Notes on the Analysis of Geographical Distributions

W. R. Tobler
Department of Geography
The University of Michigan
Ann Arbor

ABSTRACT

Theory implies the existence of a geographical periodicity in the arrangement of cities and other orders of service facilities, and to some extent ever predicts the frequencies of these geographical cycles. Even the casual cross-country tourist recognizes this geographical repetitiveness of events. Appropriate methods of investigation for the analysis of periodicities in geographical phenomena can be organized about the concept of a spatial series.

Paper delivered at the Institute on Emerging Concepts and Methods in Urban and Regional Analysis, East Lansing, 16 July 1965.

Notes on the Analysis of Geographical Distributions

Increasing attention is being devoted to the solution of regional and metropolitan problems. A common feature, recognized by the individuals responsible for such studies, is that the processes involved take place in a geographical context. A result is that vast amounts of geographical information are being accumulated. The convenient storage and retrieval of this information is being made possible by the introduction of large computing centers. A great deal of effort is being expended in this retooling for an increased informational capacity. At the same time effort is being devoted to consideration of methods of analysis appropriate to the expanding storage, retrieval, and processing capabilities. Some progress in this direction has been made in the field of geographical (spatial, regional, or areal) information processing. One of the encouraging developments has been the recent rise in the amount of economic, social, and cultural information identified in a locatively useful manner, as for example, the recent provision on the part of the U.S. Bureau of the Census of latitude and longitude coordinates for population enumerations. This allows direct computer analysis in greater volume and at less cost than previously employed methods and eliminates the need to rely heavily on nomographic approximations obtained from maps, the geographical version of the graph. At the same time the identification of data by geographical coordinates allows the automatic preparation of maps for those types of visual pattern recognition and speculative hypothesis-generation operations which the human can perform far better than current computers. A merging of the advantages of both formulations, the analytical and graphical, is represented by "sketch-pad" computing facilities.

Given the complexity of many geographical problems, and the fact that we would hope to be able to solve at least some of these in automatic real-time environments in the future, it is clear that this convenient locative identification is a step in the direction of the geographical information processing capacity required to cope with the magnitudes of data becoming available from "information centers" or from such recently developed geographical information collection systems as earth satellites. The megalopolis region may well find it to be economically beneficial to invest $4,000,000 for an orbiting (that is, stationary) sensor as a part of a larger information processing system.

There are many modes of organization for the analysis of geographical distributions. One which is common in the physical sciences but which is rarely applied in the social sciences considers space to be continuous (not compartmentalized into "regions"). By analogy with time series analysis, one can define a subject called geographical (or spatial) series analysis. Almost every textbook on economic statistics has a section on time series analysis; none has a section on spatial series analysis. There are many texts and monographs on time series. Again, there are none on geographical series. For each of the methods of analyzing time series data it seems possible to consider the equivalent for a geographical series. A first problem is that the mathematical extension often is difficult. Secondly, it is at least as critical that a geographical interpretation be supplied for the suggested manipulations. Generally this is simpler than would be imagined. Central place theory, for example, clearly implies that there exists a geographical periodicity in the arrangement of cities and other orders of service facilities, and to some extent even predicts the frequency of these cycles. It is apparent that methods of spectral (or harmonic) analysis would be of great assistance in the empirical verification of this important theory.

Table I is a presentation of some temporal data, arranged in alphabetical order. The first step in the analysis of this information might be to rearrange the table into its correct temporal sequence, and this is

TABLE II **

Anhwei	30	54
Chekiang	23	39
Chinghai	2	278
Fukien	13	48
Heilung-Kiang	12	179
Honan	44	65
Hopei	36	82
Hunan	33	81
Hupei	28	72
Kansu	13	167
Kiangsi	17	64
Kinagsu	41	41
Kirin	11	72
Kwangsi	20	85
Kwantunk	35	89
Kweichow	15	67
Lisoning	19	58
Shansi	14	61
Shantung	49	59
Shensi	16	76
Sinkiang	5	635
Szechwan	63	219
Yunnan	18	168

** China, population in millions and area in thousands of square miles.

TABLE I *

April	20
August	115
December	2
February	11
January	4
June	112
July	121
March	14
May	50
November	5
October	43
September	90

* Calcutta, average rainfall in tenths of inches.

rather easily done. Table II is a presentation of some geographical data, and the reader is asked to perform the comparable operation; arrange the data in its correct spatial sequence. The geographical ordering is the more difficult. Virtually everybody is familiar with the temporal sequence of the partitioning of the year into months, but few people would know the correct spatial sequence of the political partitioning of even as important a country as China. Another difficulty arises from the fact that the geographical sequence is quite irregular. Possibly one reason that geographical information is so frequently given in alphabetical order is that the printer would find a "geographical table" (with every entry in its correct spatial position) too irregular and with too many blank spaces for his symbol positioning techniques.

Related to the foregoing considerations is the disparity in size of intervals. For temporal data one would hardly collect information in a manner such that the first interval spans 30 days, the next 60 days, the next 25 seconds, the next 2 days, then 15 minutes, and so on. Instead one uses intervals of equal, or nearly equal, size. A similar strategy has been proposed (Hägerstrand, 1955) for geographical data collection but is rarely found.

A next step (recommended by most textbooks) in the analysis of temporal data is the preparation of a graph or histogram. Typically the data are assigned to the midpoint of the interval and represented with an ordinate proportional to the observations. A smooth curve is then drawn through the resulting points. Comparable procedures are employed for geographical distributions, with the results shown as a perspective diagram or in a plan view. The contour map of smooth level curves is of course the two-dimensional equivalent of the smooth curve on the ordinary graph. A difficulty arises here in choosing the center of the spatial interval, which has a shape as well as a size, and is therefore more complicated than in the case of temporal intervals. The two-dimensional interpolation required to obtain smooth level observations is also more complicated than the one-dimensional interpolation necessary for the temporal graphs.

From this simple introduction we can continue the analogy with time series in a straight forward manner. A continuing complication is the spacing of the observations, however. Several situations may arise: observations are obtained at irregular intervals (either by the sampling technique employed, or by the data collection agency, or because of the inherent nature of the events), or observations are obtained or assigned (by some interpolation technique) to regular grid points. These grid points may be either a square or a trigonal lattice on a plane, or polar coordinates, or an adaptation of latitude and longitude on a sphere or spheroid. It suffices here to consider only plane phenomena, and there is great temptation to consider only square grids since the mathematics are simplified. The greatest differences are between the square grid and the irregular scatter of observations. It is generally a simple matter to adapt the techniques to a trigonal (or hexagonal) grid, even on a computer, once they have been developed for a square grid. Data obtained by reference to polar coordinates can be converted to rectangular coordinates and then analyzed as are the irregularly scattered observations, though this is a rather unnatural procedure.

A simple smoothing technique often applied to time series data is the moving average. Here each observation is averaged with its neighbors, which may or may not be weighted. In the geographical case, and for a square grid, this may involve 5,9,13,17,21,25...points, or more typically, the k^{th} weighted average involves $(2k + 1)^2$ points. This can be written as

$$\overline{Z}_{ij,k} = \frac{\sum_{m=-k}^{m=k} \sum_{n=-k}^{n=k} W_{m,n} Z_{i+m,j+n}}{\sum_{m=-k}^{m=k} \sum_{n=-k}^{n=k} W_{m,n}}$$

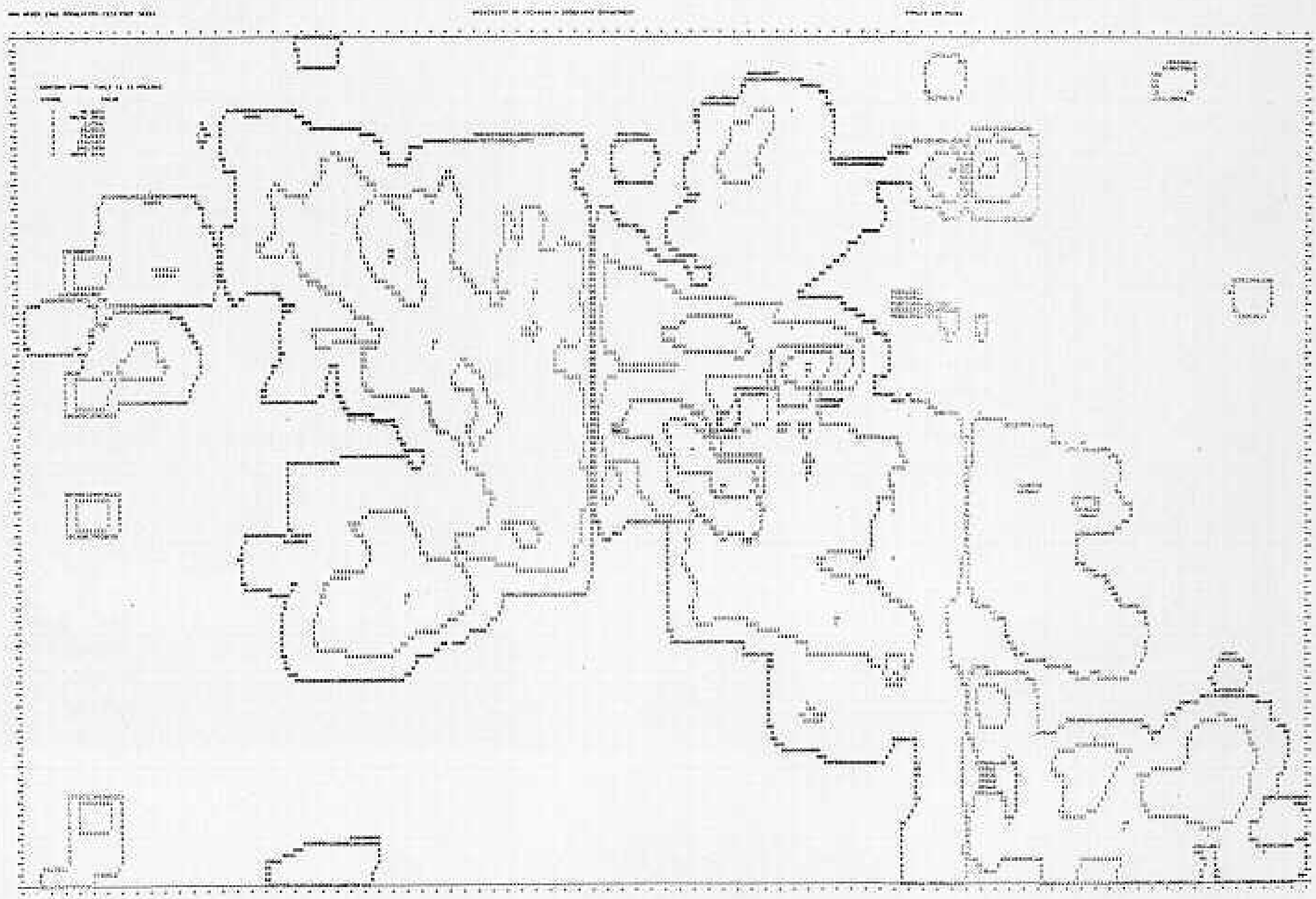
where W is an appropriate weighting factor. This operation is repeated for all grid points, except boundary points. For irregularly scattered observations a moving average might be established by averaging nearest (weighted or unweighted) neighbors, which is again rather simple.

As an example of the geographical use of a moving average, Mr. Yuill, a graduate student in the Department of Geography, recently compiled population figures for the city of Ann Arbor on a 300 foot grid. The population of each city block (or fraction thereof) was assigned to the grid point(s) which fell on that block. Grid points which landed in the street were assigned a value of zero, which caused extreme fluctuations in the data. A moving average was employed to smooth these observations enabling the computer to produce a population density contour map. (Another strategy of course would have been to use a floating template of larger than grid size in the initial determination of the population densities).

The moving average given above is a local or neighborhood operator (in this case a smoothing) in which k specifies the region considered to be local. As this neighborhood is enlarged the operator tends to become more and more global (the mathematical concepts of local and global are vaguely akin to the geographical concepts of site and situation). Other local operators may be of interest. For example,

$$\sigma^2_{ij,k} = \frac{\sum_{m=-k}^{m=k} \sum_{n=-k}^{n=k} (\overline{Z}_{ij,k} - Z_{i+m,j+n})^2}{(2k+1)^2}$$

will define a local variance, from which a contour map might be produced. This is of some theoretical interest since the variance may be regarded as the integral of the power spectrum over all frequencies. The local variance may also be regarded as a measure of texture (on aerial photographs) or a measure of roughness (in topographical situations). The local range, known from geographical studies of terrain as the "relative relief", can be defined



in a similar manner. Numerous additional local operators of interest can be defined.

Two asides may be appropriate here. One concerns the development and construction of a "parallel" computer at the University of Illinois. In contrast to most serial computers, which would have to apply the foregoing operations to all points in sequence, the parallel computer can apply the operations to all points simultaneously, with a resultant reduction in cost. The second point is of more theoretical interest and concerns the boundaries. It is clear that the smoothing operation cannot work in the vicinity of the edges of the region of observations. This is true of several of the operators discussed. The implication is that the region of observations should exceed the region of interest. For example, an agency concerned with the state of Iowa should also have observations from a strip completely surrounding Iowa.

In the analysis of temporal phenomena one is often concerned with the rates of change. A similar technique is available for geographical phenomena. This might, for example, be the rate of change of land values per mile. Clearly this depends on the direction it is natural to consider the maximum rate of change per unit distance (the gradient, orthogonal to the contours). For gridded data the gradient at the ij^{th} point can be calculated from

$$|\Delta Z| = \left(\left(\frac{\partial z}{\partial x} \right)^2 + \left(\frac{\partial z}{\partial y} \right)^2 \right)^{1/2}$$

where

$$\frac{\partial z}{\partial x} = \frac{Z_{i,j+1} - Z_{i,j-1}}{2\Delta x}$$

$$\frac{\partial z}{\partial y} = \frac{Z_{i+1,j} - Z_{i-1,j}}{2\Delta y}$$

This can again be contoured or can be used to compute the rate of change of the rate of change. This second spatial derivative, along with discriminatory analysis, information theory, and several other techniques, is useful in boundary delineation. In addition to the gradient, it may be of interest to calculate the directional derivative, as for example, the rate of change per mile of the population density from the central business district.

A fundamental question in the analysis of temporal data is whether or not there is a trend. This generally involves fitting a straight line to the graph of the observations, though a curvilinear trend is occasionally employed. Similar techniques can be employed for geographical phenomena. For example, the trend of population density (persons per square mile) in Michigan is

$$P = 6314 - 68.7*\text{latitude N} - 36.9*\text{longitude W},$$

or, in words, the population declines towards the northwest, which is not a particularly astounding result, though useful. For example, knowledge concerning the trend can be employed to adjust for population density in a rather simple manner, and the departures from the trend (residuals or anomolies) can be examined in greater detail. The technique of trend analysis, as employed in practice, is basically a linear or curvilinear least-squares curve-fitting procedure (multiple regression with geographical coordinates as the independant variables).

As another example of the utility of this technique consider the Von Thünen model of agricultural land use. The statement of the model is:

a) Commodities are chosen to maximize income.

b) $I = M-P-T$,

where

I = income

M_i = market price of the i^{th} commodity

P_i = production costs

T_i = transportation costs from farm to market

This model has been programmed for computer solution by Mr. A. Schwarz, a sophomore student at the University of Michigan. As formulated, the computer program allows many market locations and commodities. Transportation costs are linear functions of the distance from the market, and production costs depend only on the commodity. An obvious improvement in the program would be to allow production costs to be a function of the productiveness (fertility, etc.) of each location. A simple procedure is to enter the production costs as geographical trend equations. A climatic gradient could easily be represented as an equation similar (perhaps non-linear) to that given for population density in Michigan. Simple sinusoids could provide a first approximation to rolling topography, and so on. The final restrictions are basically limitations on data availability. The curve fitting technique can be extended to more and more complicated equations so that one may arrive at a "complete" mathematical equation which represents the observations in question to a high degree. Such "numerical maps" are employed in several fields. An alternate, but related direction is the fitting of theoretically derived curves, such as the S-shaped curve in demographic studies or the exponential decay function for the distribution of population within cities.

A practical application of these techniques might be the greater acceptance (and availability) of regional and metropolitan data which have been "adjusted" for regional trends. The net effect of this would be a reduction of regional complexity toward the "homogeneous plain" concept now employed in theoretical work of some elegance (this is perhaps somewhat comparable to the meteorologists "reduction to sea level", an effort to have comparable data despite the vagaries of topography). Presumably such adjustments for geographical trends would simplify some aspects of the current complexity of the environment which in turn might aid in the solution of numerous important practical problems of development and planning.

The economist often considers that events at time t may be related to events at some previous time t-k. It is not unreasonable to expect that events at one location (ij) may be related to events at some other geographical location (i+m, j+n). We are thus lead to a consideration of serial (or auto-)

correlation of each observation with, for example, its neighbor immediately to the east, i.e.

$$r\ (Z_{ij},\ Z_{i+1,j})$$

or with the neighbor immediately north

$$r\ (Z_{ij},\ Z_{i,j+1})$$

or with the neighbor two removed to the east

$$r\ (Z_{ij},\ Z_{i+2,j})$$

and so on. Consider all possible such lagged correlations. Now plot a contour map of the correlation coefficients as a function of the spatial lags in the two directions (i.e., $r = r(m,n)$, where $m = \pm 1, \ldots, n = \pm 1, \ldots$). "Peaks" on this correlogram indicate similarities of events, and the symmetry (or lack of symmetry) about the origin of the contoured correlation coefficients relates to the isotropicity of the space. This is somewhat difficult to interpret, but an appropriate question might be: Does the (economic, cultural, social) landscape repeat itself, and if so, how often and in which direction? As another aside, attention is drawn to the recent development of optical computers which perform this correlation operation (and several of the operations to be discussed shortly) extremely rapidly, and which involve no complicated computer program in the conventional sense.

For scattered and irregular observations the serial correlation procedure might be attempted in several ways. The most obvious is to correlate each observation with its nearest neighbor, perhaps adjusting for their distance apart. Obviously the process is symmetrical, which reduces the number of samples available. A second approach might be to correlate only with nearest neighbors within some sector. This should bring out directional tendancies more clearly than the previous method. Finally, correlation of all neighbors within some distance increment could be attempted. This involves a variable number of observations at each stage of the process, but seems appropriate for geographical situations in which some phenomenon is spreading from central places (city, or university, etc.).

It is often of interest to compare two sets of time series data with each other (e.g., stock market prices and consumer price index; or tree rings with weather records, etc.). This is known as cross-correlation. The geographical equivalent is the comparison of two maps, of either the same region or of different regions. A caveat should be entered here. Geographical patterns may be considered to consist of points, lines, areas, intensities, or flows. This analysis, if correct, would lead to 25 distinct types of intercomparisons. Among these might be, for example, comparison of the road pattern of the U.S. (lines) with the pattern of cities (points), or with the pattern of railroads (lines), or with the pattern of "depressed" areas (areas), or with the pattern of population densities (intensities), and so on. Most of the statistical techniques now available allow only comparison of like entities. Comparisons of pairs of point patterns, for example, has recently been shown to be possible via ordinary correlation and regression techniques, but extended to the domain of complex numbers. Similarly a two-dimensional version of rank correlation has been devised for the comparison of areal patterns, and a tensor correlation has been suggested for comparisons of intensity patterns. The cross-correlation technique, as employed in time-series analysis, really applies only to sets of observations which are considered (sampled) continuous functions. The geographical equivalent might be two maps showing, say, rainfall intensities and per acre value of agricultural land. The most common technique is to superimpose the two maps and calculate the immediate correspondences, that is, with both sets of data (Z and W) on a grid, calculate $r = r\ (Z_{ij}, W_{ij})$ for all ij.

The two dimensional cross-correlation goes one step further. Like the autocorrelation technique it shifts one of the maps around and produces a correlogram giving the correlation coefficients as a function of the spatial lags in two directions, viz

$$r = r(m,n) = r\ (Z_{ij}, W_{i+m,j+n}).$$

As a geographical example, the highest incidence of irritation from air polution might be downwind from the highest intensity of automobile traffic. Another application might be to attempt to find the "best" positioning of two maps relative to each other in order to "compare" two different cities. Comparisons of topographical map sheets with a file of "standard" physiographic types could also be performed via the cross-correlation technique.

The previous comments regarding autocorrelation for irregularly spaced observations also appear to apply to cross-correlation functions. The usual cross-correlation technique for gridded data involves only translations (no reflections or rotations) and therefore has several limitations, as workers in the field of pattern recognition have become aware: Letters which are upside down cannot be recognized, for example, and there may be geographical situations of interest which would not be detected because of this limitation. The use of two-dimensional convolutions may be required, and the statistical validity of cross-correlations must be evaluated by use of the relatively complicated coherence measures.

Seasonal trends, or business cycles, are often of interest in the analysis of economic time series. One approach to these phenomena is to adjust the data by taking out these fluctuations. For cyclical events the use, and empirical fitting, of Fourier series is common. Computer programs are now available for the fitting, to geographical distributions, of so-called double Fourier series. These will yield descriptive equations involving trigonometric functions. The next step requires some rather delicate analyses to decide whether periodicities actually exist or whether the data simply contain quasi-periodicities, or only apparent periodicities, or none of these. The mathematical details become rather complicated here, but the autocorrelation function can be shown to be related, via the Fourier transform, to the power spectra of the "waves". The interpretation, of course, is that central place theory does suggest a geographical periodicity in the arrangement of cities, stores, hospitals, and so on. This can be seen if one imagines and considers the highly regular and repeating pattern of the population density contours implied by Christaller's original statement of central place theory. In consequence, the search for geographical periodicities is not simply a vacuous empirical or mathematical exercise. A complicating factor, of course, is the geometry of geography, which is warped by transportation facilities and other real conditions. As a remark, the analysis of data for periodicities seems to be practical only when the observations are available (so collected or assigned) at uniform intervals. Choice of the appropriate grid interval appears to be related to the frequency expected, and the effects of aliasing must be considered.

Assume that central place periodicities could be estimated (at some level of statistical confidence) from empirical observations of, say, the population distribution in the United States. Aside from the obvious expectations regarding wave lengths and phases implied by the theory, additional validity tests are available. One can consider the population to consist of two sectors, urban and rural. These then should correspond rather well to a distinction between central place and non-central place populations. Consequently one would expect that subtraction from the total population of that portion of the population which could be attributed to the geographical cycles should leave as a remainder the "rural" population. Since the U.S. Bureau of the Census distinguishes between urban and rural population the empirical information is available for the test.

In summary, it appears that valid interpretations can be attached to geographical extensions of the methods of analysis developed for the study of time series.

Select References

J. L. Alward, "Notes on Spatial Discrimination", Engineering Summer Conference on Advanced Infrared Technology, lecture notes, Ann Arbor, June 1965.

R. Bachi, "Standard Distance Measures and Related Methods for Spatial Analysis", Papers, Regional Science Assn., X(1962), pp. 83-132.

M. S. Bartlett, "The Spectral Analysis of Two-Dimensional Point Processes", Biometrika, Vol. 51, Parts 3 and 4, Dec. 1964, pp. 299-311.

B. J. L. Berry, and A. Pred, Central Place Studies, Philadelphia, Regional Science Research Institute, 1961.

R. A. Bryson, and J. A. Dutton, "The Variance Spectra of Certain Natural Series", Quantitative Geography, W. Garrison (ed.), forthcoming.

Jon Cunnyngham, "The Spectral Analysis of Economic Time Series", Washington D.C., U.S. Bureau of the Census Working Paper #14, 1963.

L. J. Cutrona, E. N. Leith, C. J. Palermo, "Optical Data Processing and Filtering Systems," IRE Trans. Inform. Theory, Vol. IT-6, No. 3, June 1960, p. 391 et seq.

O. D. Duncan, R. P. Cuzzort, and B. Duncan, Statistical Geography: Problems in Analyzing Areal Data, The Free Press of Glencoe, Illinois, 1961.

E. S. Epstein, and J. A. Leese, "Application of Two-Dimensional Spectral Analysis to the Quantification of Satellite Cloud Photographs", Ann Arbor, University of Michigan (UM 05068-1-F), 1963.

T. Hägerstrand, "Statistika Primaruppgifter, Flygkartering Och 'Data Processing' Maskiner: Ett Kombinerings projekt", Meddelanden Fran Lunds Geografiska Institution, Nr 344, Lund, University of Lund, 1955, pp. 233-255.

J. W. Harbaugh, and F. W. Preston, "Fourier Series Analysis in Geology", Computers and Computer Applications in Mining and Exploration, J. Dotson and W. Peters (eds.), Tucson, University of Arizona, 1965.

J. L. Holloway, "Smoothing and Filtering of Time Series and Space Fields," Advances in Geophysics, Vol. 4, New York, Academic Press, 1958.

IBM Corporation, "Numerical Surface Techniques and Contour Map Plotting", E. 20-0117-0, undated (1964?).

W. C. Krumbein, "Trend Surface Analysis of Contour-Type Maps with Irregualr Control-Point Spacing", Journal of Geophysical Research, 67, 7(1959), pp. 823-834.

____________, and F. A. Graybill, An Introduction to Statistical Models in Geology, New York, McGraw-Hill, 1965.

Y. W. Lee, Statistical Theory of Communication, New York, J. Wiley, 1960, 501 pp.

M. Masuamya, "Correlation between Tensor Quantities", Proceedings, Physico-Math. Soc. Japan, 3rd Series, 21 (1939), pp. 638-647.

H. H. McCarty, and N. E. Salisbury, "Visual Comparison of Isopleth Maps as a Means of Determining Correlations Between Spatially Distributed Phenomena", Iowa City, Department of Geography, State University of Iowa, 1961.

R. L. Miller, and J. S. Kahn, Statistical Analysis in the Geological Sciences, New York, J. Wiley, 1962.

R. F. Minnick, "A Method for the Measurement of Areal Correspondence", Papers, Michigan Academy of Science, Arts, and Letters, XLIX (1964), pp. 333-344.

W. D. Montgomery, and P. W. Brome, "Spatial Filtering", Journal of the Optical Society of America, 52, 11 (1962), pp. 1259-1276.

H. A. Panofsky, and G. W. Brier, Some Applications of Statistics to Meteorology, University Park, Pennsylvania State University, 1958.

W. J. Pierson, ed., "The Directional Spectrum of a Wind Generated Sea as Determined from Data Obtained by the Stereo Wave Observation Project", Meteorological Papers, New York University, Vol. 2, No. 6.

A. H. Robinson, Mapping the Correspondence of Isarithmic Maps", Annals, Assn. Am. Geographers, 47 (1957), pp. 379-391.

A. Rosenfeld, and J. Pfaltz, Sequential Operations in Digital Picture Processing, TR-65-18, Computer Science Center, University of Maryland, 1965.

S. M. Simpson, Jr., "Least Squares Polynomial Fitting to Gravitation Data and Density Plotting by Digital Computers", Geophysics, XIX (1954), pp. 250-257.

P. Switzer, C. M. Mohr, and R. E. Heitman, "Statistical Analyses of Ocean Terrain and Contour Plotting Procedures," Cambridge, A. D. Little (AD 601538), 1964.

E. N. Thomas, "The Stability of Distance - Population - Size Relationships for Iowa Towns from 1900-1950," IGU Symposium in Urban Geography, K. Norborg (ed.), Lund, Gleerup, 1962.

E. N. Thomas, "Maps of Residuals from Regression: Their Characteristics and Uses in Geographic Research", Iowa City, Department of Geography, State University of Iowa, 1960.

W. R. Tobler, "A Polynomial Representation of Michigan Population", Papers, Michigan Academy of Science, Arts, and Letters, XLIX, 1964, pp. 445-452.

____________, "Computation of the Correspondence of Geographical Patterns", Ann Arbor, 1964 Regional Science Association Meeting.

J. W. Tukey, and R. B. Blackman, The Measurement of Power Spectra, New York, Dover, 1959.

N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, Cambridge, MIT Press, 1949.

A. Winder, and C. Loda, Introduction to Acoustical Space-Time Information Processing, ONR report ACR-63, Washington, GPO, 1962.