Using Zone Design as a Spatial Analysis and Modelling Tool

Seraphim Alvanides
(Makis)
Stan Openshaw
(Stan)

School of Geography
Leeds University

Contents

- Background
- A Conjecture
- A generic Zone Design Process
- Case Study 1
    - visualisation
- Case Study 2
    - modelling
- Case Study 3
    - spatial analysis
- Conclusions

Objectives

- to investigate the extent to which Zone Design can be used as a geographical visualization, analysis and modelling tool
- to perform some preliminary empirical tests to illustrate this approach
- to identify ways of developing this approach to geographical analysis further

Why Bother with a new approach?

because

- ✔ current tools are either OLD or DEFICIENT or otherwise INADEQUATE for exploiting GIS environments
- ✔it is worth investigating how the explicit geographical component in spatial analysis may be strengthened
- ✔there has been relatively little innovation in the last few decades
- ✔a growing demand for useful geographic tools

Geographers have been slow to develop explicitly geographic geographical analysis methods

Some reasons for the neglect of geographic geographic analysis tools

- ✓ it is easier to use statistical techniques developed by others and widely available
- ✓ a failure to understand the geography of the geographical analysis task
- ✓ the high degree of difficulty in attempting something different
- ✓ a reluctance to go out on a limb!

Virtually the only widely used uniquely geographic geographical analysis (GGA) tool is.....

the MAP

Map based Visualization and Analysis is a simple but fundamentally flawed technology

- X maps can tell lies
- X map stories can be manipulated
- X the analysis task is left to the viewer's eyeballs
- X it is NOT analysis!

The Modifiable Areal Unit Problem (MAUP) is an ADDED complication

- scale changes the level of generalization and thus what you see on maps
- aggregational variability is far more devastating since often far more than a billion different sets of results for each scale!

Ecological Fallacies are essentially a great irrelevancy which geographers can probably do

✔ only a problem if the focus is on making inferences about people from aggregate data for places

✔ ecological inference is a prime candidate for illegality under DPA etc legislation

the analysis of aggregated data is currently the principal need for Geographers

Aggregate data created for user defined flexible geographies is on the increase

✔ avoids the release of personal or sensitive individual data

✔ preserves confidentiality

✔ reduces data volume

the problem is....

BUT

BUT

we are not very good at the geographic analysis of SPATIALLY AGGREGATED data

Flexible user defined geographies and GIS create mega additional problems!

- ✗ no more FIXED data sets
- ✗ no more small or inadequate data sets that can be blamed
- ✗ an increase in analysis needs
- ✗ enhanced MAUP effects

The principal problem is an almost complete absence of suitable Geographic Geographical Analysis Technology (GGAT)

Lets return to a basic proposition...

the

MAP

is a useful

GGAT

and

then see what needs to be done to improve it as a geographic geographical analysis tool

There are various ways of tackling this challenge

- ➡ via interactive map and linked graphics based exploratory analysis systems
- ➡ via visualization and animation
- ➡ via zone design

This paper is about developing a Zone Design based approach to GGAT

Curry, Tobler et al pointed out long ago that the map acts as a kind of higher order filter on the data it shows

- their error was to treat it as an electronic engineer might by seeking to apply various mathematical transformations
- they neglected the essential geography of the problem!

Openshaw (1984) revived this filter idea by regarding the zoning system as a kind of aerial that could be tuned to detect specified patterns or relationships

Two BASIC ideas

(1) the zoning system could be calibrated or engineered or designed to detect specific phenomena

(2)the shape and configuration of the resulting optimal zoning systems might itself be of interest as a pattern or relationship visualization

An interesting conjecture or a daft idea?

- needed to wait for GIS before it can be tested
- needed digital map data
- needs fast computers and powerful algorithms for zone optimization
- needs a rationale for why such an unconventional approach may be worthwhile

the
time
is
now
RIPE
to
give
it
a
try

The Zone Design Problem

- The task involves the aggregation of data for **N small areas** (points or individuals) to **M regions** (M < N) subject to a set of constraints to ensure that N areas are assigned to only one region and that all the members of a region are geographically connected

- If N is large then there is almost an inifinity of alternative possible aggregations that will yield M regions

Zone Design can be viewed as a special type of combinatorial optimization problem

optimise **F(Z)**

where

F(Z) is any function defined on the data produced by applying zonal definitions (**Z**) to some data

Z is the aggregation of **N** zones into **M** regions

subject to two types of constraints:

(a) to ensure coverage and within zone contiguity

(b) any other user imposed design restrictions on the properties of the M zones or the data they generate

What type of optimisation is it?

- **Its characterised as being**
  - integer variables
  - composed of many variables (N*M)
  - nonlinear
  - probably discontinuous
  - non-convex
  - constrained

- **In other words...**

its..

HARD!

to solve and lies far outside what conventional optimisation methods can handle

Many spatial planning problems can be handled via zone design merely by changing the F(Z) term and the nature of the constraints

- Political redistricting
- School Districting
- Design of sales territories
- Network optimization
- Location-allocation
- Identification of policy areas
- Functional regionalisation

Basic algorithms were first produced 20 years ago

- a mild steepest descent heuristic (AZP) worked well on small problems (N=73) in 1977
- new algorithms developed by Openshaw and Rao (1995) based on Tabu search and Simulated annealers
- Simulated Annealing worked best

BUT

BUT

Even on simple no constraint problems it takes two orders of magnitude more compute time and ***three or four orders of magnitude longer*** on more realistic user constrained problem formulations

A simple Simulated Annealer

Step 1. Set temperature **t** high

Step 2. Perform local search

accept improving moves but also

accept non-improving moves with probability dependent on **t**

Step 3. Continue for a while

Step 4. Reduce temperature **t**

Step 5. Repeat Steps 2 to Step 4 for 100 times or so iterations

Simulated Annealing is a highly serial method of optimization

parallel versions needed to speed it up

Openshaw-Schmidt hybrid genetic simulated annealer

Step1. Initialisation

- Generate a population of P temperatures (P > number of processors)
- Compute best result for each temperature

Step 2. Repeat until convergence

- when a processor is free
  - update population values
  - use Genetic Algorithm to breed a new temperature
  - apply annealer at that temperature

Case Studies use 1991 census data for Leeds - Bradford region

- there are 2,315 small census areas (enumeration districts)
- these can be aggregated into 63 wards and 2 local authorities (Leeds, Bradford)
- the results you get for these areas depend on the zoning system you use

Conclusions

- the map in the form of a zoning system is a wonderfully powerful GGAT

- we just need to spend more time discovering how best to use it by developing appropriate and complementary spatial data processing tools

ZDES and related developments will not instantly solve all the problems that exist

BUT

it does appear to provide a framework for at least thinking further about the possibilities

there is also still much to be done!