Linear Operators Applied to Areal Data[†]

W.R. Tobler

Regional forecasting. spatial filtering, and map generalization have been treated advantageously by linear analytical methods. In general these methods have been applied to data assembled on a regular lattice. The objective of this study is the extension of these techniques to the case in which the data are assembled in geographical areas with irregular boundaries such as countries, counties. parishes, census tracts, or school districts. These curiously shaped regions are usually the domains of substantive interest and cannot he discarded by subdividing the world into kilometer or centimeter squares. At the same time one would like to retain the analytical power that is as available when working with geographical data in gridded form. To simulate the geographical spread of ideas using Hägerstrand's model of the diffusion of innovations. for example. generally requires the use of geographical data in the form of square cells superimposed on some part of the Earth (Hägerstrand 1968). One concern of this study is to explore how this model might he implemented using data on political units. As a second example. it would he desirable to be able to apply the same enhancement techniques to arbitrary areal data (choropleth maps to the cartographer) as are used to filter contour maps (Holloway 1958, Tobler 1966).

To review briefly, assume that geographical data have been collected at regular spatial intervals in two independent directions in a region small enough to allow neglect of the Earth's curvature. These data will be arranged in the form of a geographical matrix **Z**, with states z_{ij} using the conventional positional notation. Entries in the matrix will now be changed in some manner to become new entries z^*_{ij}. The modification might be. for example completely random, or could depend on the state itself, and could he deterministic or stochastic. The concern here is with processes that are such that the modification of a state z_{ij} depends only on its own state and on the states of the other observations. Symbolically **Z*** is a function of its own state and the collection of remaining states. **Z*** = f(**Z**). Generally, it is convenient to assume closure so that the mapping **Z** → **Z*** has as its result objects of the same type. Such a general function has as many arguments as there are elements in the geographical matrix. The function is now restricted to he a local or neighborhood operator, a function that defines the modification of each z_{ij} in terms of its spatial neighbors only. Such an operation might be defined using a neighborhood that consists of the given element and its eight immediate neighbors. In this case the function has only nine elements and is of the form

$$z^*_{ij} = f(z_{i-1,j-1}, z_{i-1,j}, z_{i-1,j-1}, z_{i,j-1}, z_{i,j}, z_{i,j-1}, z_{i+1,j}, z_{i+1,j+1})$$

Neighborhoods which are larger or of a different shape can be defined in an analogous manner. It is implicit in the notation that every cell has the same definition for its neighborhood, a type of spatial stationarity generally found only in hoard games (Clowes 1970) and mathematics. If the number of states is S and the number of entries in the neighborhood is N. there are obviously S^N situations which the function must cover.

From a substantive, geographical point of view the algebra of the technique also is of concern. For example. if **Z*** = f(**Z**) then in symbolic notation one might write **Z**** = f(**Z***) = ff(**Z**) or $\mathbf{Z}^k = f^k(\mathbf{Z})$. a logical product. repeating the function k times if this is meaningful. An obvious question is what happens as k becomes very large

$$\lim_{k \to \infty} f^k(\mathbf{Z}) = ?$$

Can the procedure be run backwards. $\mathbf{Z} = f^{-1}(\mathbf{Z}^*)$; or again symbolically. given **Z** and **Z*** find f = **Z*/Z** in which we ask whether the *process* can be inferred from empirical instances. Alternatively, and from a planning point of view, given a current set of states **Z** and a desired set **Z*** what are the requirements on f that **Z*** be realizable in k iterations?

To give this substance, suppose that the states are land-use types. The geographical matrix **Z** might be a land-use map at time t and **Z*** might be a land-use map at some earlier or later time period t + Δ**t.** Alternatively. **Z*** might be a generalization of the map **Z** and $(\mathbf{Z}^*)^{-1}$ might represent "ungeneralization" or deconvolution. If there are five land-use categories and a neighborhood consists of a cell and its north, east, south. and west adjacent cells, then a deterministic transition rule would have to cover 3,125 cases. It is standard procedure to assume spatial isotropy. which reduces the number of cases considerably. There are several models of this type that might be used in geography and cartography (Burks 1970, Codd 1968, Feldt 1968, Gardner 1970, Haubaugh and Bonham-Carter 1970, MacDougall 1972, Minskv and Pappert 1969: Ratliff. Hartline. and Miller 1964, Smith 1971**,** Switzer 1969).

In special cases the states can be represented by numbers. with the usual arithmetic properties. Integers might he used for numbers of people by kilometer squares. Real numbers could represent elevations taken from a topographic map or grey tones measured on an aerial photograph. Complex numbers at each grid position can be used to define a two-dimensional vector field. Triplets of numbers at each point might define the tristimulus values for a colored map. or the dimensions of ellipsoids. A wind rose at each point requires a complete function at every point, and so on. The methods described below can he used for all of these examples, even if the number of states is essentially unbounded. The infinity of states does not cause serious problems because the system is closed. Of all possible ways of converting one set of numbers into another only linear or log linear functions are considered here. In the spatially continuous case one such function can he written as

$$z^*(x,y) = \iint w(x, v, u, v)\, z(x, y)\, du\, dv, \text{ integrating over the entire space.}$$

If one invokes translational invariance this becomes

$$Z^*(x,y)= \iint w(u,v)\, z(x - u, y - v)\, du\, dv$$

and in the discrete case

$$z^*_{ij} = \sum_p \sum_q w_{pq} \, z_{i+p,\, j+q}$$

where the summation and weighting are taken over the p. q neighborhood.

One interpretation (Tobler 1970) that can he given to this equation is that it is a partial difference approximation to the partial differential equation

$$\partial z/\partial t = M\{ \partial^2 z/\partial x^2 + \partial^2 z/\partial y^2) + (\beta - \delta)z$$

describing spatial diffusion with sources and sinks. An alternate point of view (Tobler 1969) is to assume that the data have been approximated by a Fourier series, that z^*_{ij} has been similarly represented and that the difference between the two can he interpreted as the effect of a filter whose two-dimensional response function can he calculated from the weights. This is nothing but a change of basis in a linear vector space (Andrews 1970, Rosenfeld 1969, Tobler 1967) and can also he treated as a Markov process (Howard 1971, Woods l972)
.

The foregoing concepts have wide application in map generalization, geographical forecasting, spatial modeling, and picture processing. In virtually all of the literature it has been assumed that data are taken at regular spatial intervals. This is a rare event in geographical practice. Data are either taken at an irregularly distributed set of points in two or more dimensions, or they are aggregated within a set of irregularly shaped regions such as census tracts. Two strategies are used to handle these cases. The more common is to convert the data to a lattice and to proceed from there. The second strategy is to attempt to devise generalized operations which when specialized to regularly spaced observations are equivalent to those normally used.

A simple example should illustrate these two procedures. Suppose topographic elevations are known at randomly distributed locations and we wish to obtain an estimate of the slope at each of these locations. One procedure is to interpolate elevations to a lattice of points and to use standard finite difference methods to obtain derivatives. In the other instance we would attempt to use the observations directly to obtain an estimate of the slope. In principle these two approaches should give the same results (both are linear): in practice only the former strategy is used.

There are basically two problems. The first is to decide on the neighborhood. Given a random scatter of points in two dimensions, which points are neighbors of which other points? The second problem is that of choosing weights for particular purposes.

More formally. assume n observations at points P_1. i = 1...n. in a two-dimensional Euclidean space and label the positions with coordinates x_i,.y_i. At each P_i we have a numerical observation z_i and the set of all of these observations can he called S. In a manner comparable to that given earlier, the objective is controlled modification of the values z_i at each of the observation points. The modification at a point P_i can depend on the value z_i at that point and on the values at all of the other points. Symbolically. we can say that z_i is some function of the value at i and of the collection of remaining values

$$Z^*_i = f(z_k),\ z_k \text{ in } S.$$

As before it is required that the function be linear

$$Z^*_i = \sum_k w_k z_k$$

with as yet undetermined weights. The next restriction is to require that the function he a neighborhood operator. Denote the set of all points that are neighbors of P_i by N_i. With this geographical restriction the modification process can be written as

$$z^*_i = \sum_k w_k z_k,\ \ z_k \text{ in } N_i$$

In cartography a slightly different version of this problem frequently occurs. This is the interpolation problem of estimating a value z^*_o at a point P_o of known location x_o, y_o at which there is no observed value. It is required that this value he estimated from the known values,

$$z^*_o = f(z_k), \qquad z_k \text{ in } S.$$

which is immediately specialized to a linear form

$$z^*_o = \sum_k w_k z_k, \qquad z_k \text{ in } S$$

and is often restricted to be a neighborhood function

$$z^*_o = \sum_k w_k z_k, \qquad z_k \text{ in } N_o$$

The value of z^*_o is usually chosen so that its departure from the true unknown value z_o is minimized generally in the averaged least-squares sense

$$1/m \sum (z^*_o - z_o)^2$$

where the summation is over all locations for which an estimate is desired. Typically the z^*_o form a lattice and these estimated values are then used for subsequent operations in place of the original scattered observations, which are now ignored. One such operation is the drawing of a contour map usually by linear interpolation within the lattice. The similarity of the interpolation problem to the problem of modifying values at points for which values are already known should be apparent. Mathematically the two problems have virtually identical structures. Both are linear weightings of the given observations. The one operation is called prediction and the other is called filtering. but they are almost indistinguishable mathematically.

Consider first the neighborhood question. An adjacency or contiguity table can be constructed by listing all of the points in n rows and n columns and placing a one in the i^{th} row and j^{th} column when points i and j are considered neighbors, inserting a zero otherwise. Clearly there are 2^n ways of doing this. The number of possibilities is reduced if the relation is symmetric and if a point is considered to be its own neighbor. Neighbors are usually chosen by one of the following rules

(I) All points within distance d of the point in question are taken to be neighbors, using some agreed definition of distance.
(2) The degree of neighborliness is defined by the distance from the point in question. Typically this results in weights being a function of distance and these approach zero at the distance D. If this bounding distance is larger than the field of observations then the neighborhood includes the complete set of points.
(3) Thiessen neighbors (Rhynsburger 1973). A second-order neighbor is the neighbor of a neighbor: This can be extended to higher orders. Normally all neighbors of the same order receive the same weight.
(4) Gabriel neighbors (Sokal and Gabriel, 1969. p. 266 -270).
(5) The nearest k points, using some agreed definition of distance, are taken to be neighbors.
(6) Triangulation: each point has associated with it three neighbors and the area is partitioned into triangles. The number of topologically distinct triangulalions is given by Brown (1965) to be

$$\frac{2(2m+3)!(4n+2m+1)!}{(m+2)!m!n!(3n+2m+3)!}$$

where m is three less than the number of points in the convex hull making up the outer boundary of the region and n is the number of interior points. The number of geometrically distinct triangulations is somewhat less than this topological value. Among these there will be some whose cumulative edge lengths are minimal.
(7) Quadrangulation: each point has associated with it four neighbors and the area is partitioned into quadrangles (Brown, 1965).
(8) An assignment of neighbors using additional knowledge: *a posteriori* empirical or *a priori* theoretical knowledge is used.

Several of these definitions can be made to coincide for a lattice, which is clearly a special case. The important property is that each lattice point or cell always has the same number of neighbors: this lends considerable homogeneity to (he processing operations. The concept of spatial stationaritv is especially easy to visualize in this context. Try inventing a chess-like game in which the cells have variable numbers of neighbors. In the ensuing discussion it will be assumed that the second neighbor definition is used. In this case the controlled modification of data pertaining to irregularly spaced observations is fairly simple.

Suppose first that the filter kernel is known to be K(u, v). This kernel is centered on the position P_i at which the observation z(x, v) = z_i to be modified is located, which assumes spatial stationarity. The weights to be applied to the individual observations are now given by the value of this kernel at those points

$$w_k = K(u - x, .v - y)$$

where point k is located at (u, v). This value is then normalized so that

$$\sum w_k = \iint_{-\infty}^{+\infty} K(u.v)\, du\, dv.$$

The modified value z^*_i of z_i is then given as a linear combination of these weights and the observations

$$z^*_i = \sum w_k z_k.$$

The kernel will vary for different types of linear operations. Two common types are differentiation and smoothing (low-pass filtering). Most important is “optimal'' filtering whereIn it is assumed that the observations themselves may be contaminated by measurement errors (Moritz 1962 1963 1967 1970) has shown that the Wiener-Hopf equation leads to a kernel which involves covariance functions. Identical optimal filters have been derived for use in oceanography, in astronomy. in meteorology, in photogrammetry. and in mining geology (Bracewell 1955, Burr 1955, Kraus 1971, Krige 1966, Switzer. Mohr. and Heitman 1964, Thompson 1956). The same theory also leads to “optimal” least-squares prediction (interpolation), where it can be shown that the best predictor, when the observations are given as deviations from the mean, when there is no trend, and when the process has a stationary covariance, is

$$\mathbf{Z}^* = \mathbf{A}\mathbf{B}^{-1}\mathbf{Z}.$$

Z* is the k by 1 vector of values to be predicted. **Z** is the i by 1 vector of known observations. **B** is the covariance matrix relating the observations, and **A** is the covariance between the values at the locations to be predicted and the known values. Thus the optimum prediction requires that something (the covariances) be known about the phenomena being studied. It is usually assumed that the covariances are functions only of the distances between the observations. Interestingly this theory also shows that the expected error of the prediction is

$$\text{error} = 1/k \,\Sigma\, (\mathbf{Z}^* - \mathbf{Z})^2 = \sigma^2 - \mathbf{A}\mathbf{B}^{-1}\mathbf{A}'$$

where σ^2 is the variance of the observations (Moritz 1967). Thus it is possible to provide confidence and consequence maps that can he used to estimate the value of additional observations (Epstein 1969, Rapp 1964).

The next step is to extend these methods to areal data given, for example, by census tracts. Here contiguous units are usually recognized as neighbors and only the question of weights remains. In several cases the data are simply assumed to he concentrated at some point within the area: usually the centroid is taken (Pitts 1967). From then on we can proceed as if the data were initially observed at points thus falling back on the case discussed above. The result of any operation is assumed to apply without change to the entire area with which the point is associated.

A second obvious approach is to use a set of weights proportional to the length of the common boundary (Cliff and Ord 1969). Call this L_{ij} and use for L_{ii} the total length of the boundary of area i. Clearly $L_{ij} = L_{ji}$ a theoretically desirable feature since asymmetry is prohibited. The weights can now be normalized

$$L'_{ij} = L_{ij} / \sum_{j} L_{ij}$$

and the n by l vector of modified areal values is obtained from the n by 1 vector of observed values by the multiplication

$$\mathbf{Z}^* = \mathbf{LZ}$$

where **L** is the n-square matrix of weights. Another attempt has been made to combine both boundary lengths and centroid distances (Cliff and Ord 1970). The reason is obvious. Irregularly shaped regions may consist of long narrow countries, with long boundary contact or of countries of large size making boundary contact only through narrow necks or corridors. But both size and border contact would seem to be important.

One procedure which captures these two effects is to assume that the kernel K(u, v) is given by a continuous function. Next assume that the geographical data are given by an at least piecewise continuous function z(x, y). This function is of the form z(x,. y) = z_i if (x, y) is in region i. Figure 1 shows an important example. Until about 1820 it was thought that it was impossible to deal with such “pathological’ graphs. But we now know better. As an example the equation and its derivation for a one-dimensional histogram are given by Lanczos (1966 p. 37). Suppose we write our geographical data as a continuous function, using the notation z(x. v). expandable as an infinite Fourier series in two dimensions

$$z(x, y) = \int\int_{-\infty}^{+\infty} A(\alpha, \beta) \exp [i \pi x + i \pi y] \, d\alpha \, d\beta$$

The linear weighting procedure would then be written as the spatial convolution

$$z^*(x, y) = \int\int_{-\infty}^{+\infty} k(u, v) \, z(x-u, y - v) \, du \, dv.$$

It is perhaps clearer in the Fourier domain where the weights induce a filter.

An approximation to this theory can be developed numerically in a direct fashion. Figure 2 shows a population profile through a latitudinal (43°N) tier of counties in Michigan from Lake Michigan to Lake Huron (87°W to 82°W longitude). To this we have fitted a high order (100 harmonics) Fourier series, which does quite well except for the Gibbs phenomena. The sharp edges of the function would in fact require the entire infinity of Fourier coefficients. Walsh's (1923) functions probably would be more appropriate but the particular expansion is not of concern here. Finally we show the low-order expansion of the series (5 harmonics) which might correspond to the results of a low-pass filter. But this is no longer a step function. Clearly to obtain the appropriate step function we must set the new value for region i equal to

$$z^*_i = 1/(b - a) \int_a^b z^*(x) \, dx$$

where the integration extends from one edge of the county to the other. In this instance it is clear from Figure 2 that regional values are diminished if neighbors have low values and the converse is also true. This is expected of a low-pass filter. It is tempting to ask what the original function was from which the county aggregation presented to us was derived. This question is of fundamental importance and suggests the use of a type of geographical enhancement. Making patterns in census data stand out more clearly is addressed in this paper only indirectly. (See Harmon and Julesz 1973, Nordbeck and Rystedt 1970, Tobler. 1969.)

Analogously, in. two dimensions the original function z(x, y) is sampled by multiplying it by a Dirac brush, the two-dimensional equivalent of the comb function. This yields values on a lattice. They are weighted and summed in the usual manner in the spatial domain

$$g^*_{ij} = \sum_p \sum_q w_{pq} \, g_{i+p,j+q}$$

or multiplied in the frequency domain. All values falling within region k are now integrated over the irregular shape and averaged

$$z^*_k = 1/A_k \sum_{R_k}\sum g^*_{ij}$$

This entire process can be represented as a sequence of operations

$$z_k \rightarrow z(x, y) \rightarrow \text{(sampling)} \rightarrow g_{ij} \rightarrow \text{(convolution)} \rightarrow g^*_{ij} \rightarrow \text{(summing)} \rightarrow z^*_k$$

This is the procedure to be used for data such as per capita income by area! units. However, if the data are number of persons within an areal unit, then the averaging should take place before the convolution, thus working with density data. This is the case treated here. It is worth examining these steps in greater detail with an example.

Suppose there are three regions, labeled **A. B.** and **C**. in the following discretized configuration:

A	**A**	**B**	**B**	**B**	**B**
A	**A**	**A**	**B**	**B**	**B**
A	**A**	**A**	**A**	**B**	**B**
A	**A**	**A**	**A**	**B**	**B**
A	**C**	**C**	**B**	**B**	**B**
A	**A**	**C**	**C**	**C**	**B**
A	**A**	**A**	**A**	**A**	**A**

We quickly tabulate that

15 *A's* have an A to their east
4 A's have a B to their east
2 A's have a C' to their east
1 A has an external area to its east
15 A's have an A to their west
0 A's have a B to their west
0 A's have a C to their west
7 A's have an external area to their west
13 A's have an A to their north
3 A's have a B to their north
4 A's have a C to their north
2 A's have an external area to their north
3 A's have an A to their south
1 A has a B to its south
2 A's have a C' to their south
6 A's have an external area to the south

0 B’s have an A to their east
9 B’s have a B to their east
0 B’s have a C to their east
6 B’s have an external area to the east
4 B’s have an A to their west
9 B’s have a B to their west
2 B’s have a C to their west
0 B’s have an external area to the west
1 B has an A to the north
10 B’s have a B to the north
0 B’s have a C to the north
4 B’s have an external area to the north
3 B’s have an A to their south
10 B’s have a B to their south
2 B’s have a C to their south
0 B’s have an external area to the south

0 C’s have an A to their east
2 C’s have a B to their east
3 C’s have a C to their east
0 C’s have an external area to the east
2 C’s have an A to their west
0 C’s have a B to their west
3 C’s have a C to their west
0 C’s have an external area to their west
2 C’s have an A to their north
2 C’s have a B to their north
1 C has a C to the north
0 C’s have an external area to the north
4C’s have an A to the south
0 C’s have a B to the south
1 C has a C to the south
0 C’s have an external area to the south

There are 22 A’s, 15 B’s, and 5 C’s in all. Suppose that the value in each region is to be modified by weighting using a neighborhood of five cells. Let these weights he labeled E, W. N. S. and X. First convert the values to densities. using a prime to denote a value obtained after this operation. Then the modified value A* for area A is given by

$$\begin{aligned} A^* = {} & (22X + 15E + 15W + 13N + 13S)\, A' \\ & + (4E + 3N + 1S)\, B' \\ & + (2E + 4N + 2S)\, C' \\ & + (1E + 7W + 2N + 6S)\, O' \end{aligned}$$

and similarly,

$$
\begin{aligned}
B^* = {} & (0X + 0E + 4W + 1N + 3S)\,A' \\
& +(15X + 9E + 10W + 10N + 10S)\,B' \\
& +(0X + 0E + 2W + 0N + 2S)\,C' \\
& +(0X + 6E + 0W + 4N + 0S)\,O'
\end{aligned}
$$

$$
\begin{aligned}
C^* = {} & (0X + 0E + 2W + 2N + 4S)\,A' \\
& +(0X + 2E + 0W + 2N + 0S)\,B' \\
& +(5X + 3E + 3W + 1N + 1S)\,C'
\end{aligned}
$$

or, with $\mathbf{U}^t$ given by

$$
\begin{bmatrix}
22 & 0 & 0 \\
15 & 0 & 0 \\
15 & 4 & 2 \\
13 & 1 & 2 \\
13 & 3 & 4 \\
0 & 15 & 0 \\
4 & 9 & 2 \\
0 & 9 & 0 \\
3 & 10 & 2 \\
1 & 10 & 0 \\
0 & 0 & 5 \\
2 & 0 & 3 \\
0 & 2 & 3 \\
4 & 0 & 1 \\
2 & 2 & 1 \\
0 & 0 & 0 \\
1 & 6 & 0 \\
7 & 0 & 0 \\
2 & 4 & 0 \\
6 & 0 & 0
\end{bmatrix}
$$

and **V** given by

$$\begin{bmatrix} X & 0 & 0 & 0 \\ E & 0 & 0 & 0 \\ W & 0 & 0 & 0 \\ N & 0 & 0 & 0 \\ S & 0 & 0 & 0 \\ 0 & X & 0 & 0 \\ 0 & E & 0 & 0 \\ 0 & W & 0 & 0 \\ 0 & N & 0 & 0 \\ 0 & S & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & E & 0 \\ 0 & 0 & W & 0 \\ 0 & 0 & N & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & X \\ 0 & 0 & 0 & E \\ 0 & 0 & 0 & W \\ 0 & 0 & 0 & N \\ 0 & 0 & 0 & S \end{bmatrix}$$

then

$$\begin{bmatrix} A^* \\ B^* \\ C^* \end{bmatrix} = \mathbf{UV} \begin{bmatrix} A' \\ B' \\ C' \end{bmatrix}$$

If the conversion to density values is explicitly included in this, dividing by the area of each region, then

$$\begin{bmatrix} A^* \\ B^* \\ C^* \end{bmatrix} = \mathbf{UVS} \begin{bmatrix} A \\ B \\ C \\ O \end{bmatrix}$$

where

$$\mathbf{S} = \begin{bmatrix} 1/22 & 0 & 0 & 0 \\ 0 & 1/15 & 0 & 0 \\ 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 1/26 \end{bmatrix}$$

This is conveniently written, using **T = UV** as

$$\mathbf{Z}^* = \mathbf{TSZ},$$

If the weight field is given specific values then **T** is completely determined by the interaction of the weight field, the sampling density. and the shapes and adjacencies of the regions. For example. if

$$W = \begin{matrix} & N & \\ W & X & E \\ & S & \end{matrix} = \begin{matrix} & 1/8 & \\ 1/8 & ½ & 1/8 \\ & 1/8 & \end{matrix}$$

then

$$\mathbf{T} = 1/8 \begin{bmatrix} 144 & 8 & 8 & 16 \\ 8 & 98 & 4 & 10 \\ 8 & 4 & 28 & 0 \end{bmatrix}$$

This is clearly dependent on the external cells. If in-migration had been allowed

$$\mathbf{T} = 1/8 \begin{bmatrix} 144 & 8 & 8 & 16 \\ 8 & 98 & 4 & 10 \\ 8 & 4 & 28 & 0 \\ 16 & 10 & 0 & 182 \end{bmatrix}$$

would have been obtained. Since a symmetrical kernel has been used the transformation rule is also symmetrical Suppose that the values for the external area are not known. Then the weights in the last column of **T** cannot he applied. A boundary rule or condition must be supplied, which can he of several types. One rule is to assume all external values to he zero. Then **T** can be taken to he square and, if the weight field is symmetric, then **T** is also symmetric. **T** is usually diagonally dominant, which increases the likelihood of $\mathbf{T}^{-1}$ being uniquely determined. It may also be desirable to normalize the weights to take into account the “power” leaking through the boundary. Alternatively, a boundary condition may he chosen which is reflexive by modifying the weight held near the edges of the domain of interest. In the NE corner of this simple example. the weight field might he modified as follows,

```
        1/8                          0
 - - - - - - - - |            - - - - - - - - |
 1/8        ½  | 1/8           ¼     ½  |    0
           1/8 |                     ¼  |
```

so **T** now becomes

$$1/8\begin{bmatrix}155 & 9 & 12\\ 9 & 106 & 5\\ 8 & 4 & 28\end{bmatrix}$$

An alternative would be to use

```
         0
 - - - - - - - - |
 1/8       ½     | 0
                 |
           1/8   |
```

Clearly different boundary conditions lead to different results. The operation

$$\mathbf{Z^*} = \mathbf{TSZ}$$

can be applied sequentially. This repeated convolution can be written as

$$\mathbf{Z^k} = \mathbf{TSZ^{-1}} = \mathbf{(TS)^kZ}.$$

If the substantive interpretation is such that densities are not used then $\mathbf{Z^k} = \mathbf{(ST)^kZ}$ in which the regional averaging is applied after each convolution. These two interpretations yield somewhat different results. The one-dimensional matrix multiplication is usually much quicker than the two-dimensional convolution because the number of areal units is generally much smaller than the number of points in a spatial lattice of data.

The important result in this discussion is that even when the data of interest are given by irregular spatial regions an equivalent to spatial filtering can be derived. Furthermore this filtering takes a simple linear form. To apply this result to Hägerstrand's model, it is only necessary to perform one computation convolving the mean information field with a discretized map of the area. This yields the weights which convert the two-dimensional contact field into the transformation rule for the irregularly shaped regions.

The method can similarly he applied to migration studies, using a mean migration field, or other types of interaction fields. The same result holds for any problem in which weights are applied to spatial data taken on a lattice, whether this he deterministic or stochastic. Map generalization and regional forecasting are two realizations of such processes. These are illustrated using Ann Arbor as the empirical instance in Figure 3 and Tables 1, 2, and 3. It is worth noting that the transformation matrix **T** is also an aggregation operator. For example. conversion from census tracts to school districts. or grouping into larger regions can be seen as a type of spatial filtering with particularly simple kernels,

Conversely. given a matrix equation relating spatial units. this equation may he interpreted as two-dimensional spatial filtering with an unknown spatial kernel. although this kernel is not necessarily spatially invariant.

This second important result provides a direct connection between spatial filtering techniques and linear analytical regional models of recognized importance. These include demographic and exchange models and regional input output models (Gould. 1967; Isard. 1960: Jutila. 1971; Masser. 1972; Rogers. 1971; Stone, 1968; Tinline, 1970). All of these suggest investigation of the ultimate stability of the systems, their modifiability for planning purposes. and empirical estimation procedures. Suppose that Z and Z* are known. For example. we will take the population of Ann Arbor as given by census tracts for 1960 and 1970 (Table 4). The model is

$$\mathbf{Z^* = TSZ}$$

with inverse

$$\mathbf{Z} = \mathbf{S}^{-1}\mathbf{T}^{-1}\mathbf{Z}^*$$

assuming that $\mathbf{T}^{-1}$ exists. The problem is to obtain an estimate of **T** given **Z*** and **Z**, taking **S** and $\mathbf{S}^{-1}$ as knowns. The formal least-squares solution is immediately

$$\mathbf{T} = \mathbf{Z^*Z}^t(\mathbf{ZZ}^t)^{-1}\mathbf{S}^{-1}.$$

But **T** = **U·V** where **U** is determined by the spatial configuration and **V** consists of neighborhood weights. Formally, using the least-squares principle again

$$\begin{aligned}\mathbf{V} &= (\mathbf{U}^t\mathbf{U})^{-1}\mathbf{U}^t\mathbf{T}\\ &= (\mathbf{U}^t\mathbf{U})^{-1}\mathbf{U}^t\mathbf{Z^*Z}^t(\mathbf{ZZ}^t)^{-1}\mathbf{S}^{-1}.\end{aligned}$$

For estimation purposes it may be convenient to rewrite **T** as a column vector **T**’ = **U**’**V**’, $\mathbf{T}_{ij} \rightarrow \mathbf{T}'_k$ with k = (i -1)c + j. Here c is the number of columns of $\mathbf{T}_{ij}$, and (for example)

$$\mathbf{V}' = \begin{bmatrix} X \\ E \\ W \\ N \\ S \end{bmatrix}$$

and

$$\mathbf{U}' = \begin{bmatrix} 22 & 15 & 15 & 13 & 13 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 2 & 0 & 4 & 2 \\ 0 & 1 & 7 & 2 & 6 \\ 0 & 0 & 4 & 1 & 3 \\ 15 & 9 & 9 & 10 & 10 \\ 0 & 0 & 2 & 0 & 2 \\ 0 & 6 & 0 & 4 & 0 \\ 0 & 0 & 2 & 2 & 4 \\ 0 & 2 & 0 & 2 & 0 \\ 5 & 3 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

using the present simple illustrative case. Then. by least squares,

$$\mathbf{V}' = (\mathbf{U}'^{t}\mathbf{U}')^{-1}\mathbf{U}^{t}\mathbf{T}$$

In this form the (assumed) spatial stationarity of the kernel is made more explicit. The matrix **U**' is obtained as a direct count of adjacencies in the geographical data matrix. In this illustrative system there are six observations and five unknowns so that a solution is possible in principle. (Table 4 tentatively suggests that the 1970 arrangement of the Ann Arbor population might have derived from the 1960 arrangement by a neighborhood diffusion process. The kernel would not he of unit weight. however, since the 1960 population totaled 74.700 people and the 1970 value was 102.451 A rotationally invariant model would he more appropriate for city growth than a translationally invariant one [Tobler. 1970, p. 239].)

The suggestion is that it may be possible to determine the filter kernel from aggregated before and after data. which was not entirely obvious *a priori.* A spatial interpretation can thus he given to models which deal with areal units in a superficially aspatial manner. Hypotheses can also he tested concerning spatial neighborhoods and spatial invariances of various types from aggregated data. It should also he possible to use these procedures to filter categorical data given by irregularly shaped spatial regions. Changing a geological map, land-use map. or agricultural crop map into a smoothed version of the same map might proceed as follows. The new category to be assigned to an area is selected by noting the amount of ground in various categories at the neighbors of every place in the region. Next, weight these amounts by a function of the degree of neighborliness and choose the dominant weighted category to represent the region on the generalized map. The general methods discussed here can thus be extended to non-numerical situations. Extension to the rnultivariate spatio-temporal case also seems feasible.

Acknowledgment

Support for this study was provided in part by NSF grant GS 34070X and by a grant from the Institute of Environmental Quality. University of Michigan.

References

Andrews. H. 1970. *Computer techniques in image processing,* Academic Press. New York. 187 p.

Bracewell. R.. 1955. Simple graphical method of correcting for instrumental broadening, *Jour. Optical Soc. America, .* v. 45. p. 873-876.

Brown. W.. 1965. Historical note on a recurrent combinatorial problem, *American Mathematical Monthly.* v. 72. no. 9. p. 971-977.

Burks. A.. ed . 1970. *Essays on cellular automata*, Univ. Illinois Press, Urbana .375 p.

Burr. F., 1955, Sharpening of observational data in two dimensions, *Australian Jour. Physics.* v. 8, p. 30-53

Cliff. A., and J. Ord, 1969. The problem of spatial autocorrelation, in A. Scott. ed.. *Studies in regional science,* Pion. London. p. 25 55.

Cliff. A . and J. Ord. 197(1. A regression approach to univariate spatial forecasting. in M. Chisholm. ed.. *Regional forecasting,* Butterworths. London. p. 47-70.

Clowes. M., 1970. On the description of board games, in S. Kaneff, ed. *Picture language machines:* Academic Press. Ness York. p. 397-420.

Codd, E., 1968. *Cellular automata:* Academic Press. New York. 226 p.

Epstein. E., 1969. Stochastic dynamic prediction: *Tellus.* v. 21. no. 6. p. 734-751), figures 8 and 10. p. 753.

Feldt, A.,1968. *The community land use game,* Cornell Univ. Press. Ithaca. 125 p.

Gardner. M.. 1970. Mathematical games, *Scientific American.* Oct., p. 120-123.

Gould. P., 1967. On the geographic interpretation of eigenvalues: *Trans. Inst*i. *British Geographers.* V. 42. p. 53-86.

Hägerstrand. T., 1968. A Monte Carlo approach to diffusion, in B. Berry and D. Marble. eds.. *Spatial analysis,* Prentice-Hall. Englewood Cliffs. New Jersey. p. 368-384.

Harbaugh. J., and G. Bonham-Carter. 1970. *Computer Simulation in geology, W*iley Interscience. New York. p. 435-457.

Harmon. L.. and B. Julesz, 1973. Masking in visual recognition, *Science.* v. 180. no. 4091. p. 1194-1197.

Holloway. J. L., 1958. Smoothing and filtering of time series and space fields*, Advances in Geophysics.* v. 4. p. 351-389.

Howard. R.. 1971. *Dynamic probabilistic systems:* John Wiley and Sons. Inc.. New York. 576 p.

Isard. W.. 196(1. *Methods of regional analysis,* John Wiley and Sons. Inc.. New York. 784 p.

Jutila, S., 1971, A linear model for agglomeration, diffusion, and growth of regional economic activity, *Regional Science Perspectives.* v. 1. no. 1, p. 83-108.

Kraus. K.,1971. Automatische Berechnung digitalen Höhenlinien, *Zeitschrift fuer Vermessungswesen, v.* 96. no. 6, p. 234-293.

Kraus. K.. and E. Mikhail. 1972. Linear least-squares interpolation: *Photogrammetric Engineering, v.* 38. no. 10. p. 1016-1029.

Krige. D., 1966. Two dimensional weighted moving average trend surfaces for ore valuation, *Symp. on mathematical statistics and computer applications in ore valuation,*

Johannesburg. p. 13-38,
Lanczos. C., 1966. *Discourse on Fourier series:* Oliver and Boyd. Edinburgh. 286 p.
MacDougall. E.. 1972. Optimal generalization of mosaic maps, *Geographical Analysis.* v.4. no.4. p. 417-423.
Masser. I., 1972, *Analytical models for urban and regional planning,* David and Charles. Newton Abbot. 164 p.
Minsky. M, and H. Pappert. 1969. *Perceptrons,* MIT Press.. Cambridge, Massachusetts, 258 p.
Moritz. H.. 1962. Interpolation and prediction of gravity anomalies and their accuracy: Inst. Geodesy. Photogrammetry, and Cartography, Ohio State Univ., Columbus, *Rept. No*. 24, 69 p.
Moritz. H.. 1963. Interpolation and prediction of point gravity anomalies: *Annals Academiae Scientiarum Fennicae. v.* 69 IIIA, p. 1-32.
Moritz. H., 1967, Optimum smoothing of aerial gravity measurements, Ohio State Univ. Columbus. *Rept No.81, 51* p. Available from NTIS as AD654008.
Moritz, H.,1970. Eine allgemeine Theorie der Verarbeitung von Schweremessungen nach kleinsten Quadraten : *Deutsche Geodätische Kommission.* München, Heft Nr. 67A. 56 p.
Moritz, H., and W. Heiskanen. 1967, *Physical Geodesy*, W. Freeman. San Francisco. Chapter 7.
Nordbeck. S.. and B. Rystedt. 1970. Isarithmic maps and the continuity of reference Interval functions, *Geografiska Annaler,* B. 2, p.92-123.
Pitts. F.. 1967. MIFCAL and NONCEL: Two computer programs for the generalization of the Hägerstrand model to an irregular lattice: Office Naval Research. Geographic Branch, *Tech. Rept. No. 7,* ONR Task No. 389-140. Contract Nonr-l228(33), 23p.
Rapp. R., 1964, The prediction of point and mean gravity anomalies through the use of a digital computer: Inst. Geodesy. Photogrammetry and Cartographv. Ohio State Univ.. Columbus. *Rept. No.* 43. 135p.
Ratliff. F., H. Hartline. and W. Miller. 1964. Spatial and temporal aspects of retinal inhibitory interaction : *Jour. Optical Soc.,.* v. 53. no. 1 . p. 110-120.
Rhynsburger, D.. 1973. Analytic delineation of Thiessen polygons: *Geographical Analysis.* v. 2. p. 133-144.
Rogers, A.,1971. *Matrix methods in urban and regional analysis* : Holden Day. San Francisco, 508 p.
Rosenfeld. A.,1969. *Picture processing by computer,* Academic Press. New York,196p
Smith. A.. 1971, Two dimensional formal languages and pattern recognition by cellular automata, *12*th A*nnual Symp.on Switching and Automata Theory.* I.E.E.E.. Michigan State Univ.. Ann Arbor. p. 114-152.
Sokal. R., and K. Gabriel. 1969. A new statistical approach to geographic variation analysis, *Systematic Zoology,* v, 18. no. 3. p. 259-278.
Stone. L.. 1968. Stable migration rates from the multiregional growth matrix operator:*Demography.* v. 5. p. 439-442.
Switzer, P.,1969. Mapping a geographically correlated environment: Stanford Univ., *Tech. Rept 145.* Contract Nonr-225(52), (NR-342-022). 38 p. Available from NTIS as AD693982.
Switzer, P., C. Mohr. and R. Heitman. 1964. Statistical analyses of ocean terrain and

contour plotting procedures: Dept. Navy. Bureau of Ships. NObsr-8l564. SS-050. *Rep:. No.1440464, 85* p. Available from NTIS as AD601538.
Thompson. P.,1956, Optimum smoothing of two-dimensional fields: *Tellus.* v. 8. p. 384-393.
Tinline. R.,1970. Linear operators in diffusion research, in M. Chisholm. ed.. *Regional Forecasting,* Butterworths, London. p. 71-91.
Tobler. W., 1966. Numerical map generalization: Michigan Inter-Univ. Community of Mathematical Geographers. Ann Arbor. *Discussion Paper No.8,*. p.1-24. Available from University Microfilms as OP-33067.
Tobler. W.,1967. Of maps and matrices: *Jour. Regional Science Assoc..* v. 7. no. 2. p. 275-280.
Tobler. W.,1969. Geographical filters and their inverses: *Geographical Analysis.* v. I. no. 3. p. 234-253.
Tobler. W.,1970. A computer movie simulating urban growth in the Detroit region: *Economic* Geography, v. 26. no. 2, p. 234-240.
Tobler. W.. 1973. Regional analysis: time series extended to two dimensions: *Geographica Polonica, v.* 25. p. I 03-106.
Walsh, J.,1923. A closed set of orthogonal functions. *American Jour Mathematics,* v. 45. p. 5-24.
Woods, J.,1972. Two dimensional discrete Markovian fields: *I.E.E.E. Trans. on Information Theory.* v. IT-18, no. 2. p. 232-240.

†Pages 14-38 of J.C.Davis & M.J. McCullagh, *Display and Analysis of Spatial Data*, NATO Advanced Study Institute, Geography Department, University of Nottingham, J. Wiley, London

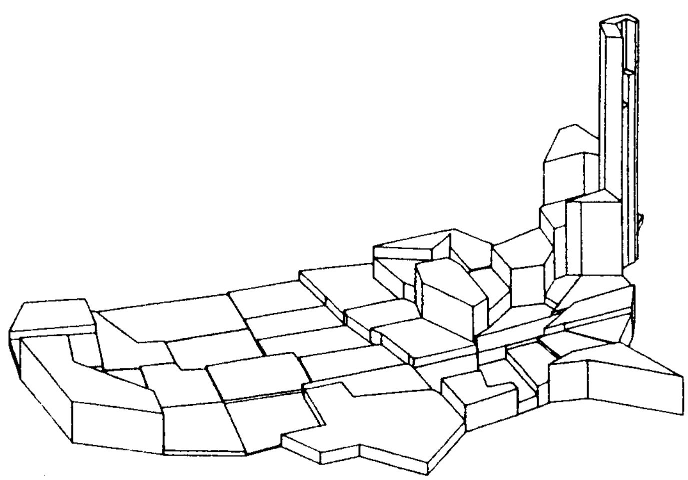
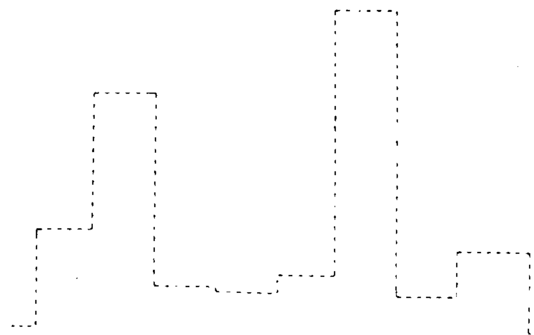
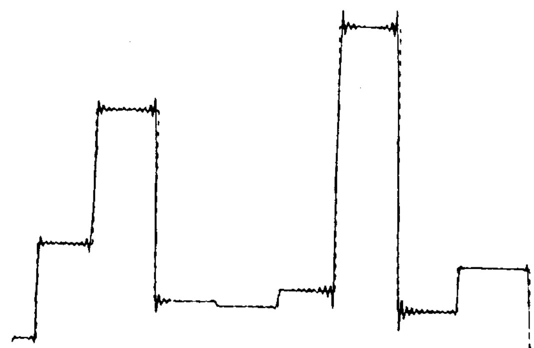


Figure 1 A piecewise continuous function z = f(x, y); 1970 population density by states


(a)


(b)

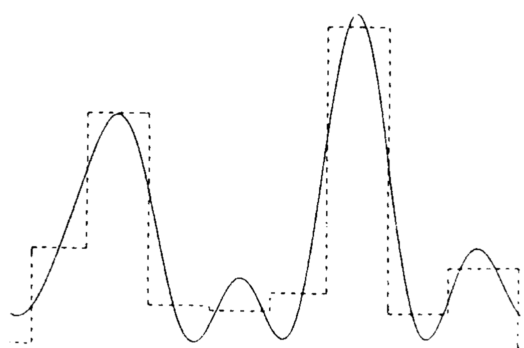
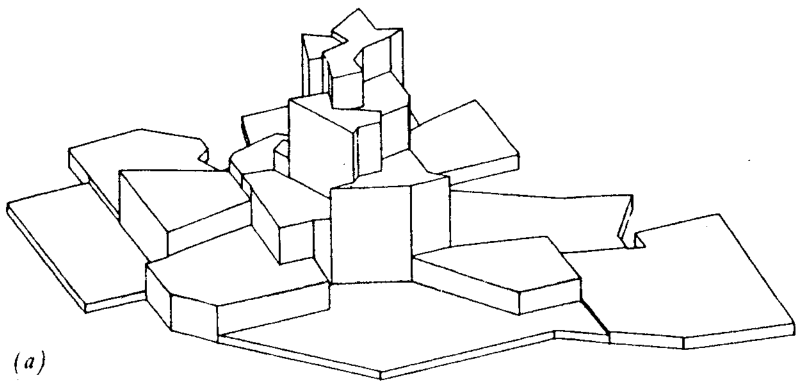
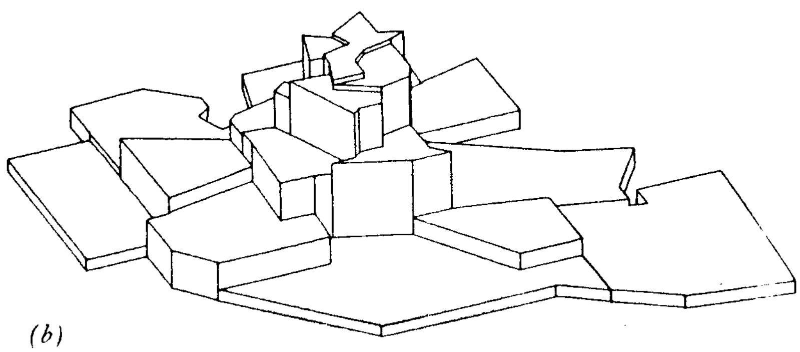

(c)

Figure 2 Latitudinal (43°N) profile (87°W to 82°W) of population density using Michigan county data, and its representation by Fourier series: 100 and 5 harmonics, respectively. Computed by J. Dozier.



(a)



(b)

Figure 3 Ann Arbor 1960 population density by census tracts shown in perspective as a piecewise continuous function. Original and linearly modified values.

Table 1 Ann Arbor Population by Census Tracts

Tract No.	1960 Population
1	1250
2	8003
3	5474
4	4042
5	4827
6	4993
7	3209
8	2736
9	2498
10	4087
11	4673
12	3596
13	3005
14	6393
15	6496
16	2075
17	4445
18	2898

Table 2. Transformation Matrix

14.1	1.2	0.0	0.0	2.5	1.5	2.7	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.2	64.9	6.9	0.0	2.5	0.0	0.0	4.3	4.2	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	0.0
0.0	6.9	54.2	2.2	2.9	0.0	0.0	0.0	0.0	4.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	2.2	55.8	3.9	1.1	0.0	0.0	0.0	4.3	0.0	2.5	5.2	0.0	0.0	0.0	0.0	0.0	0.0
2.5	2.5	2.9	3.9	44.7	6.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.5	0.0	0.0	1.1	6.4	107.9	2.1	0.0	0.0	0.0	0.0	0.0	1.7	8.7	1.5	0.0	0.0	0.0	0.0
2.7	0.0	0.0	0.0	0.0	2.1	69.6	3.1	0.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	4.8	2.1	0.0
1.0	4.3	0.0	0.0	0.0	0.0	3.1	19.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0
0.0	4.2	0.0	0.0	0.0	0.0	0.0	0.0	287.8	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.7	22.2
0.0	1.5	4.7	4.3	0.0	0.0	0.0	0.0	2.0	524.0	5.3	7.9	0.0	0.0	0.0	0.0	0.0	0.0	21.7
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.3	726.0	5.9	4.6	0.0	0.0	0.0	0.0	0.0	38.4
0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.0	0.0	7.9	5.9	200.6	14.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	5.2	0.0	1.7	0.0	0.0	0.0	0.0	4.6	14.0	693.2	10.8	0.0	0.0	0.0	0.0	29.4
0.0	0.0	0.0	0.0	0.0	8.7	0.0	0.0	0.0	0.0	0.0	0.0	10.8	305.3	7.1	2.2	0.0	0.0	8.7
0.0	0.0	0.0	0.0	0.0	1.5	2.6	0.0	0.0	0.0	0.0	0.0	0.0	7.1	184.1	8.7	7.8	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	8.7	405.2	5.8	0.0	21.6
0.0	0.0	0.0	0.0	0.0	0.0	4.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.8	5.8	468.3	5.8	22.0
0.0	3.4	0.0	0.0	0.0	0.0	2.1	1.9	10.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.8	506.0	24.4
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	22.2	21.7	38.4	0.0	29.4	8.7	0.0	21.6	22.0	24.4	3817.5

Transformation matrix for Ann Arbor using a separable symmetric Gaussian field of 49 weights obtained as the two-dimensional product of (0.016, 0.094, 0.234, 0.312, 0.234, 0.094, 0.016) with itself. The last row and column of the transformation matrix apply to the area external to the 18 census tracts. The area of each tract is proportional to the row (or column) sum for that tract since the kernel is of unit value.

Table 3. Ann Arbor 1960 Population by Census Tracts, after Filtering

Tract No.	Smoothing No. 1	Smoothing No. 2	Smoothing No. 3
1	1313	1313	1280
2	7031	6326	5700
3	5180	4861	4539
4	3611	3267	2989
5	4467	4160	3890
6	5040	5041	5005
7	3209	3179	3125
8	2364	2079	1856
9	2838	3128	3377
10	4831	5474	6029
11	4729	4790	4856
12	3406	3250	3123
13	3679	4279	4818
14	6459	6520	6572
15	6039	5655	5331
16	2369	2637	2880
17	4720	4975	5211
18	3491	4011	4470

Smoothed using the transformation **Z*** = **TSZ** as described in the text, with **T** given in Table 2, but normalized to adjust for the leakage to the external area. **S** is also implicit in Table 2.

Table 4. Percentage Arrangement of Population by Ann Arbor Census Tracts

Tract No.	1960	1960 thrice smoothed	1970
1	1.67	1.71	1.23
2	10.71	7.59	6.85
3	7.33	6.05	6.82
4	5.41	3.98	3.83
5	6.46	5.18	5.87
6	6.68	6.67	5.07
7	4.30	4.16	2.83
8	3.66	2.47	2.63
9	3.34	4.50	6.47
10	5.47	8.03	4.88
11	6.26	6.47	9.69
12	4.81	4.16	5.25
13	4.02	6.42	7.06
14	8.56	8.76	8.75
15	8.70	7.10	7.28
16	2.78	3.84	3.74
17	5.95	6.94	4.91
18	3.88	5.96	6.83