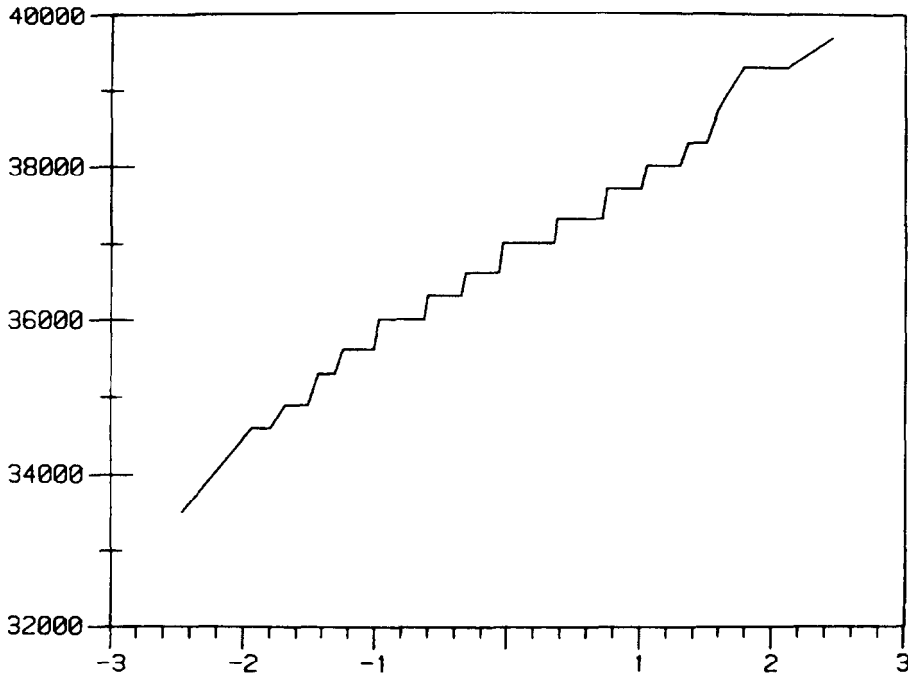


Fig. 8. Normal probability plot for TR, 25×25 grid networks, using uniformly generated node densities.

values on n , a number of regression models were investigated. Quite adequate fits for the mean $E(\text{TR})$ and the standard deviation $\text{SD}(\text{TR})$ were obtained using the following models:

$$E(\text{TR}) \cong \frac{1}{a} (n^2 + cn + d)^b, \quad (11)$$

$$\text{SD}(\text{TR}) \cong \frac{1}{e} n^f, \quad (12)$$

where

$$\begin{aligned} a &= 10.06, \\ b &= 1.984, \\ c &= .6115, \\ d &= 1.652, \\ e &= 13.11, \\ f &= 2.989. \end{aligned}$$

These estimates were based on 14 grid sizes ($n = 5$ to $n = 35$) for uniformly distributed node densities. Thus, $E(\text{TR})$ grows approximately as $(n^2)^{1.984} = n^{3.968}$, while $\text{SD}(\text{TR})$ grows approximately as $n^{2.989}$. As n increases, then, the relative standard deviation $\text{SD}(\text{TR})/E(\text{TR})$ must decrease—a desirable situation.

6. COMPARISON OF TEST STATISTICS UNDER VARIOUS ALTERNATIVES

6.1. Patterns of inhomogeneity

In evaluating test statistics for one-dimensional problems, one applies them to non-null situations and determines their success in recognizing such situations (i.e., establishes a "power curve"). In two-dimensional problems, the non-null situation can be any of an infinite

Table 6. Deterministic inhomogeneous patterns (5×5 case).

(1) Increasing horizontally					(2) Alternating high and low				
1	2	3	4	5	1	14	2	15	3
6	7	8	9	10	16	4	17	5	18
11	12	13	14	15	6	19	7	20	8
16	17	18	19	20	21	9	22	10	23
21	22	23	24	25	11	24	12	25	13
(3) Increasing diagonally					(4) L-shaped peaks and valleys				
1	2	4	7	11	1	16	2	19	7
3	5	8	12	16	18	17	3	20	8
6	9	13	17	20	6	5	4	21	9
10	14	18	21	23	25	24	23	22	10
15	19	22	24	25	15	14	13	12	11

variety of inhomogeneous patterns. It is impossible to generate all possible patterns, but we suggest a few schemes in order to evaluate the three approaches discussed in this paper.

The first four alternatives are deterministic patterns; i.e., no random generation is involved. Since all of our test statistics are applied to ranked data, these patterns merely place the numbers 1 through $N = n^2$ in a set fashion. They are shown in Table 6 for the 5×5 case. The first alternative sets the ranks down horizontally, row by row. The second alternative places the first $(N + 1)/2$ ranks alternately with the last $(N - 1)/2$ ranks. The third alternative places the ranks in increasing order diagonally, and the fourth generates L-shaped ridges.

Since no problem will ever involve nicely generated patterns such as those in Table 6, the remaining two alternatives have been simulated so as to create patterns which have increasing amounts of association within them. The fifth alternative is based on generating correlated values within the grid. Given $N = n^2$ independent and uniformly distributed random numbers, we transform them so that the correlation between any two numbers is constant:

$$\text{corr}(X_i, X_j) = 1/\sqrt{m}, \quad m = 2, 3, 4, \dots$$

Our choice of $1/\sqrt{m}$ was suggested by the analogy to the situation in which alternatives in one-dimensional tests are considered for the purpose of power comparisons (e.g., see [13]). The data are then ranked as before. We should see a progression of increasingly significant test statistics as m decreases. The calculations that were used to generate these pairwise correlated random values are given in the Appendix.

The sixth alternative is similar to the previous one, except that the grid of uniformly generated random numbers is smoothed by replacing each value by the average of its adjacent neighbors. Each successive smoothing results in a grid whose values are increasingly correlated. Statistics are calculated on grids after one, two, and three smoothings. In this alternative (as well as in Alternative 5), a total of 100 repetitions were performed by varying the initial set of N (randomly selected) grid values.

6.2. Additional test statistics

We also compare the approaches of this paper with three other statistics which have been suggested in the geographical literature [23, 24, 25]:

$$\text{TM} = \frac{12}{E(N^2 - 1)} \sum_{\text{edges}} (R_i - \bar{R})(R_j - \bar{R})$$

$$\text{TG} = \frac{6}{EN(N + 1)} \sum_{\text{edges}} (R_i - R_j)^2$$

$$\text{TS} = \sum_{\text{edges}} |R_i - R_j|,$$

Table 7. Comparison of normalized test statistics on inhomogeneous patterns.

	TM (Moran)	TG (Geary)	TS (Royalty <i>et al.</i>)	TR (MST)	N_T (Cluster)	T_K (KW)
Alternative 1: Increasing horizontally						
5 × 5	6.442	-10.567	-8.181	-5.029	-1.320	—
9 × 9	14.090	-18.460	-19.629	-10.742	-3.040	16.000
15 × 15	25.883	-30.536	-37.800	-18.647	-5.081	27.713
19 × 19	33.798	-38.582	-50.154	-23.689	-6.337	29.693
Alternative 2: Alternating high and low						
5 × 5	1.127	-1.981	-1.046	-2.146	15.160	—
9 × 9	2.934	-3.944	-3.623	-6.487	31.488	2.885
15 × 15	5.833	-6.947	-7.993	-13.496	55.824	4.293
19 × 19	7.782	-8.945	-11.021	-18.218	72.017	4.702
Alternative 3: Increasing diagonally						
5 × 5	6.849	-10.414	-8.371	-5.029	-1.320	—
9 × 9	14.804	-18.409	-20.024	-10.742	-3.040	16.011
15 × 15	26.776	-30.517	-38.329	-18.647	-5.081	27.716
19 × 19	34.753	-38.570	-50.733	-23.689	-6.337	31.417
Alternative 4: L-shaped peaks and valleys						
5 × 5	-1.678	3.271	1.665	-1.292	0.927	—
9 × 9	-3.257	4.978	3.844	-5.639	0.059	0.974
15 × 15	-5.057	6.822	6.564	-12.844	-1.515	1.926
19 × 19	-6.182	7.891	8.265	-17.664	-2.617	3.244
Alternative 5: Correlation = 0.1						
5 × 5	1.447	-2.169	-1.348	-1.152	1.714	—
9 × 9	6.460	-8.158	-6.496	-3.756	1.187	7.156
15 × 15	13.061	-17.670	-16.606	-7.463	1.322	15.077
19 × 19	19.035	-21.662	-20.155	-9.808	1.642	24.329
Alternative 5: Correlation = 0.5						
5 × 5	3.203	-5.030	-3.367	-2.344	0.560	—
9 × 9	8.134	-10.485	-8.779	-5.007	0.404	9.367
15 × 15	13.187	-17.837	-16.811	-8.283	0.886	16.410
19 × 19	19.946	-22.689	-21.437	-10.421	1.699	23.552
Alternative 6: One smoothing						
5 × 5	3.448	-5.392	-3.908	-2.417	0.126	—
9 × 9	7.839	-9.840	-8.591	-4.367	-0.477	4.507
15 × 15	13.638	-15.296	-14.146	-6.431	-1.345	4.572
19 × 19	17.782	-19.023	-53.380	-7.330	-1.840	3.576
Alternative 6: Two smoothings						
5 × 5	5.789	-8.587	-6.642	-4.012	-0.803	—
9 × 9	12.491	-14.975	-14.546	-7.395	-2.146	9.363
15 × 15	21.245	-23.676	-24.604	-11.260	-3.638	10.921
19 × 19	21.159	-29.414	-31.342	-13.465	-4.657	10.608

where $\bar{R}$ = average rank = $(N + 1)/2$; E = number of edges in grid = $4mn - 3n - 3m + 2$ [$= 2(2n - 1)(n - 1)$ if square grid, $m = n$]. Notice that these statistics use the differences in ranks between all adjacent grid points (i.e., all E edges in the grid network; cf. Fig. 6), whereas the MST uses only $N - 1$ of these edges.

In reporting the values of these statistics on the various alternatives, we shall normalize them by subtracting off expectations and dividing by standard deviations. For N_T (total

number of clusters of one type) and TR (length of minimum spanning tree), we use the approximations given in Eqs. (1) and (11)–(12), respectively. For the Kruskal–Wallis approach, several statistics are calculated, corresponding to blocks of sizes 3×3 , $4 \times 4 \dots 10 \times 10$. From these, T_K , as defined in (6), is the least significant of the statistics. To compare the value of T_K with other statistics, we report its normalized form; namely,

$$(T_K - df)/\sqrt{2 \cdot df},$$

where df = the degrees of freedom corresponding to the $K(b)$ which defines T_K . Note that this approximation is intended only for a semiquantitative comparison with the other statistics. The expectation and standard deviation for the Sokal statistic TS are derived in [25]; those for the Geary and Moran statistics (TG and TM) are given in [26]. On larger grid sizes, the central limit theorem allows us to compare the normalized values with typical 95%/97.5% critical points from the Gaussian distribution (i.e., $\pm 1.645/\pm 1.96$).

6.3. Comparison of statistics

Table 7 displays the results of comparing the normalized test statistics on various alternatives for $n \times n$ grids, where $n = 5, 9, 15, 19$. For the fixed alternatives 1 and 3, with response values that have positive spatial correlation, all test statistics except N_T are quite powerful in rejecting the null hypothesis, even for the smallest grid sizes. On the other hand, when response values exhibit negative spatial correlation (Alternatives 2 and 4), the MST statistic is the most powerful, whereas the Kruskal–Wallis and the clustering statistics are the least powerful, especially for Alternative 4. Although N_T was originally formulated to reject the null hypothesis for negative normalized values of N_T , Alternative 2 shows that N_T should rightly be used in a two-tailed test, as in Eq. (3).

For the increasingly correlated response values of Alternative 5, all test statistics except N_T are reasonably powerful, especially for grid sizes 9×9 and larger. The Geary statistic is the most powerful, followed by the Kruskal–Wallis statistic. In Alternative 6 (successive smoothings), all test statistics except for N_T generally reject the null hypothesis quite strongly. The Geary, Sokal, and Moran statistics are the most powerful, followed by the MST and Kruskal–Wallis statistics.

7. AN APPLICATION

Despite the differences in the mechanics of the three approaches (e.g., definition of neighbor), the three approaches are, surprisingly, somewhat related in concept. Their similarities may be more clearly recognized by considering binary (0–1) data. The Kruskal–Wallis test on subareas indirectly counts average number of one's within local regions. It has the advantage that it uses the (ranked) numerical data but the final answers depend upon the specific geometric nature of the subareas. When diagonal adjacencies are included in the definition of a cluster, the value of TR from the minimum spanning tree gives the number of “+” clusters plus the number of “–” clusters less one. In fact, if there are clusters of several different types, then

$$\text{TR} = \text{total number of clusters of all types} - 1.$$

That this is so is seen by the fact that the total cost TR is augmented only by connecting clusters of different types without producing any cycle (hence the “– 1”). As a result, we may expect the standard deviation of TR to increase as it is computed for increasingly discrete data. The approach to a given problem, therefore, largely depends on the discreteness of the data.

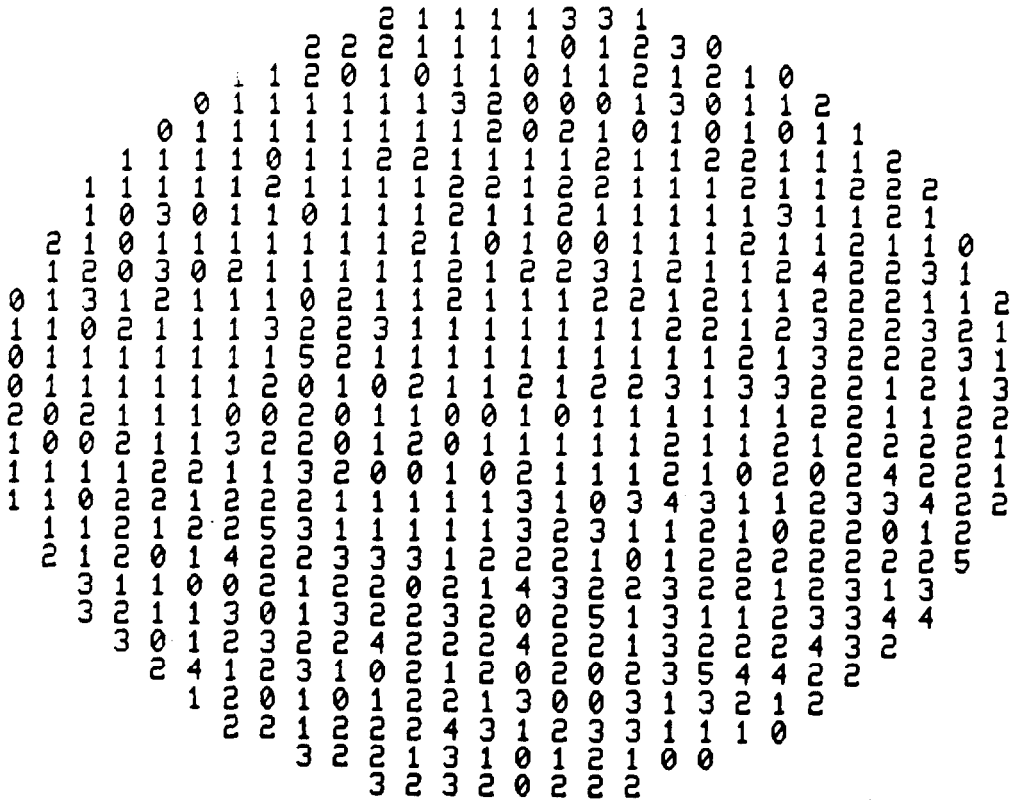


Fig. 9. Rounded percentages of austenite in subareas (0.75×0.75 mm) for a sample austenite/ferrite standard reference material.

Table 8. Three approaches applied to a standard reference material of austenite.

(A) Number of clusters of size k (see Sec. 3)															
k	1	2	3	4	5	6	7	8	9	10	11	12	13	...	184
N_k	22	3	8	1	1	2	0	1	2	0	0	1	1	...	1

$N_T = 43$: cf. $E(N_T) \cong 49.9 \pm 5.9$ from (1)

$S_{\max} = 184$: cf. $E(S_{\max}) \cong 67.7 \pm 26.2$ from (2)

(B) Kruskal-Wallis statistic (see Sec. 4)

Block Size $b^{(*)}$	n_b = No. of Blocks	Kruskal-Wallis Statistic	$\chi^2_{n_b-1}(0.0005)$
9	67	124.98	110.45
16	37	99.65	70.59
25	22	75.20	49.01
36	15	61.34	38.11
49	11	64.47	31.42
64	10	43.17	29.67
81	8	47.76	26.02
100	6	59.22	22.12

(C) Normalized length of minimum spanning tree (see Sec. 5)

Ranked Data: TR = 19237.5, cf. $E(TR) \cong 35904.9 \pm 1125.7$ [from (11)–(12)]

*Due to the circular nature of the austenite specimen, a block was required to have at least $b/2$ observations. Thus, the number of observations in each block of nominal size b is really k_b , $i = 1 \dots n_b$, where $b/2 < k_b \leq b$. This fact is used in all subsequent calculations.

To test the efficacy of these three approaches, we use the data provided by Dr. George Hicho at the Center for Materials Science at the National Bureau of Standards. Figure 9 shows the density of austenite/ferrite in a given standard reference material after a certain blending time. Dr. Hicho's preliminary judgment refuted the homogeneity assumption. The results of applying the three approaches described in this paper are shown in Table 8. (Since this material is circular, we chose $n = \sqrt{\text{number of values}} = \sqrt{616} = 24.8$, even though it is not an integer.) All three approaches concur with scientific judgment; i.e., there is definite lack of homogeneity in the sample.

The cluster approach finds a cluster of considerably large size (184) and thus clearly indicates inhomogeneity. The Kruskal-Wallis approach (Panel B) rejects the assumption (equal means in all blocks of size b) for eight block sizes [$\sqrt{b} = 3(1)10$] at the 0.0005 level of significance. (The simultaneous evaluation of eight such tests indicates an overall level of 0.004 at the very highest. It is probably substantially smaller due to the high degree of correlation among all tests.) The statistic calculated from the minimum spanning tree (Panel C), being considerably smaller than the expected value obtained from the simulation study, also firmly rejects the assumption of homogeneity. Finally, Fig. 10 compares, via box plots, the estimates of the standard errors of the medians of each of the n_b blocks of size b ; e.g., the box plot at $\sqrt{b} = 3$ is constructed using the estimates of the standard error of M_j , $j = 1 \dots n_b$, from Eq. (7). The general trend of this plot (as represented by the asterisks in the boxes) is surprisingly constant. In connection with the other three tests, we conclude that this test is probably the least powerful of the approaches we have discussed.

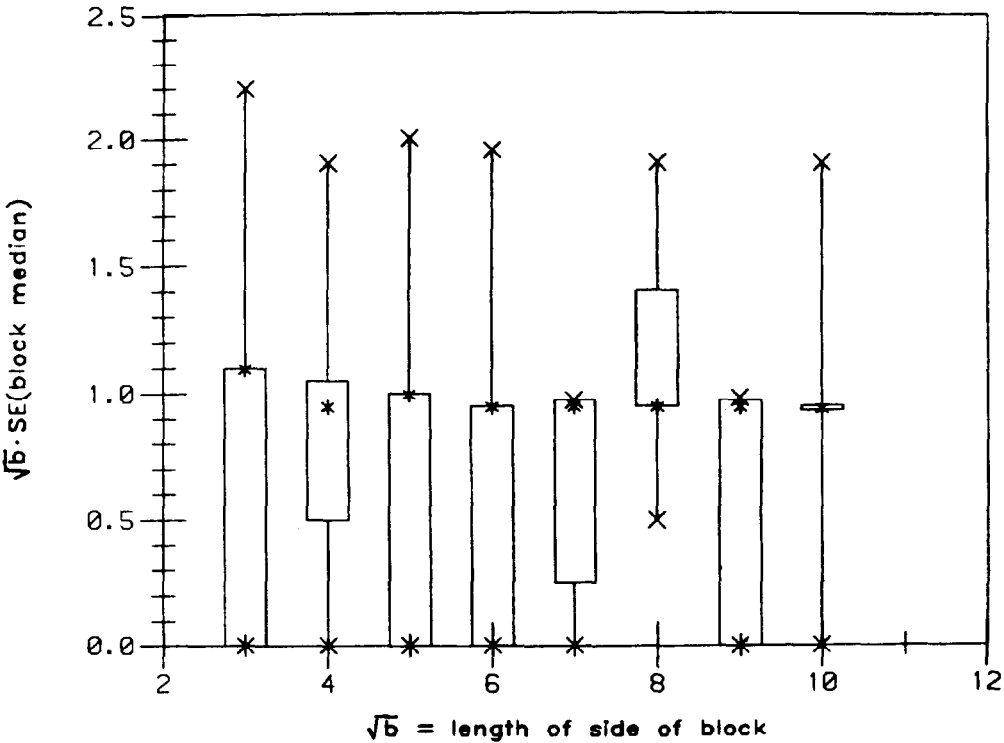


Fig. 10. Comparison box plots of normalized estimates of standard errors of block medians (ideally constant).

8. CONCLUSIONS AND EXTENSIONS

The assessment of homogeneity may be a straightforward problem with no simplistic solutions. We have proposed three procedures: any may be used separately or in conjunction with other tests and/or scientific judgment. None of these procedures can be expected to replace graphs or visual representations of the data: the goal is to aid the scientist in making a judgment about these graphs. Nor should the role of the scientist himself be undermined, for only by combining these techniques with knowledge of the underlying physical mechanism can accurate judgments be made.

Two of the three methods are fundamentally statistical in nature. The cluster approach is simple and is probably the most directly related to the given problem, but, for most types of alternatives, the statistic on which we concentrated, N_T , is the least powerful of the techniques. Perhaps an alternative formulation would suggest a more powerful statistic. The Kruskal–Wallis test, while being applied to “blocks” which have not been obtained as in a strict experimental design situation, nonetheless offers an objective means of assessment and is generally sensitive to inhomogeneity. One may also examine the Kruskal–Wallis statistics on other block sizes, in addition to the least significant one, and compare medians in the different blocks. The third approach uses a graph-theoretic tool: minimum spanning trees, and is also quite powerful on most alternatives, but it is more expensive to compute. This application to the homogeneity problem is just one of many problems to which the minimum spanning tree approach has been applied. Both of these latter two approaches appear to be competitive with the statistics suggested by Sokal, Moran, and Geary in geographical contexts. The relationship of any of these three methods to more conventional clustering algorithms (e.g., dendrograms) may suggest additional solutions.

There remain many open questions. We have been concentrating on results for square grids, “interpolating” the results to nonsquare grids (as in Sec. 7); this extension ought to be verified. Other sequences of alternatives might provide additional power comparisons among the various methods. One might also consider a two-dimensional spectrum analysis, as suggested in [5, Chap. 5]. Finally, it may prove worthwhile to smooth the original data in order to obtain a clearer view of the clusters involved. Several methods for two-dimensional smoothing have been suggested, particularly in the geology literature [27, 28, 29]. The techniques described in this paper may then be more usefully applied to such smoothed data.

APPENDIX: GENERATING CORRELATED RANDOM VARIABLES FROM INDEPENDENT ONES

Suppose $X_1, X_2, \dots, X_N$ are independent variates from an arbitrary distribution having mean 0 and variance 1. We wish to find the appropriate linear combinations of these variates, viz.,

$$U_i = \sum_{j=1}^N a_{ij} X_j,$$

such that

$$\text{correlation}(U_i, U_j) = 1/\sqrt{m} \quad i \neq j.$$

Since

$$\begin{aligned} E(U_i) &= 0 \\ \text{var}(U_i) &= \sum_{j=1}^N a_{ij}^2, \end{aligned}$$

we will normalize the coefficients so that

$$\text{var}(U_i) = \sum_{j=1}^N a_{ij}^2 = 1 \quad i = 1 \dots N.$$

If we denote the matrix of coefficients by A , having row vectors $\mathbf{a}_i' = (a_{i1}, a_{i2} \dots a_{iN})$, then we require that

$$\begin{aligned} \mathbf{a}_i' \mathbf{a}_i &= 1 \quad \text{for all } i; \\ \mathbf{a}_i' \mathbf{a}_j &= 1/\sqrt{m} \quad j \neq i; \end{aligned} \quad (*)$$

that is, the inner product of each row with every other must be the same. Clearly, we need to specify only $N(N+1)/2$ of the coefficients in A ; hence, take A to be a lower triangular matrix. If we let R denote the reciprocal of the common correlation between the variables, then the equations (*) can be solved:

$$a_{ij} = \begin{cases} 0 & \text{for } j > i \\ a_{i-1,j} & \text{for } j = 1 \dots i-2 \\ \left(\frac{(R-1)}{R(R+i-3)(R+i-2)} \right)^{1/2} & \text{for } j = i-1 \\ \left(\frac{(R-1)(R+i-1)}{R(R+i-2)} \right)^{1/2} & \text{for } j = i \end{cases}$$

Acknowledgement—We would like to thank Dr. George Hicho of the National Bureau of Standards for providing the data in Sec. 7 and for initiating this pursuit of research. We benefitted from valuable discussions from many of our friendly colleagues, notably Professors John Rice, J. Scott Provan, Steven Grotzinger, Charles Johnson, and Dr. Charles Fenimore. We especially thank Stefan Leigh of the Statistical Engineering Division, National Bureau of Standards, for help in the definition and research of the problem, and for carrying out the simulations in Sec. 3.

REFERENCES

1. H. Cramér and M. R. Leadbetter, *Stationary and Related Stochastic Processes*, Wiley, New York (1967).
2. A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, Prentice Hall, Englewood Cliffs, NJ (1962).
3. R. L. Miller and J. S. Kahn, *Statistical Analysis in the Geological Sciences*, Wiley, New York (1962).
4. B. Matérn, Spatial variation. *Comm. Swed. Forestry Res. Inst.* **49**, 1-144 (1960).
5. B. Ripley, *Spatial Statistics*, Wiley, New York (1981).
6. E. L. O'Neill, *Introduction to Statistical Optics*, Addison-Wesley, Reading, MA (1962).
7. M. E. Fisher and J. W. Essam, Some cluster size and percolation problems. *J. Math. Phys.* **2**, 609-616 (1961).
8. M. F. Sykes and J. W. Essam, Exact critical percolation probabilities for site and bond problems in two dimensions. *J. Math. Phys.* **5**, 1117-1120 (1964).
9. E. A. Trabka, Crowded emulsions: Granularity theory for monolayers. *J. Optical Soc. Am.* **61**, 800-810 (1971).
10. I. J. Good, The generalization of Lagrange's expansion and the enumeration of trees. *Proc. Camb. Phil. Soc.* **61**, 499-517 (1965).
11. P. Erdős and A. Rényi, On the evolution of random graphs. *Bull. Inst. Int. Statist.* **38**, 343-347 (1961).
12. T. L. Austin, R. C. Fagen, W. F. Penney, and J. Riordan, The number of components in random linear graphs. *Ann. Math. Statist.* **30**, 747-754 (1959).
13. E. L. Lehmann, *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco (1975).
14. J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA (1977).
15. R. McGill, J. W. Tukey, and W. A. Larsen, Variations of box plots. *Am. Statistician* **32**, 12-16 (1978).
16. J. A. Hartigan, *Clustering Algorithms*, Wiley, New York (1975).
17. R. E. Osteen and P. P. Lin, Picture skeletons based on eccentricities of points of minimum spanning trees. *SIAM J. Comput.* **3**, 23-40 (1974).

18. C. T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. Computers C* **20**, 68–86 (1971).
19. J. B. Kruskal, Jr., On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* **7**, 48–50 (1956).
20. R. C. Prim, Shortest connection networks and some generalizations. *Bell Systems Tech. J.* **36**, 1389–1401 (1957).
21. R. E. Haymond, J. P. Jarvis, and D. R. Shier, Computational methods for minimum spanning tree algorithms. Technical Report 354, Department of Mathematical Sciences, Clemson University (1981).
22. J. Johnston, *Econometric Methods*, McGraw Hill, New York (1972).
23. P. A. P. Moran, Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 (1950).
24. R. C. Geary, The contiguity ratio and statistical mapping. *Incorporated Statistician* **5**, 115–145 (1954).
25. H. H. Royaltey, E. Astrachan, and R. R. Sokal, Tests for patterns in geographical variation. *Geograph. Anal.* **7**, 369–395 (1975).
26. A. D. Cliff and J. K. Ord, *Spatial Autocorrelation*, Pion, London (1973).
27. F. T. Agterberg, *Geomathematics*, Elsevier, Amsterdam (1974).
28. J. C. David, *Statistics and Data Analysis in Geology*, Wiley, New York (1973).
29. J. W. Tukey, Smoothing and smelting, lecture given at Department of Statistics, Princeton University (July 26, 1979).