Title
Unit 41 - Spatial Interpolation II

Authors
Unit 41, CC in GIS
Waters, Nigel M.

Publication Date
1990

Peer reviewed

UNIT 41 - SPATIAL INTERPOLATION II

UNIT 41 - SPATIAL INTERPOLATION II

Compiled with assistance from Nigel M. Waters, University of Calgary

UNIT 41 - SPATIAL INTERPOLATION II

Compiled with assistance from Nigel M. Waters, University of Calgary

A. INTRODUCTION

- this unit continues the examination of spatial interpolation by looking at areal interpolation techniques and some applications
- areal interpolation is the problem of transferring data from one set of areas (source reporting zones) to another (target reporting zones)
  - this is easy if the target set is an aggregation of the source set, but more difficult if the boundaries of the target set are independent of the source set

- later we look at applications that do not fall easily into either point or areal interpolation categories

B. AREAL INTERPOLATION - NON-VOLUME PRESERVING

- e.g. interpolating population counts from census tracts to school districts

Procedure

overhead - Non-volume preserving areal interpolation

- calculate the population density for each source census tract by dividing population by area
- identify a centroid for each region
  - assign to the point located at each centroid, the population density value determined for its enclosing area
- using this set of points, interpolate a gridded population density surface using any of the methods described previously
- convert each grid cell's value to a population by multiplying the estimated density by the cell's area
- overlay the interpolated grid on the target map and assign each grid value to each its target region (school district)
- calculate the total population in each target region
- this method is criticized because:
  - choosing the center point is ill-defined
  - inadequacy of point based interpolation methods
  - most importantly, the total value of each zone is not conserved
    - e.g. if a source zone is divided into two target zones, the total population of the target zones after interpolation need not equal the population of the source zone

C. AREAL INTERPOLATION - VOLUME-PRESERVING

1. Overlay

- discussed by MacDougall (1976) and Goodchild and Lam (1980)
- procedure involves:
  - overlay of target and source zones
  - determining the proportion of each source zone that is assigned to each target zone
  - apportioning the total value of the attribute for each source zone to target zones

according to the areal proportions

- assumes uniform density of the attribute within each zone
    - e.g. uniform population density if the attribute is total zone population

2. Pycnophylactic

- see Tobler (1979) for the original algorithm

- the technique has two objectives: 1. create a smooth surface, no steps
        - attribute values should not change suddenly at zone boundaries 2. the total value of the attribute within each zone must be correct

- procedure: 1. overlay a dense raster on a choropleth map 2. divide each zone's total value equally among the raster cells that overlap the zone

  3. smooth the values by replacing each cell's value with the average of its neighbors 4. sum the values of the cells in each zone 5. adjust the values of all cells within each zone proportionally so that the zone's total is the same as the original total

        - e.g., if the total is 10% low, increase the value of each cell by 10% 6. repeat steps 3, 4 and 5 until no more changes occur

- does not require an assumption of homogeneity within zones but rapid variation within zones may affect the quality of interpolation

- output is a contour or continuously shaded map

Boundary conditions

- at the boundary of the reporting zones, pixels will have neighbors outside the study area and therefore without values
    - some decision must be made about the behavior of the surface outside the study area
        - e.g. population density equals zero (a lake or rural area)
        - e.g. population density unknown, assumed equal to the values of the outermost pixels of the study area

D. SPECIAL CASES OF SPATIAL INTERPOLATION

1. Mapping populated areas

- objective is to create a map showing "populated areas", given point population values for a number of cities and towns

- this problem arises frequently when populated areas are represented as points

- it arises for small reporting zones when boundary files are unavailable, but data includes centroid locations e.g. US or UK census data

- are several methods that could be used

- a simple approach would be to estimate the populated area using an empirical relationship like:

  A is proportional to p0.84

  - and draw a circle around the point, of radius:

    / (A / p)

- Bracken and Martin (1989) have developed methods for replacing ED centroids by disks, the radius of each disk being estimated from the distances to neighboring centroids
  - the method works very well with UK ED data

- an alternative approach might proceed as follows:
  - establish a critical population density for defining an urban area
  - spread the population over each urban area so that population density is highest in the center and decreases gradually outwards
    - e.g. use a normal distribution function

  - interpolate densities to a raster, accumulating values where the population spread from two urban areas overlap

  - draw contours at the critical value to define the boundaries of the populated areas

- both of these methods fall within the general heading of density estimation
  - a density is being estimated from a collection of points
  - see Silverman

2. Estimating trade areas

- in marketing, it is often desirable to plot the boundary of a trade area for e.g. a store, given information on the home locations of customers

- simplest case is when the location of all customers and non-customers is known
  - simply draw a boundary contour between them

- if the location of non-customers is not known: 1. calculate the average distance to all customers and draw a circle or 2. divide the area into sectors, average the distance to customers within the sectors and draw a distance arc for each sector (see Huff and Batsell, 1977)

    - these techniques do not pick up islands or holes in the trade area

    or 3. give each customer a small probability surface

    - accumulate values as in the populated areas example set critical value for delimiting trade area

E. A GIS PERSPECTIVE ON INTERPOLATION

- both point and areal interpolation try to estimate a continuous surface
  - in the point case, the surface has been measured at sample points
  - in the areal case, the surface of population density is estimated from total population counts in each reporting zone

- in other cases it is impossible to conceive of a continuous surface
  - e.g. if each point is a city and the attribute is city population
    - if city A has population 1 million and city B 100 km away has population 2 million, there is no reason to believe in the existence of a city half way between A and B with population 1.5 million
  - in this case, the variable population exists only at the points, not as a continuous surface
  - in other cases the variable might exist only along lines e.g. traffic density on a street network

- we must distinguish here between layer and object views of the world
  - a continuous surface of elevations is a layer view of the world - there is one value of elevation at an infinite number of possible places in the space
  - the point map of cities is an object view of the world - the space in between points is empty, and has no value of the population variable
  - the street map is an object view of the world - the world is empty except where there are streets - only along streets is traffic density defined

- spatial interpolation implies a layer view of the world, and it requires special techniques (e.g. density estimation) to apply it to objects such as store customers

Expert systems for spatial interpolation algorithms

- a good GIS should include a range of spatial interpolation routines so that the user can choose the most appropriate method for the data and the task

- ideally, these routines should provide a natural language interface which would lead the user through an appropriate series of questions about the intentions, goals and aims of the user and about the nature of the data

- a number of prototype expert systems for guiding the choice of a spatial interpolation algorithm have been developed

- these may be written in the form of:
  - an expert system shell (Waters, 1988)
  - in one of the artificial intelligence languages such as Prolog or LISP (see Dutton-Marion, 1988)
  - or in a high level language such as Pascal (Maslyn, 1987)

Conclusion

- if computer contouring and surface generation techniques are to be incorporated successfully into GIS, they must be easy to use and effective
    - "easy to use" implies that those without a detailed knowledge of the mathematical and statistical characteristics of the procedure should be able to choose the correct technique for displaying a particular data set for a particular purpose
        - note: statisticians argue that this is not an ideal goal as people may use techniques without a proper understanding of the underlying assumptions
    - "effective" means that these techniques should be informative, highlighting the essential nature of the data and/or surface and serving the purpose of the researcher/analyst
        - the researcher's measure of success will be largely subjective and visual - does the result look right?
- this purpose may vary from an attempt to model all the "real" intricacies of the surface to simply trying to highlight the general, spatial trend of the data in order to aid in the decision-making process

REFERENCES

Bracken, I. and D. Martin, 1989. "The generation of spatial population distributions from census centroid data," Environment and Planning A 21:537-44.

Dutton-Marion, K.E., 1988. Principles of Interpolation Procedures in the Display and Analysis of Spatial Data: A Comparative Analysis of Conceptual and Computer Contouring, unpublished Ph.D. Thesis, Department of Geography, University of Calgary, Calgary, Alberta.

Goodchild, M.F., and Lam, N., 1980. "Areal Interpolation: A Variant of the Traditional Spatial Problem," Geo- Processing 1: 297-312.

Huff, D.L. and R.R. Batsell, 1977. "Delimiting the areal extent of a market area," Journal of Marketing Research 14:581-5.

MacDougall, E.B., 1976. Computer Programming for Spatial Problems, Arnold, London.

Maslyn, R.M., 1987. "Gridding Advisor: An Expert System for Selecting Gridding Algorithms," Geobyte 2(4):42-43.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.

Tobler, W.R., 1979. "Smooth pycnophylactic interpolation for geographical regions," Journal of the American Statistical Association 74:519-30.

Waters, N.M., 1988. "Expert Systems and Systems of Experts," Chapter 12 in W.J. Coffey, ed., Geographical Systems and Systems of Geography: Essays in Honour of William Warntz, Department of Geography, University of Western Ontario, London, Ontario.

DISCUSSION AND EXAM QUESTIONS

1. What are the main considerations to be aware of in computer contouring? What are the key aspects for the design of an expert system to aid in choosing a computer contouring algorithm within a GIS? How long do you think it will be before such expert systems become widely available?

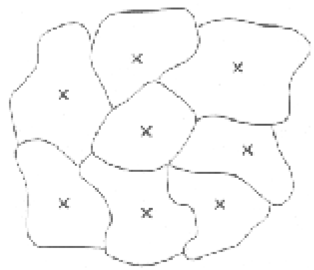
2. Describe how Tobler's pycnophylactic method differs from volume-preserving overlay. What model of the underlying spatial distribution is assumed by each? Give examples of phenomena and application which fit each method's assumptions.

3. Describe the application of areal interpolation in political districting.

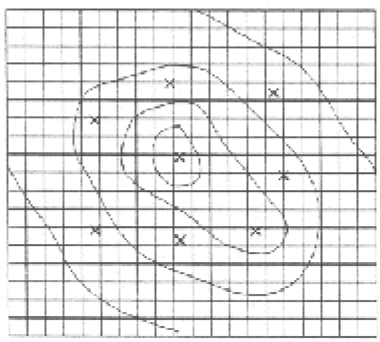
4. One test of a spatial interpolation method is that its results would be judged as equal or better to hand contouring by a specialist, e.g. a field geologist, with detailed knowledge of the phenomenon being mapped. How well do the methods discussed in these two units do against this criterion?

Last Updated: August 30, 1997.

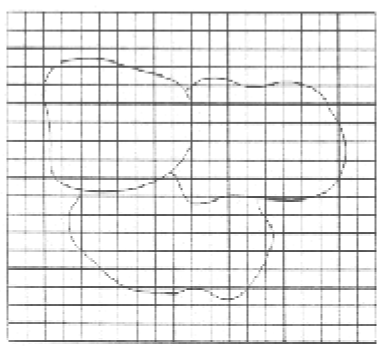
UNIT 41 IMAGES



Source Zones with Centroids and Population Density



Interpolated Population Density Surface with Grid



Target Zones on Grid