Dierk Rhynsburger

*Analytic Delineation of Thiessen Polygons**

INTRODUCTION

Imagine a finite set of $n \geqslant 3$ points, which we shall call centers, to be distributed in some manner over an infinite plane surface. The number of centers is not important, but each center must be distinct from every other. Define the spanning or bounding polygon for this distribution of centers as the convex polygon with m centers as vertices, where $3 \leqslant m \leqslant n$, constructed so that all other centers are contained within it. At a certain time, each center begins to propagate a circle at the same constant rate. The center of a circle is the center which generates it. Growth in a certain direction stops only when contact is made with a neighboring circle; otherwise it continues undiminished until the entire plane has been covered. After a while, only the circles generated by the m centers of the bounding polygon are still growing, and it is easy to see that, in the absence of a bound on the plane, they will continue to do so. The remaining circles have expanded into a set of closed polygons, one polygon about each center, which completely partitions the plane in the immediate vicinity of the distribution. These closed polygons, together with the open polygons generated by centers on the bounding polygon, partition the entire plane into n polygons. These space-filling polygons are called Thiessen polygons.

Several interesting properties of Thiessen polygons should be noted. Since the

* This paper is a revised portion of the author's unpublished Master's thesis "Analytic Delineation of Thiessen Polygons" (Department of Geography, The University of Michigan, 1972). Thanks are due to Professors Waldo R. Tobler and John D. Nystuen, and to John A. Griffin, for suggestions leading to the algorithm developed here, and to Harry F. Lins, Jr., for drawing the diagrams. Publication has been approved by the director of the United States Geological Survey.

Dierk Rhynsburger is a geographer working with the Geographic Applications Program of the U.S. Geological Survey in Washington, D.C.

circles expand at the same rate, the first contact between two circles must occur at the midpoint of the line joining their centers. It follows also that every point of continuing contact between these two circles is equidistant from the two centers; hence, the locus of these points is the perpendicular bisector of the line joining the centers. This locus is the edge between the Thiessen polygons about the two centers. If the edge is of nonzero length, the centers are said to be Thiessen neighbors through the edge. An endpoint of the edge is obtained when the developing circles of the neighboring centers encounter a circle developing about a third center. This point must be equidistant from all three centers and is, therefore, the circumcenter of the triangle defined by the three centers. Since the circles now fill the space within the triangle, no further growth occurs in the vicinity of this point. It becomes a vertex of each of the three polygons. (In the special case of half-neighbors, to be discussed later, more than three centers will have a common vertex.) By defining edges as lines equidistant from two centers and vertices as points equidistant from three or more, we have partitioned the plane so that all locations closer to a center than to any other will be included in the Thiessen polygon about that center. It is this property which renders Thiessen polygons most useful. This paper presents an unambiguous way of assigning Thiessen polygons to a set of centers, thereby allowing areal interpretation of point data.

THE NOTION OF "NEIGHBOR" IN GEOGRAPHY

Many geographical processes may be thought of as operating through neighbor relations. Only one of the many possible classes of spatial neighbors is considered here, that of Thiessen neighbors. The neighbor idea with respect to sets of points has been applied to problems of regionalization and to interpolation of values to grid intersections.

Regionalization of areas in two-space can be considered a matter of joining contiguous areas to each other. In this case, the areas are usually well defined, and contiguity is implied by their spatial relationships. When the phenomenon under consideration is punctiform, some other definition of neighbor is required. In this case, the neighbor relation can be confined to the nearest point, to the nearest several points, to the nearest point of each sector, or to points whose theoretical areas of influence adjoin that of the central point. A neighbor relation is required as well when the regionalization is performed in n-dimensional characteristic space. Here the neighbor is usually the nearest point by the Euclidean distance formula for n-dimensions, although it is possible to consider as neighbors the points whose n-dimensional hyperspheres adjoin the hypersphere about the central point.

In another class of problems, interpolation is required in order to approximate the value of a given variable at a grid point that is not an observation point. The distribution of the variable, of course, must satisfy certain assumptions of continuity. It is necessary to select from nearby observations a set which, when properly weighted, will produce the best approximation. The choice of the particular set depends on the definition of neighbor which is applied.

Unfortunately for geographers, most areal data come from observations made at points, and in most cases these points are not regularly spaced. To facilitate analysis of such data, two courses of action are available. The irregular areas over which the point observations can reasonably be expected to reflect reality can be transformed into a regular set of triangles, parallelograms, or hexagons. This is regionalization to an areal data base. Alternatively, the observation points themselves can be transformed into a regular grid. This is interpolation to a point data base. Both methods require a definition of neighbor; both can profitably utilize the space-filling polygons discussed in this paper. An application of these polygons to the first method is presented in Johnson and Thomas [*7*]; an application to the second, by an inefficient algorithm, can be found in computer program GENEB [in Tobler *14*, pp. 79–88].

ORIGIN AND USES OF THIESSEN POLYGONS

An early practical application of space-filling polygons was in the determination of precipitation averages over particular areas. The simplest method of estimating rainfall over a drainage basin is to average the rainfall at all the recording stations within the basin. This average is the arithmetic mean. If one assumes, however, that

> the amount of rain recorded at any station should represent the amount for only that region inclosed by a line midway between the station under consideration and surrounding stations [*13*, p. 1083],

then greater accuracy can be expected. This suggestion, along with a geometric example, was presented by A. H. Thiessen in the *Monthly Weather Review* for July, 1911. He did not elaborate upon his suggestion by including rules for apportioning the area, and in fact there is an error in the computations for his example; the idea, however, had been published.

Six years later R. E. Horton presented a fully developed procedure for the construction of space-filling polygons in drainage-basin precipitation situations [*6*]. He chose to call them Thiessen polygons, crediting Thiessen with the independent suggestion of the idea, and it is this work which first defined the Thiessen mean.

Whitney suggested that even the Thiessen mean can be improved upon [*15*]. Its assumption that rainfall is distributed equally over the entire polygon area will yield errors which can be lessened by considering the triangles originally used for the polygon construction to be the component planes of a rainfall surface. This, the inclined plane method, allows for spatial variation of rainfall as well as for the influence of stations near but not in the basin.

These three methods for areal averaging of precipitation are described in varying detail in textbooks of hydrology. For the Thiessen method, most of these texts include only an example with a diagram and a short discussion [*3*, pp. 28–32; *10*, pp. 224–27]. Johnstone and Cross, although including no diagram in their discussion of the method, allude to the uncertainty in polygon construction:

> Although there is a multiplicity of choice in drawing the triangles initially, there is no ambiguity in the finally resulting polygons. . . . If the polygons are drawn correctly, there will be no re-entrant angles. [*8*, p. 46]

That is, the polygons will be convex. Wisler and Brater take this a step further by using an ambiguous situation as a graphic example [*16*, pp. 85–89]. Their conclusion is that trial and error is the best approach to resolving the ambiguities.

Thiessen polygons have been utilized by geographers as well. Snyder has applied them to the department capitals of Uruguay in an effort to devise a more flexible approach toward defining an urban hierarchy than Christaller's hexagons provide [*12*]. Dahlberg has constructed "dot polygons" about the points of a dot distribution, using these as the basis for a choropleth map of the distribution's density [*2*]. Bogue has adapted Thiessen polygons as a first approximation to the hinterlands of metropolitan centers in the United States [*1*, pp. 17–18], while Johnson and Thomas suggest their use as a reasonable approximation to areas for which a point observation can be assumed to be representative [*7*].

Early Algorithms

Traditionally, Thiessen polygons have been constructed with compass and straightedge. The generating points were connected with straight lines into a set of nonintersecting triangles upon the sides of which perpendicular bisectors were erected. With judicious choice of triangles the bisectors would "merge to form polygons" [Kopec, *9*, p. 25], but occasionally the generating points would be so arranged that quadrilaterals occurred which could not be divided unambiguously into triangles. The rule usually given in this case was that the shorter of the two diagonals would properly subdivide the quadrilateral. Although it was admitted that this rule did not always apply, no clear statement of the conditions under which it did was ever made. Kopec, using the example given in Wisler and Brater [*16*, pp. 88–89], did find that by imagining the existence of the triangles instead of drawing them, the actual construction of polygons was much neater; but this contribution succeeded in obscuring rather than in resolving the ambiguity, and students continued to discover overlapping polygons and re-entrant angles in their exercises.

With the development of the computer, several approaches which previously had been impractical became feasible. For example, the space containing the observations can be overlaid with a fine grid at each intersection of which the distances to all observations are calculated. The minimum distance then gives the observation in whose polygonal neighborhood the intersection is located. Intersections with two or more closest observations are then on edges or vertices of Thiessen polygons, and these can be joined to approximate the desired boundaries.

A more refined algorithm, which has been programmed by R. Gambini, increments along the perpendicular bisectors of the lines joining all $n(n-1)/2$ generating pairs [*4*]. For any pair a point from which the distance to all other generators is greater than the distance to both members of the generating pair is sought; if one is found, a polygon edge has been encountered, and the program proceeds to locate within a given tolerance the coordinates of the endpoints of this edge. Although the procedure is time consuming, polygon edges and intersections are produced. This method was recently used by the author with the comment that no better procedure to find geographical neighbors is known [*14*, p. 79].

A better procedure was discovered independently of this work by B. Shelton, who produced a computer program to define all nonzero length edges of Thiessen polygons [*11*]. It incorporates a routine for the elimination of centers clearly not neighbors of the center in question, a potential improvement in efficiency not considered in the programming of the algorithm developed here. It does not, however, isolate the half-neighbors which develop when more than three polygons have a common vertex (see Figure 5 for an example).

Some Theorems for a New Algorithm

An efficient procedure for the determination of Thiessen neighbors is presented here. Unlike Gambini's method, it is analytic rather than iterative, and it does not rely upon an arbitrary choice of search increment and error tolerance. From an initial point located within[1] a polygon edge all vertices of the polygon can be found in an efficient, precisely defined manner. The only error is the roundoff error inherent in the machine executing the algorithm.

Definition. *The Thiessen polygon about a generating center is the interior envelope of the perpendicular bisectors of the lines joining the center to all other centers.*

Let I in Figure 1 be the center in question. Erect at Z, the midpoint of IJ, a perpendicular to IJ, ZA. All locations below this line are clearly closer to center J than to center I, for the perpendicular bisector ZA is the locus of points equidistant from I and J. Construct the perpendicular bisectors of the lines joining I to the remaining centers. In each case, locations in the half plane not containing I will be closer to the other center than to I. These locations can not be within the polygon about I. Thus, the innermost closed polygon formed by these lines about center I (here termed the interior envelope) will contain all locations closer to I than to any other center. This is by definition a Thiessen polygon. The intersections P, Q, R, and S of the lines forming the envelope with each other are the vertices of the polygon as well as the endpoints of the edges of the polygon. Line segments PQ, QR, RS, and SP are the edges of the Thiessen polygon generated by I, and the points J, L, M, and N are the Thiessen neigh-

[1] To be contained within an edge is to fall in the open interval defining the edge. If the point is in the closed interval, the point is said to be on the edge.

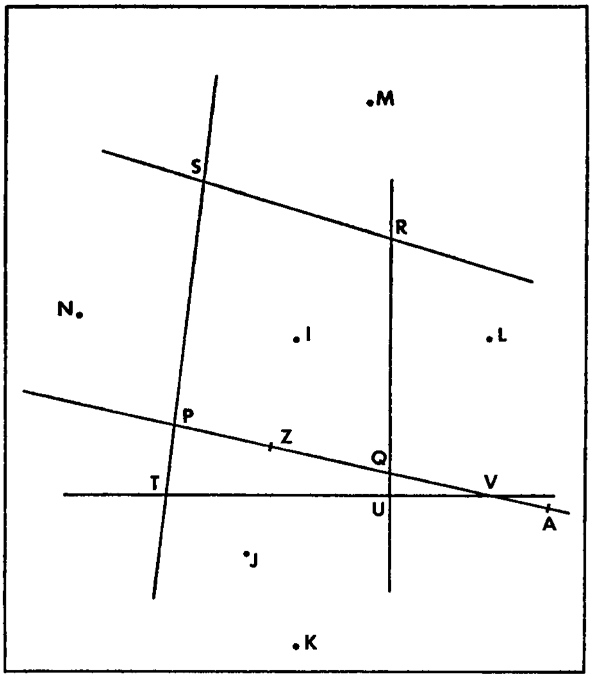

FIG. 1. Thiessen Polygon PQRS about Center I.

bors, through these shared edges, of I. Observe that intersections T, U, and V fall on the boundary between I and K. Since line TV is not part of the interior envelope about I, it does not contain an edge of the polygon. Center K is not a Thiessen neighbor of center I, and line TV can be said to contain an imaginary polygon edge.

Special case. If the center is a vertex of the bounding polygon for the point set, the perpendicular bisectors will not close into an envelope. They will, nevertheless, form the polygon edges as in the general case—only here the polygon is open.

THEOREM 1. *The nearest center to a given center is always a Thiessen neighbor, and the midpoint of the line joining the two is always contained within the polygon edge between them. This applies as well when two or more centers are equally near.*

Proof. From the definition of Thiessen polygons it follows that all polygon vertices are intersections of perpendicular bisectors of lines joining neighboring centers. Define J as the nearest neighbor of I, and draw a circle of radius IJ about center I (Figure 2). By the definition of nearest neighbor, there is no center K such that $IK < IJ$. Let VW be the perpendicular bisector of IJ through midpoint U. To divide the plane, extend indefinitely in both directions the line passing through I and J. Locate center K at any point in the right half plane

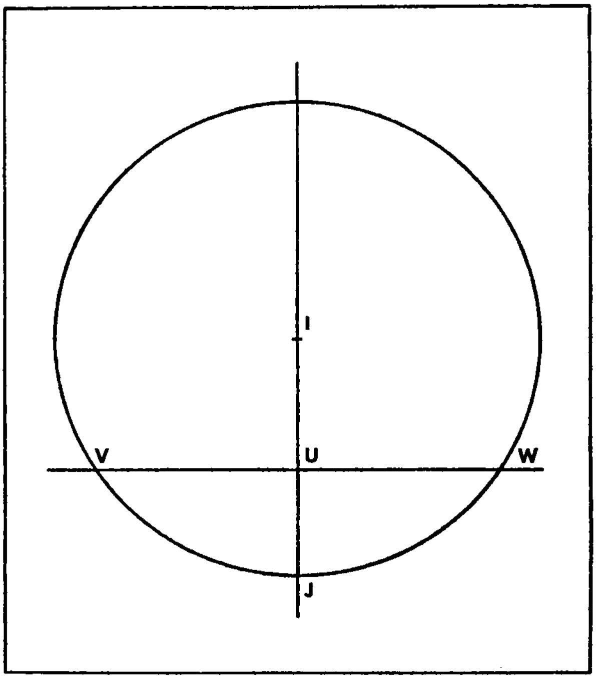

FIG. 2. Construction for the Proof of Theorem 1.

which is on or outside the circle. Construct the perpendicular bisector of IK and observe that, under the constraints imposed upon the location of K, the intersection of this bisector with VW falls to the right of U, and the edge between polygons about I and J extends in the direction of V. Correspondingly, if K is in the left half plane and on or outside the circle, the intersection falls to the left of U on VW, and the polygon edge extends toward W. If K is chosen to be on the line dividing the plane, but not inside the circle, the bisectors are parallel, and K is a neighbor of I or of J but not both. Thus, when centers I, J, and K are mutual Thiessen neighbors, the edge common to polygons about I and J has a nonzero length and must contain the midpoint of the line joining the two centers. The only ambiguous case arises when J and K are coincident, but this is disallowed by the restriction that all the centers be distinct.

THEOREM 2. *Given one point known to be within a polygon edge, all endpoints of this edge can be found.*

Proof. Suppose, as in the previous theorem, that we are describing the polygon about I and know that U is within the edge shared with the polygon about J. J need not be the nearest neighbor, but it must be a Thiessen neighbor. Since we are given that U is not an endpoint of the edge, intersections of the perpendicular bisectors of the lines joining I to all centers other than J, as A, B, C, D, and E in Figure 3, will not coincide with U. Since the Thiessen polygon is the interior envelope of the lines, and since U must be within the edge, it is clear that the necessarily nonzero minima of $UA, UB, \ldots, UE$ in each direction from U along the line AE will be the distances to the vertices of the polygon edge. In the event the edge runs to infinity, there will be no intersection on one side of U. In the example of Figure 3, B and C are the endpoints of the edge between polygons I and J.

THEOREM 3. *Given a polygon edge and one of its endpoints, the adjacent edge can be found.*

Proof. Suppose we have, as in Figure 4, a real edge between centers I and J running from some known point U to vertex C. U may be the midpoint of line IJ, it may be the other vertex of the edge, or it may be an arbitrary point known to be within the edge. The case in which there is only one perpendicular bisector intersecting AE at C is trivial. Suppose that there are two perpendicular bisectors intersecting each other and AE at C, namely, CY and CZ. This is the multiple neighbor case in which one of the Thiessen neighbors will be defined as a half-neighbor of I. Since the distances along AE from U to the intersections of lines CY and CZ are equal, there is no immediate way, other than by graphic means, to determine which line contains the adjacent edge. Draw line BF through J, the last found neighbor, parallel to AE, the last found line containing a polygon edge. Draw a line from I through C, the last found vertex, to its intersection

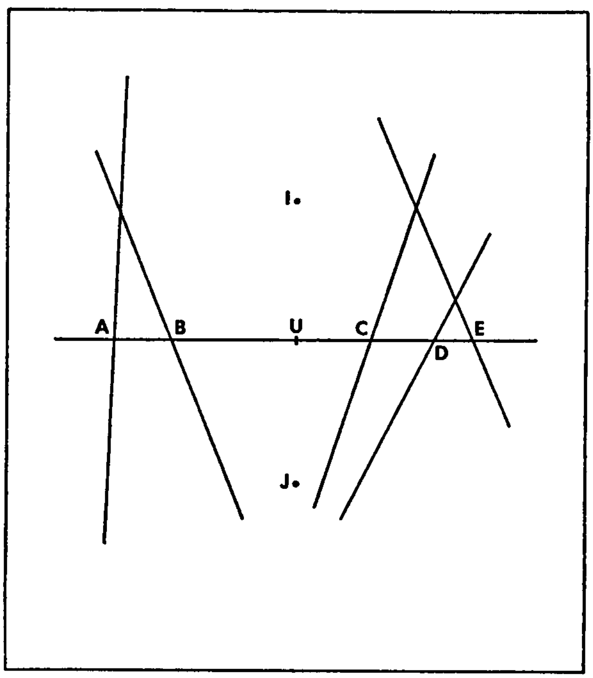

FIG. 3. Locating Polygon Vertices Given a Point within the Edge.

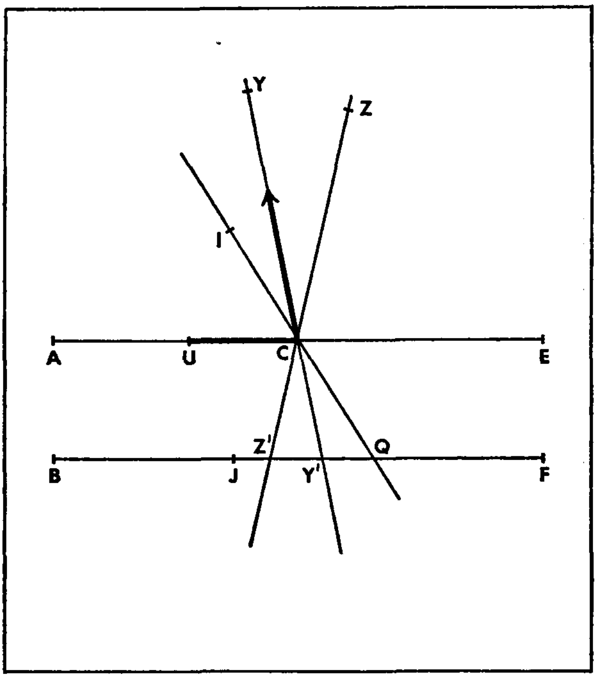

FIG. 4. Construction for the Proof of Theorem 3.

Q with BF. No edge of the polygon about I will fall within the angle UCI; thus angle ICE will include all of the lines through C which may contain the adjacent edge. Because the interior envelope about I has no lines passing through it, it is clear that the smaller of angles ICY and ICZ will indicate the adjacent side. This can be translated to the smaller of distances QY' and QZ'. These distances are always finite, since the polygons are convex, and always greater than zero, since the polygon edge may not pass through the polygon center. Equal distances are precluded by the requirement that all centers in the point set of interest be distinct. We have here a simple linear method for isolating the one adjacent edge from the (possibly many) zero length edges between the polygon and its half-neighbors, as well as a method for determining the direction of the next vertex. The adjacent edge will run from vertex C in the direction away from Y' or Z'.

THE ALGORITHM

The foregoing theorems allow unambiguous delineation of Thiessen polygons. Thus, for closed polygons, all vertices can be found by proceeding in one direction around the polygon beginning with the midpoint of the line to the nearest neighbor. Polygons with two semi-infinite edges (edges with only one end-point)

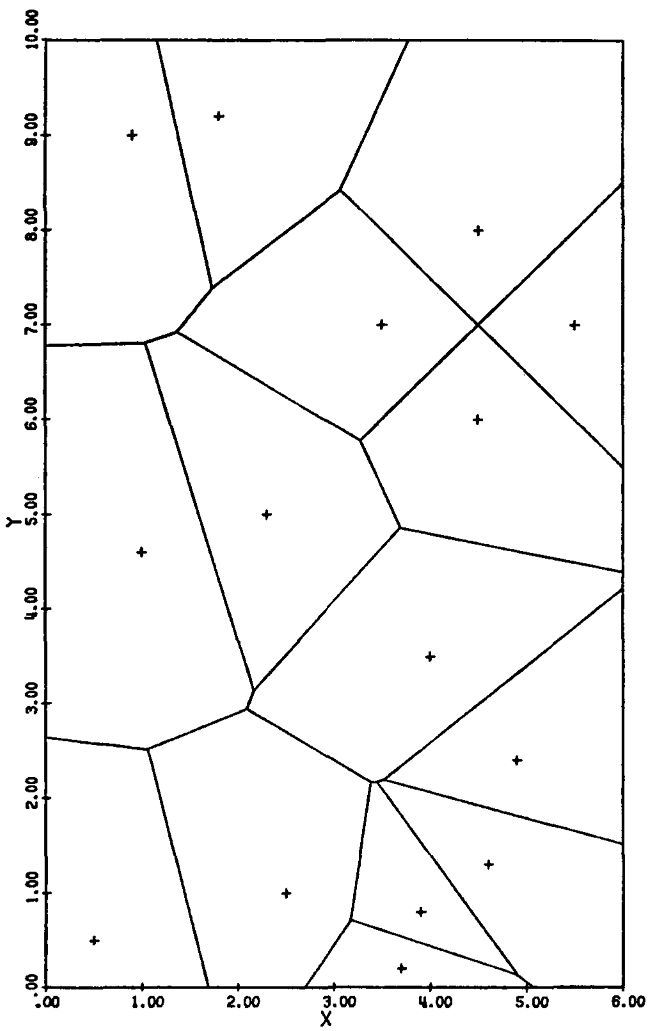

FIG. 5. Sample Plot of Thiessen Polygons Including Two Pairs of Half-Neighbors.

can be bounded by proceeding in both directions from the known midpoint until no additional vertex is encountered.

A computer program, which can be obtained from the Geography Program Exchange at Michigan State University, has been written in FORTRAN for the execution of the algorithm. It is a first approximation to an elegant solution to the Thiessen polygon problem, and as such it does not reflect the highest state of the programmer's art. It does, however, incorporate several interesting relationships for purposes of streamlining and greater efficiency of operation.

As an example, each line is referred to by the three parameters a, b, and c

obtained when the equation of the line is written in the form $ax + by + c = 0$ [*5*]. This has the advantage over the traditional form $y = mx + b$ of having no singularities: every possible line, regardless of slope, can be expressed as an ordered triple of finite numbers.

By use of the relationship $u = ay - bx$, every point (x,y) on a single line (a,b,c) can be assigned a unique value u. Comparing the magnitude and sign of a difference in u values is analogous to, and faster than, the calculation of distance and direction in the traditional Cartesian distance formula.

The point of intersection of line (a,b,c) with line (p,q,r) is given by

$$x = \frac{br - cq}{aq - bp}, \qquad y = \frac{cp - ar}{aq - bp}. \tag{1}$$

Clearly, as the denominator approaches zero, the lines become parallel. The parameters for line (l,m,n) parallel to (a,b,c) through the point (x,y) are likewise easily represented. It can easily be shown that (l,m,n) is equivalent to $(a,b,-ax-by)$.

It can be seen that this modified use of homogeneous coordinates leads to the elimination of the special case of lines with infinite slope, as well as to an overall increase in the efficiency of the algorithm.

To plot semi-infinite edges a surrogate endpoint is required. Since the convexity of the bounding polygon for the point set assures that perpendicular bisectors of the lines joining adjacent vertices will not intersect outside that polygon, adjacent vertices are Thiessen neighbors with semi-finite edges. Divide the plane by extending in both directions the line joining a pair of adjacent centers. Since this line and the half plane containing the bounding polygon contain all centers, the semi-infinite edge must be directed into the other half plane. The surrogate endpoint is the intersection of the edge with an arbitrary rectangular boundary in this direction.

The program plots the initial point set and all polygon edges within the rectangular boundary. An example plot is shown in Figure 5.

LITERATURE CITED

1. Bogue, D. J. *The Structure of the Metropolitan Community.* University of Michigan Contributions of the Institute for Human Adjustment Social Science Research Project, 1949.
2. Dahlberg, R. E. "Towards the Improvement of the Dot Map." *International Yearbook of Cartography,* 7 (1967), 157–67.
3. de Wiest, R. J. M. *Geohydrology.* New York: John Wiley & Sons, Inc., 1965.
4. Gambini, R. "A Computer Program for Calculating Lines of Equilibrium between Multiple Centers of Attraction." Center for Regional Studies, University of Kansas, Lawrence, no date.
5. Hille, E. *Analysis.* Vol 1. New York: Blaisdell Publishing Co., 1964.
6. Horton, R. E. "Rational Study of Rainfall Data Makes Possible Better Estimates of Water Yield. . . ." *Engineering News-Record* (1917), pp. 211–13.
7. Johnson, L. J. and A. T. Thomas. "A Method for Transforming Geographic Data." *Proceedings,* Association of American Geographers, 4 (1972).

8. JOHNSTONE, D. and W. P. CROSS. *Elements of Applied Hydrology.* New York: Ronald Press, 1949.
9. KOPEC, R. J. "An Alternative Method for the Construction of Thiessen Polygons." *Professional Geographer,* 15 (1963), 24–26.
10. MEAD, D. W. *Hydrology: The Fundamental Basis of Hydraulic Engineering.* New York: McGraw-Hill Book Co., 1950.
11. SHELTON, B. W. "A Computer System for the Generation of Thiessen Polygon Descriptions." Master's thesis, Oklahoma State University, 1971.
12. SNYDER, D. E. "Urban Places in Uruguay and the Concept of a Hierarchy." *Festschrift: Clarence F. Jones.* Northwestern University Studies in Geography, No. 6, 1962, pp. 29–46.
13. THIESSEN, A. H. "Precipitation Averages for Large Areas." *Monthly Weather Review,* 39 (1911), 1082–84.
14. TOBLER, W. R. *Selected Computer Programs.* A Michigan Geography Publication, 1970.
15. WHITNEY, E. N. "Areal Rainfall Estimates." *Monthly Weather Review,* 57 (1929), 462–63.
16. WISLER, C. O. and E. F. BRATER. *Hydrology.* New York: John Wiley & Sons, Inc.; London: Chapman & Hall, Ltd., 1949.