Geografiska Annaler: Series B, Human Geography

ISSN: 0435-3684 (Print) 1468-0467 (Online) Journal homepage: www.tandfonline.com/journals/rgab20

# The Standard Deviational Ellipse; An Updated Tool for Spatial Description

Robert S. Yuill

**To cite this article:** Robert S. Yuill (1971) The Standard Deviational Ellipse; An Updated Tool for Spatial Description, Geografiska Annaler: Series B, Human Geography, 53:1, 28-39, DOI: 10.1080/04353684.1971.11879353

**To link to this article:** https://doi.org/10.1080/04353684.1971.11879353

Published online: 08 Aug 2017.

# THE STANDARD DEVIATIONAL ELLIPSE; AN UPDATED TOOL FOR SPATIAL DESCRIPTION

BY ROBERT S. YUILL

Department of Geography, State University of New York at Buffalo

## Definition of the measure

A persistent challenge to the field of geography is the effective description of areal point data. In recent decades, this type of data has had considerably more widespread usage due to the increased use and development of quantitative measurements. The description of these data sets has concomitantly become of greater consequence to the geographer.

Geographic description of areal point data has usually been in the form of one or some combination of three measures: average location, dispersion, and orientation or arrangement in space. The first measure is a description of the data set by a single point or location, such as the mean, mode, or median. The point is functionally related to and determined by the location and value of each point of the areal set. It defines or is presumed to be the most representative single position of all locations in the area occupied by the point set. A summary of the uses and restrictions of the measures in this category can be found in Chapter IV of Neft's monograph.[1]

Areal dispersion of a point set may be measured in a number of ways. A common method is some average distance among the points such as the various nearest neighbor statistics.[2]

There are also measures of frequency distributions based upon point counts within areal units or quadrats.[3] These statistics are often used independently and indicate the general dispersion, but possess few directional properties. Bachi's 'standard distance', on the other hand, represents some degree of spatial orientation in that it can be used as a measure of dispersion from a single point, usually an average location measure.[4]

A drawback to many of these statistics and indices is that they do not represent the orientation or shape of a point set which is often one of the more significant aspects of that distribution. One measure which does show orientation as well as the other desired properties is the standard deviational ellipse. Although this measure was devised by Lefever[5] in 1926, until recently it has attracted little attention among geographers.[6]

Lefever's original definition of the standard deviational ellipse was relatively simple. His procedure was to first determine the mean areal center of the point set and define this as the origin of a new set of axes for the distribution. The standard deviational statistic ($\sigma$) was then calculated orthogonal to each axis (Equation 1) and plotted as a vector

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \mu)^2}{n}} \qquad \text{Eq. 1}$$

($\mu$ = zero since it is defined as the origin.) originating at the mean areal center, the scalar

[1] Neft, David S. *Statistical Analysis for Areal Distributions.* Philadelphia: Regional Science Research Institute, 1966.

[2] King, Leslie J. *Statistical Analysis in Geography.* Englewood Cliffs: Prentice-Hall, 1969, p. 89.

[3] See for example: Arthur Getis. "Temporal Land-Use Pattern Analysis with the Use of Nearest Neighbor and Quadrat Methods," *Annals of the Association of American Geographers,* LIV, No. 3 (1964), pp. 391–99.

[4] Bachi, Roberto, "Standard Distance Measures and Related Methods for Spatial Analysis," *Papers of the Regional Sience Association.* X (1963), p. 88.

[5] Lefever, D. Welty, "Measuring Geographic Concentration by means of the Standard Deviational Ellipse," The American *Journal of Sociology,* XXXII, No. 1 (1926), pp. 88–94.

[6] The author is indebted to Professor Waldo Tobler of the University of Michigan for bringing Lefever's work to his attention.

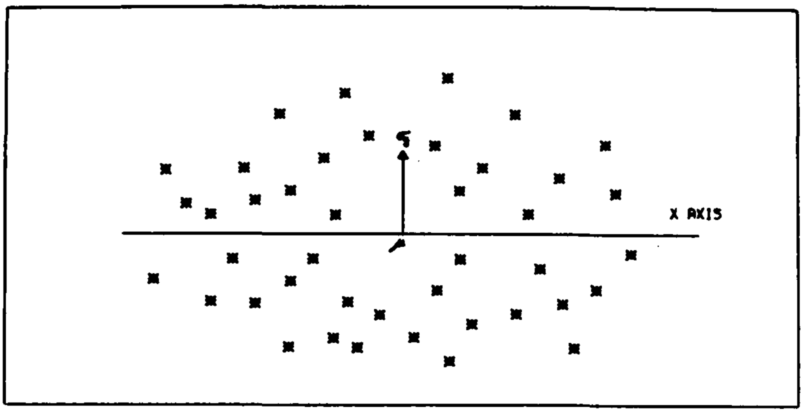


Fig. 1 Standard Deviation Vector

length being equal to the value of $\sigma$. Figure 1 shows the mean areal center ($\mu$) of a point set and the standard deviation ($\sigma_y$) calculated with reference to the $X$ axis.

The axes were then rotated about the mean with the standard deviation being calculated for each new position of the axes. For any new position of the $X$ axis ($X'$) at angle $\Theta$, Equation 1 is modified:

$$\sigma_{y'} = \sqrt{\frac{\sum_{i=1}^{n} (y'_i - \mu)^2}{n}} \qquad \text{Eq. 2}$$

$$\text{Where: } y'_i = y_i \cos\Theta - x_i \sin\Theta$$

$$\mu = 0$$

Figure 2 shows the vectors ($\sigma_y$ and $\sigma_{y'}$) plotted for the $X$ and $X'$ axes respectively.

The trace of these vectors as the axes were rotated generated an ellipse, according to Lefever, with the major axis showing the direction of maximum dispersion of the point set and the minor showing that of the minimum dispersion. The area of the ellipse indicated the concentration or scatter of the system.

Furfey, however, cast doubt on the validity of the measure by pointing out that the trace of the vectors so calculated is not necessarily an ellipse, particularly in the case where the point set has a linear distribution (Figure 3). It then followed that the standard formula for the ellipse ($\pi ab$) was obviously inappropriate for determining the area within the curve generated by the trace. Furfey concluded that as formulated, the standard deviational ellipse was a poor graphic representation of a spatial distribution.[7]

[7] Furfey, Paul Hanley, "A Note on Lefever's Standard Deviational Ellipse," *The American Journal of Sociology*, XXXIII, No. 1 (1927), pp. 94–98.

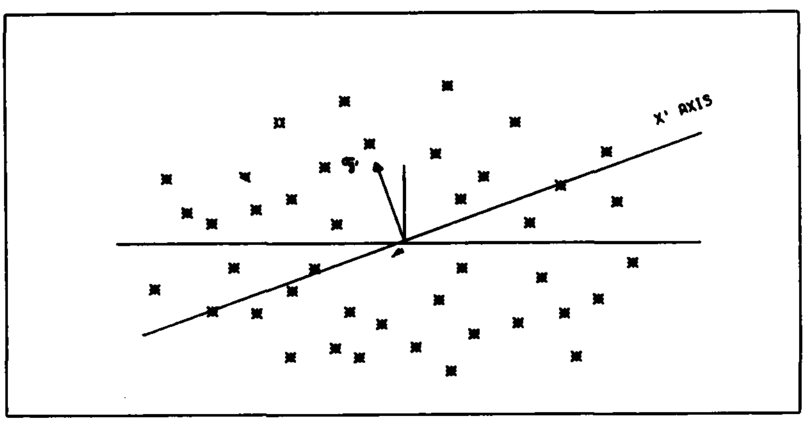


Fig. 2 Rotated Axis and Vector

An *ellipse*, however, can be a very useful graphic representation of areal data. It has the intrinsic properties of showing average location, dispersion, and orientation of a point set in a relatively simple and clear manner as its use in the several studies of Bashur, Shannon, et al. illustrates.[8] Consequently, the attempt is made here to redefine the standard deviational ellipse taking into consideration the various objections which have been raised about Lefever's measure.

The redefinition follows Lefever in the main procedure of calculation, the initial step being the calculation of an average location for the distribution. The coordinates of the point set are then transformed ($y_i = y_i - \beta_y$ and $x_i = x_i - \beta_x$) since the point of average location ($\beta$) becomes the new origin. The standard deviation from the $X$ axis is then calculated by Equation 3:

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \beta)^2}{n}} \qquad \text{Eq. 3}$$

Where: $\beta = 0$

[8] Bashur, Rashid L., Gary W. Shannon, Stanley E. Flory, and Ralph V. Smith, "Community Interaction and Racial Integration in the Detroit Area: An Ecological Analysis," Report for Project No. 2557, U.S. Office of Education, 1967. Also Rashid L. Bashur, Gary W. Shannon, and Charles A. Metzner, "The Application of Three-Dimensional Analogue Models to the Distribution of Medical Care Facilities," Paper presented to the 97th Annual Meeting of the American Public Health Association, Philadelphia, 1969.

The equation may be cast in a more general form since the standard deviation is to be calculated as the axis is rotated about the point of average location. For any angle of rotation of the $X$ axis ($\Theta$), the $Y$ distance of any $i$ th. point is expressed as ($y_i = y_i \cos \Theta - x_i \sin \Theta$). Equation 3 thus becomes:

$$\sigma_{y'} = \sqrt{\frac{\sum_{i=1}^{n} (y'_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i \cos \Theta - x_i \sin \Theta)^2}{N}} \qquad \text{Eq. 4}$$

The derivation here differs from Lefever in that we are concerned with only two values of the standard deviation (the maximum and the minimum) as the axis is rotated through 180°. It has been stated by both Lefever[9] and Bachi[10] that these values occur at an angle of 90° to each other. Therefore the maximum and minimum values of Equation 4 may be used to *define* an ellipse with the major and minor axes corresponding respectively to the maximum and minimum standard deviations. Since these values alone completely define an ellipse, the measure would be applicable to any point set. Furfey's exceptional case (Figure 3) would then become a boundary condition where the minor axis would be zero and the ellipse collapses to a line. The antithetical

[9] Lefever, *op. cit.*, p. 90.
[10] Bachi, *op. cit.*, p. 88.

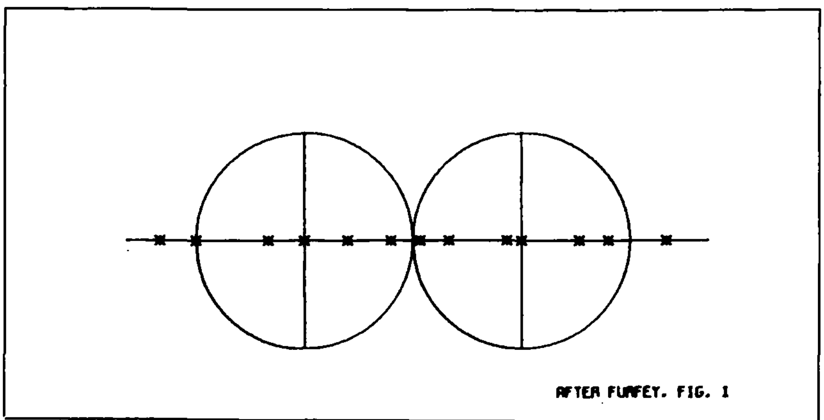


Fig. 3 Lefever's Ellipse Trace with Linear Data

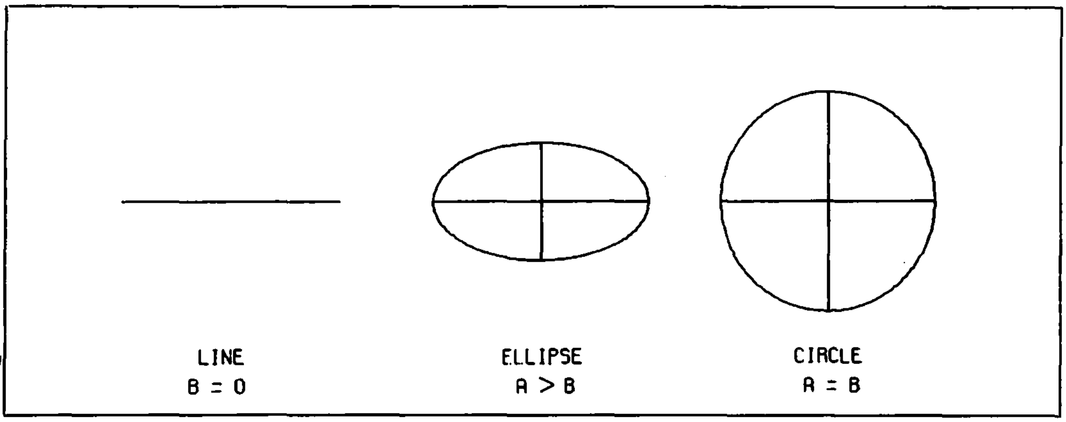


Fig. 4 Normal and Boundary Conditions of the Standard Deviational Ellipse

boundary condition would be when the point set was evenly dispersed about the average location so that the major and minor axes were equal. The ellipse in this case becomes a circle. Figure 4 shows the normal case and boundary conditions of the measure.

There is an arithmetic solution for finding the maximum and minimum standard deviations and the corresponding angles. Equation 4 is first expanded (Equation 5) and its first derivative taken. If this derivative is equated

$$\sigma_{y'} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 \cos^2 \Theta - 2 \sum_{i=1}^{n} x_i\, y_i \sin \Theta \cos \Theta}{n} + \frac{\sum_{i=1}^{n} x_i^2 \sin^2 \Theta}{n}} \qquad \text{Eq. 5}$$

to zero and solved for *0* (Equation 6), the solution will show the angle at which the

$$\frac{d\sigma_{y'}}{d\Theta} = \frac{-\sum_{i=1}^{n} y_i^2 \cos \Theta \sin \Theta - \sum_{i=1}^{n} x_i y_i \ (\cos^2 \Theta - \sin^2 \Theta) + \sum_{i=1}^{n} x_i^2 \cos \Theta \sin \Theta}{n\sigma_{y'}} = 0 \qquad \text{Eq. 6}$$

maxima and minima occur. The specific solution (as Lefever has shown) is in the form of a quadratic equation:

$$\tan \Theta = -\frac{\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right) \pm}{2 \sum_{i=1}^{n} X_i\, Y_i} \frac{\pm \sqrt{\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right)^2 + 4\left(\sum_{i=1}^{n} x_i y_i\right)^2}}{2 \sum_{i=1}^{n} x_i y_i} \qquad \text{Eq. 7}$$

Of the two solutions to this equation, one represents the angle of maximum deviation and the other, that of the minimum. These, plus the values of Θ substituted into Equation 4, yield the required information for the definition of the ellipse.

A further aspect of the redefined measure is that the formula for the area within the ellipse is the standard form ($\pi ab$). This simplifies the calculation of the various indices which show areal concentration of the data.

## Interpretations and limits

With the ellipse measure mathematically defined, its specific relationship to the properties of average location, dispersion or concentration, and orientation must be determined. This section considers the interpretations and limitations of each of the three properties as applied to and by the ellipse.

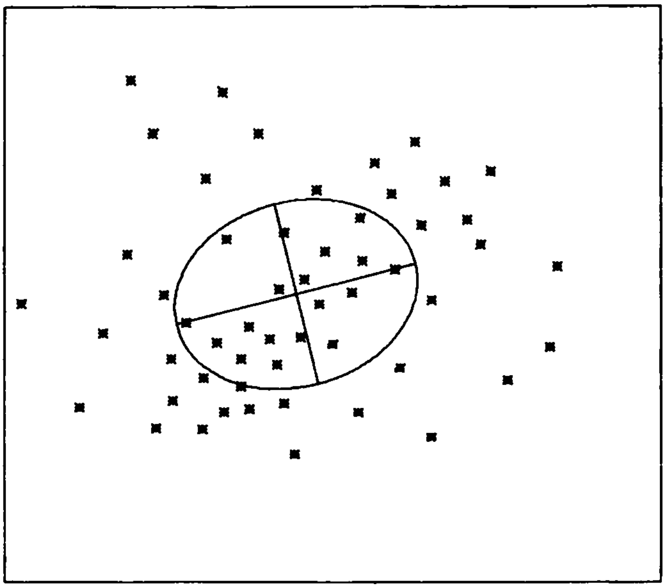


Fig. 5a Farms Producing Hay

## Average Location

There are a number of methods for calculating a point of average location for a distribution of points. Of these, the weighted mean and the median are employed here. These particular measures as defined here calculate a unique average location dependent upon only the data set in question. In addition they may be applied to either nominal or ratio scale distributions.

The weighted mean areal center is a ratio scale measure. It may be calculated for a point set using any pair of orthogonal axes. The equation has the form:

$$\beta_{x,y} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} \qquad \text{Eq. 8}$$

The value ($w_i$) represents the *quantity* of the phenomenon being measured at the $i$ th point. The point of average location is thus obviously influenced by the value as well as the location of a point.

There are two types of distributions in which the points are effectively not weighted. The first is when each point shows the location of some unit value of the phenomenon (such as one dot being equal ot *one* or *n* items). The second is a change to nominal scale where the occurrence of a phenomenon at a location is recorded. In both cases Equation 8 reduces to:

$$\beta_{x,y} = \frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{Eq. 9}$$

Equations 8 and 9 may be used to show two different facets of a distribution. Figure 5a shows the mean areal center and ellipse of the distribution of hay producing farms (nominal scale) in a hypothetical region. In contrast, the mean areal center (weighted) of hay production shows that production is more concentrated in one locality than the distribution of farms would indicate (Figure 5b). The ellipse of hay production emphasizes this difference by showing a completely different orientation than the ellipse of farm distribution.

The median areal center[11] does not correspond to the linear definition of the median. The linear definition applied to areal data fails to establish a unique location, for the

[11] Mr. Donald Brothers, a geography graduate student at S.U.N.Y. Buffalo made the initial suggestion that the median be used as a center of the ellipse.

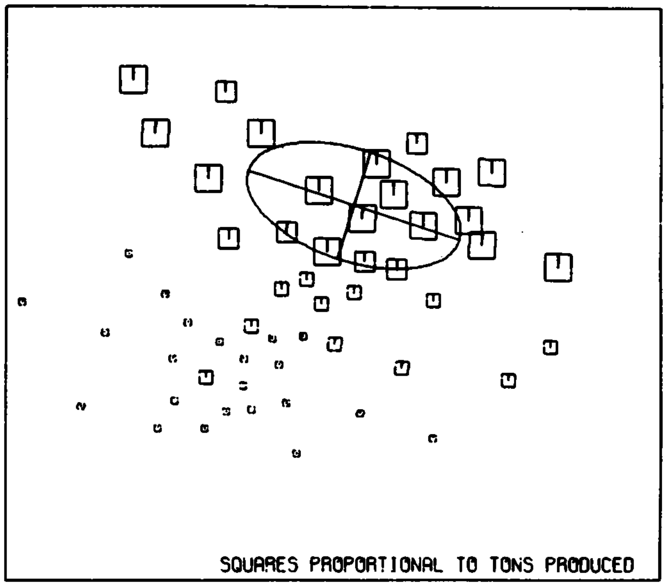


Fig. 5b Production of Hay

point so determined will vary with the position of the axes.[12] A more viable definition is that used by Bachi,[13] Court,[14] and others, the median being defined as the point of minimum aggregate travel.

The median ($\beta_m$) is expressed by finding the minimum value of the equation:

$$\beta_m = \sum_{i=1}^{n} [(x_m - x_i)^2 + (y_m - y_i)^2]^{1/2} \quad \text{Eq. 10}$$

Where: $\beta_m = m$ th. point at which summation of distances is a minimum.

Although this median is unique for an areal point set, it cannot be determined arithmetically. An approximation to its location must be made by an iterative search method such as that shown in King.[15]

As in the case of the mean, many areal point sets represent weighted data. Equation 10 is thus modified accordingly to calculate the general case:

$$\beta_{wm} = \sum_{i=1}^{n} [(x_m - x_i)^2 + (y_m - y_i)^2]^{1/2} \cdot w_i \quad \text{Eq. 11}$$

Where: $w_i$ = Weight or value of the $i$ th. point.

$\beta_{wm} = m$ th. point at which summation of weighted distances is minimum.

For many distributions, the mean and median are essentially at the same location. In a skewed nominal scale distribution such as that shown in Figure 6, however, the locations may vary significantly. In this particular case, the median may be the better representation of average location since it is more central to the major cluster of points. Also a greater percentage of the points is included within its ellipse.

The use of weighted data necessitates a change in form for the equations which determine the orientation and axes of the ellipse. Equations 4 and 7 therefore take the general forms shown respectively by Equations 12 and 13:

$$\sigma_{y'} = \sqrt{\frac{\sum_{i=1}^{n} (y_i \cos \Theta - x_i \sin \Theta)^2 \cdot w_i}{\sum_{i=1}^{n} w_i}} \quad \text{Eq. 12}$$

[12] Neft, *op. cit.*, p. 29 ff.

[13] Bachi, *op. cit.*, p. 92.

[14] Court, Arnold, "The Elusive Point of Minimum Travel," *Annals of the Association of American Geographers*, LIV, No. 3 (1964), p. 400.

[15] King, *op. cit.*, p. 94.

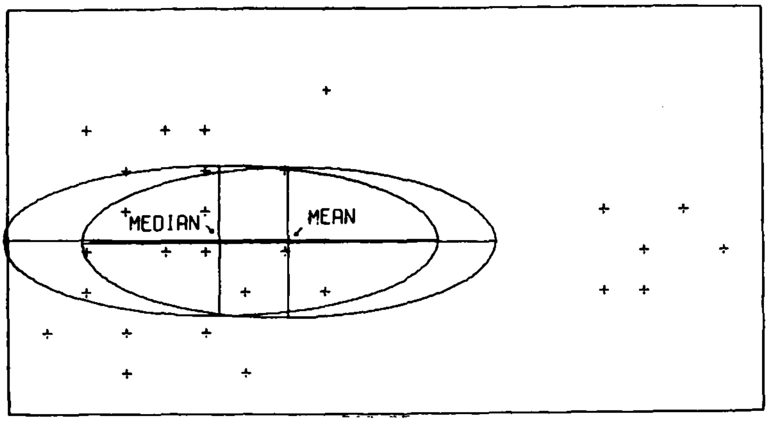


Fig. 6 Mean and Median Centered Ellipses of a Skewed Distribution

$$\tan \Theta = -\frac{\left(\sum_{i=1}^{n} x_i^2 w_i - \sum_{i=1}^{n} y_i^2 w_i\right) \pm}{2\sum_{i=1}^{n} x_i y_i w_i}$$

$$\frac{\pm\sqrt{\left(\sum_{i=1}^{n} x_i^2 w_i - \sum_{i=1}^{n} y_i^2 w_i\right)^2 + 4\left(\sum_{i=1}^{n} x_i y_i w_i\right)^2}}{2\sum_{i=1}^{n} x_i y_i w_i} \qquad \text{Eq. 13}$$

## Concentration and Dispersion of Data

The ellipse generated about a point of average location yields two measurements which may be used to examine the concentration of the data. These are the area enclosed by the curve and the number or value of the data points occurring within the ellipse. Both measurements are relative in that their interpretation comes from comparison with other data.

The area of the standard deviational ellipse gives an indication of the data concentration. When it is very small relative to the study region, the point set is concentrated or clustered. A relatively large ellipse indicates the converse, the data being more widely distributed in the region. Relative changes in a distribution over time could thus be indicated as well as differences among various point distributions.

The point count (or sum of weights) within an ellipse gives a further indication of concentration. The percent of the data within the boundary of the ellipse also gives an index of the data concentration (note Figures 5 and 6). If the area of the study region is known, the point count may be combined with the area measure to calculate densities of data inside and outside the ellipse.

The concentration measures provide a systematic method of indicating which measure of average location best describes a particular distribution. In this method, the average location whose ellipse contains the greatest proportion of the distribution (other factors being equal) would be considered the best choice. For example, in Figure 5 an ellipse centered on the weighted median might be considered the most representative since it contained 48.4 % of the data while the ellipse centered on the weighted mean (Figure 5b) contained 42.9 %.

## Orientation and Shape

The final aspects of the ellipse measure are its shape and orientation. Orientation as Bunge has stated "can be defined as the direction of the longest axis of an object."[16] For the ellipse, this is the angle of the major axis. The importance of this metric in describing a distribution varies with the eccentricity (see below); a circle ($e = 0$) has no orientation while the orientation of a line ($e = 1$) defines the locus of all points of the distribution. The orientation is a property of considerable significance to the ellipse measure for it is a primary advantage of this measure over others such as the standard circle.

16 Bunge, William. *Theoretical Geography*. Lund: Gleerup, 1962, p. 69.

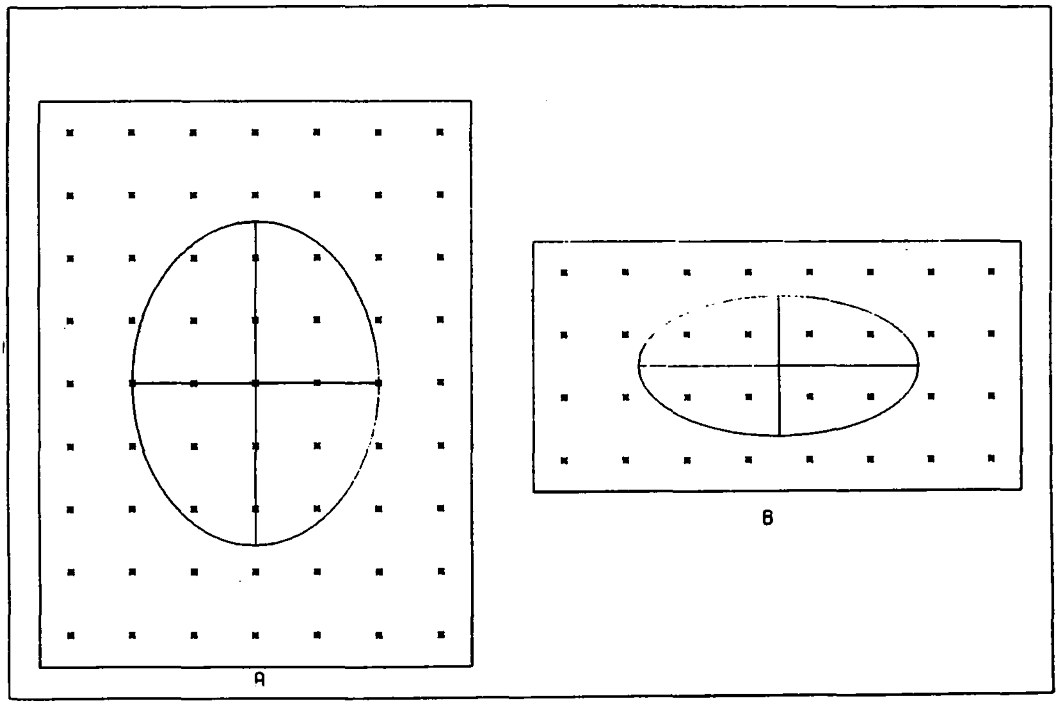


Fig. 7 Orientation Determined by Study Area Boundary

The shape or elongation of an ellipse is measured by its eccentricity:

$$e = \frac{c^*}{a} \qquad \text{Eq. 14}$$

Where: $e$ = Eccentricity.
$c$ = Length of focus.
$a$ = Length of major axis.
* after Thomas[17]

This ratio ranges from zero when the major and minor axes are equal (ellipse = circle) to one when the minor axis is zero (ellipse = line). Based upon the sample distribution here (e.g. Figures 5, 6, and 8), the eccentricity of most point sets could usually be expected to fall in the range $.5 \leq e \leq .9$.

An important factor which may affect the shape and orientation of a point distribution and its ellipse is the shape of the study region or parent distribution. If the point set being measured is distributed over only part of the base area (e.g. Figure 5), the boundary shape probably has little effect. Measures of point distributions, however, which occupy an entire study region must take the shape of that region into consideration. In fact, the eccentricity and orientation of the ellipse of a uniform distribution are determined solely by the shape of the study area boundary as Figure 7 illustrates.

A non-uniform distribution of a phenomenon covering approximately the whole of a study area poses a more serious problem. How much are the indices of the ellipse influenced by the shape of the study area or by the distribution of the parent population of which the phenomenon is a subset? One possible way to indicate the effect would be to calculate the ellipse and associated indices of the parent population. The variation of the indices (average location, orientation, eccentricity, and concentration) between the total and subset populations would show relative differences between the distributions. Considerable caution, however, would have to be used in interpreting the ellipse under these conditions.

[17] Thomas, George B. *Calulus and Analytic Geometry.* Cambridge: Addison-Wesley, 1954, p. 242.

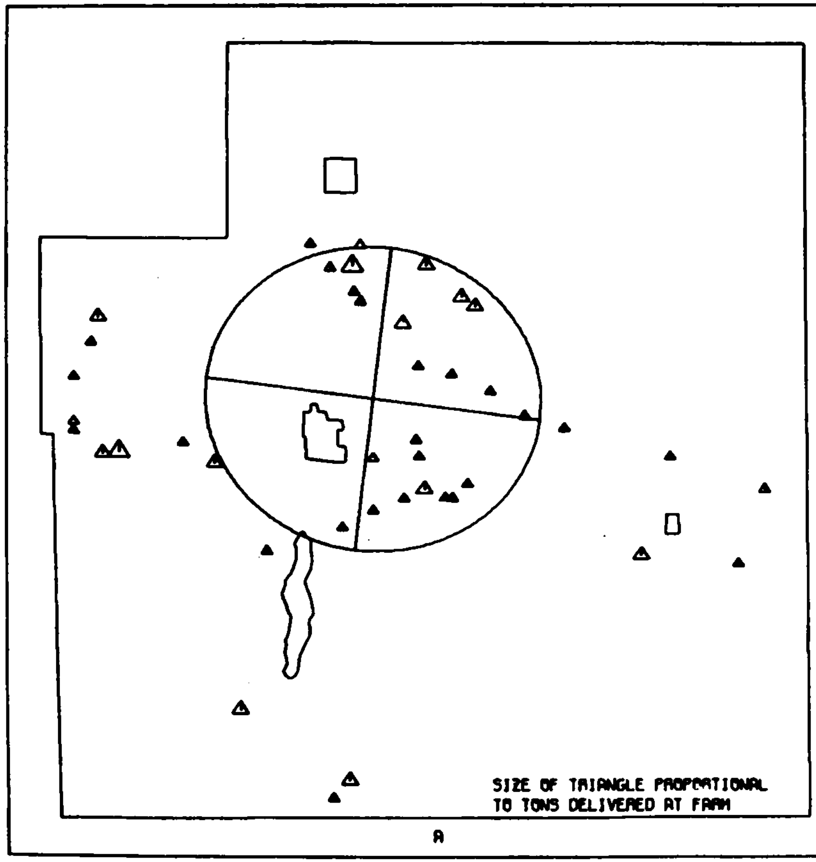


Fig. 8 A Farms Taking Limestone, Otsego County, Michigan 1950

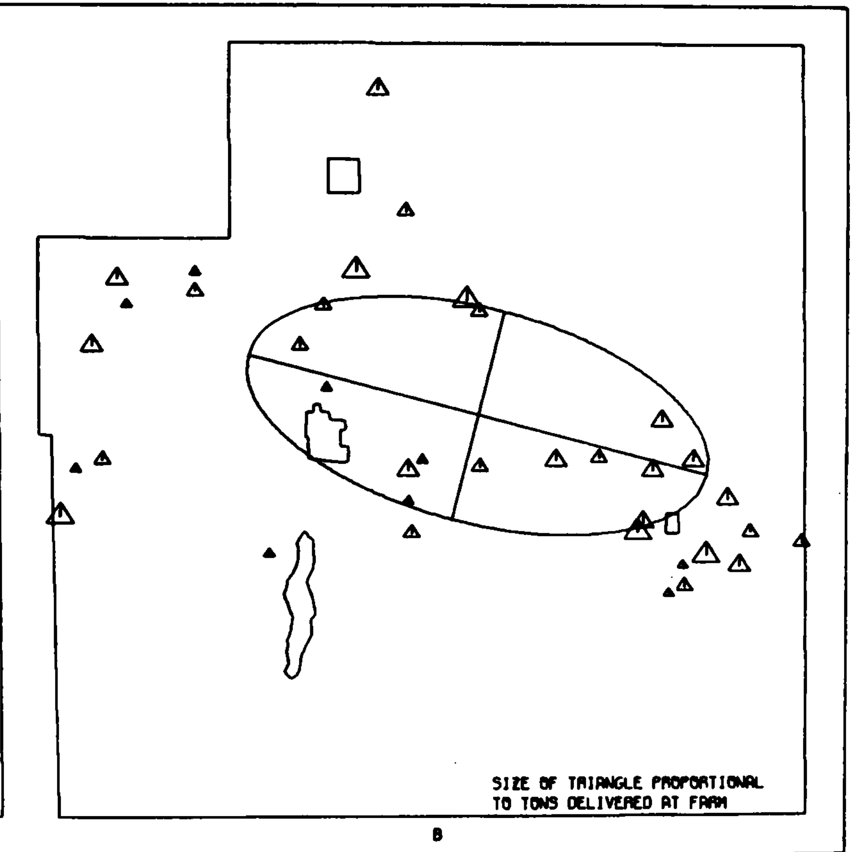


Fig. 8 B Farms Taking Limestone, Otsego County, Michigan, 1959

## Application

The general functions and effects of the standard deviational ellipse have been illustrated by generated data. The measure is now applied to several empirical points sets to indicate the variety of data to which the ellipse might be applied.

The first situation is concerned with relative change over time. It considers two point sets (Figure 8) showing the distribution of tons of limestone delivered to farms in Otsego County, Michigan in 1950 and 1959.[18] Following the procedure (above) for choice of average location, the ellipses were centered on the weighted medians. Since we are specifically concerned with relative change in the subset (farms taking limestone) and the parent distribution (all farms) remained essentially constant, the latter may be ignored here. Table I shows the indices of the ellipses for the two years.

From the area and orientation indices in Table I, it would appear that the distribution pattern did not change much from 1950 to 1959. The eccentricity and concentration indices, however, indicate otherwise. Since the data in 1959 are at the same time more linearly distributed and less concentrated at the center than in 1950, it follows that there was a relative increase in limestone delivered in portions of the periphery. This can readily be seen by comparing the distributions of Figure 8a and 8b.

A second type of data to which the ellipse

[18] Data courtesy of Mr. Stanley T. Yuill, Charlevoix Lime and Stone Company, Vanderbilt, Michigan.

Table I. Ellipse indices for figure 8.

| Index | 1950 | 1959 |
|---|---|---|
| Ellipse Center ($x$, $y$ coordinates in miles) | 9.12, 11.92 | 12.10, 11.44 |
| Ellipse Area | 65.38 mi.$^2$ | 65.08 mi.$^2$ |
| Orientation | 172.8° | 165.5° |
| Eccentricity | .4614 | .8968 |
| Concentration (Percent weights within ellipse.) | 50.7 % | 31.3 % |

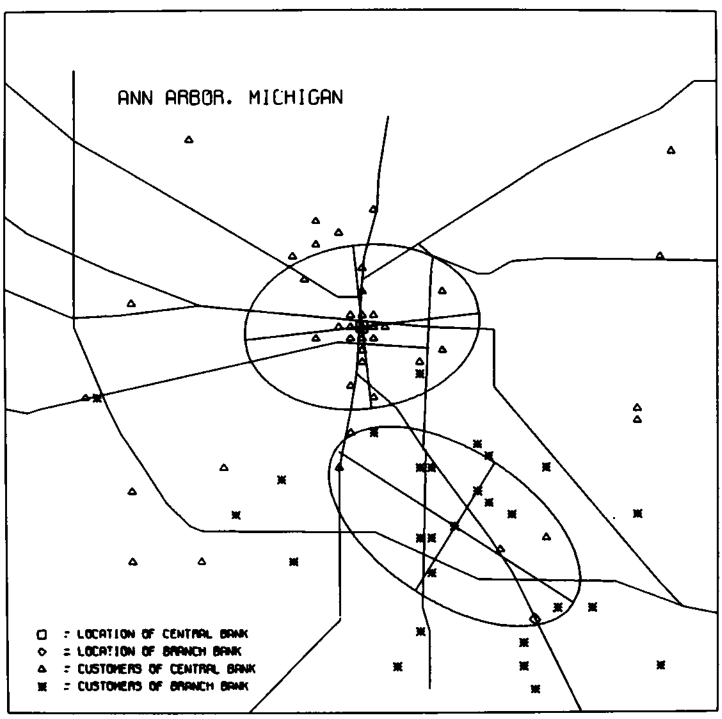


Fig. 9 Distribution of Bank Customers

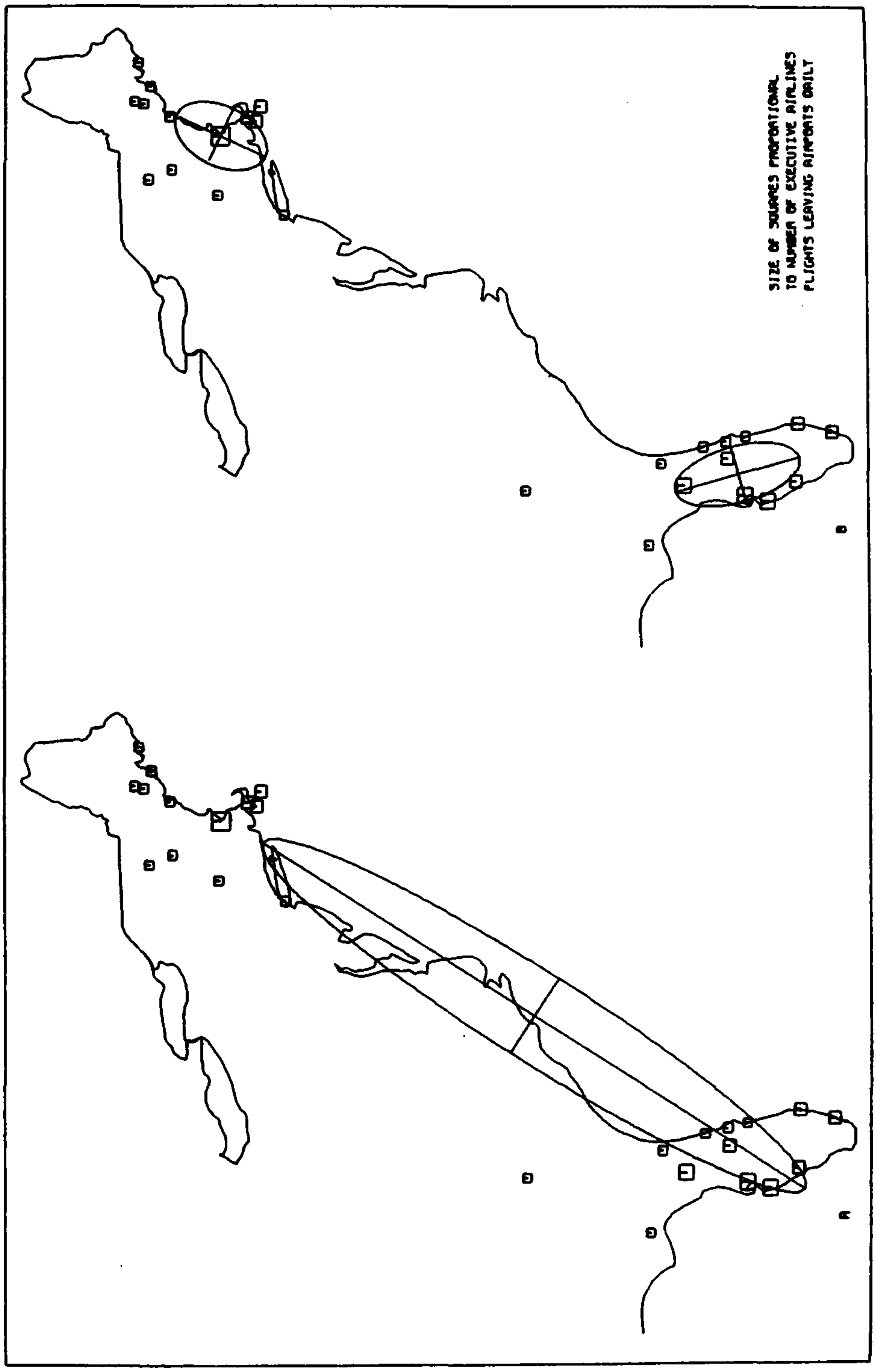


Fig. 10 A Executive airlines ellipse of total system

Fig. 10 B Executive airlines regionalized ellipses may be applied is the distribution of customers from retail or commercial establishments. Figure 9 shows the ellipses calculated for customer samples of a bank located in the CBD of Ann Arbor, Michigan and one of its branches located on a busy arterial street in the same city.[19] Table II shows the summary of the ellipse indices.

Table II. Ellipse indices for figure 9.

| Index | Main Bank | Branch Bank |
|---|---|---|
| Location | CBD | Arterial St. |
| Ellipse Center ($x$, $y$ coordinates)* | 41, 43 | 49, 26 |
| Bank coordinates | 41, 43 | 56, 18 |
| Ellipse Area | 224.3 units$^2$ | 249.3 units$^2$ |
| Orientation | 6.38° | 147.5° |
| Eccentricity | .7306 | .8510 |
| Customer Origins within Ellipse | 67.2 % | 39.3 % |

* Coordinates expressed in units of 350 feet.

As might be expected from notions of accessibility, the eccentricity index shows a more circular customer distribution at the CBD location than at the arterial location. A further indication of this is the distance between the store location and the average location of the customer distribution. For the CBD bank, this distance is negligible, but it is significantly different for the branch bank (see Figure 9). This latter difference considered in terms of the orientation eccentricity, and smaller concentration of data enclosed within the arterial location ellipse, suggests that the arterial transportation may affect the distribution of customers of the branch bank.

For data sets with great eccentricity, the ellipse measure must be applied with a certain degree of caution. A bimodal distribution in particular poses a problem for the measure, for often neither the mean nor the median measures fall within the areas of greatest data concentration. An excellent example of this is the distribution of airports (weighted by the number of flights per day) served by Executive Airlines.[20] This distribution consists of two unconnected clusters 1000 miles apart. Average location measures (Figure 10a) obviously do not accurately represent the data. The polarity is also shown by the extreme eccentricity of the ellipse ($e = .99$). The recourse to this following Bachi's example,[21] is to divide the distribution into two parts and to calculate the measure for each (Figure 10b). A further advantage of the division is to be able to spatially compare the two parts of the system.

Two important facets of the standard deviational ellipse have been demonstrated by these examples. The first is that the measure provides not just one metric to describe a distribution, but instead has a number of indices which give much more information about the point set. The second is the variety of distributions to which the ellipse may be applied. The empirical examples show the applicability to widely diverse data at greatly differing scales. The ellipse is thus a useful and versatile tool for spatial description.

[19] Data from "Ann Arbor Customer Movement Study," Ann Arbor Planning Department, 1961.

[20] Data from: *Official Airline Guide, North American Edition.* Oak Brook, Ill., Reuben H. Donnelley Corp., July 1, 1969.

[21] Bachi, *op. cit.*, p. 91.