Boundary Issues in Constructing a Hierarchical Regional Space Model for Contemporary China

Mark Henderson
Graduate student, Environmental Science, Policy, and Management, UC Berkeley
GIS analyst, G.W. Skinner Research Team, UC Davis
USA

This presentation draws on my experiences in managing a geographic information system (GIS) with Dr. G. William Skinner's research team at the University of California, Davis. Since this session is dedicated to the discussion of "boundary" topics, I thought that I would address a few practical areas related to managing boundaries with our GIS. These are:

- County boundaries. At Davis a substantial portion of our analysis is based on socioeconomic data for county-level administrative units in China. I'll talk about how we manage that data set.
- Township boundaries. A recent project at Davis has been able to compile maps of township boundaries for Shandong province, but for the most part, township boundaries are not available for GIS use. I'll discuss a couple of ways that township boundaries can be simulated.
- Boundaries of functional urban systems. One of the goals of the project at Davis is to delimit the approximate boundaries of urban systems at different hierarchical levels, from China's 9 macroregions down at least as far as regional city systems. I'll show an example of that.

In conclusion I'll briefly introduce two applications of the Hierarchical Regional Space (HRS) model that we have constructed.

HRS: A Model of Spatial Differentiation

First, a word about the HRS model itself is in order. The Hierarchical Regional Space model builds on a number of elements of modern geographical thought, including central place theory, regional systems theory, and innovation diffusion theory. The model attempts to approximate the real spatial structures that underlie socioeconomic phenomena in a predominantly agrarian society like China of 1990.

Without adequate consideration of these spatial structures, too many studies of China continue to consider very coarse aggregations and draw overly broad conclusions. If we divide China into simple categories like "urban" and "rural," or "coastal" and "interior," we mask a tremendous amount of variation within these categories. To use the term that post-Mao reformers criticized, those studies are "painting with one broad brush."

But increasingly, there is sufficient data on many topics to support a finer differentiation. The HRS model provides a logical framework for such differentiation. Furthermore, with its ability to situate any locale in China within its framework, the HRS model provides an excellent means to contextualize local-area field studies and to make valid comparisons across space and time.

County Boundaries

As I mentioned, data for county-level administrative units—*xian* and *shi*—play an important role in the construction of the HRS model. In collaboration with Lawrence Crissman and others, we have assembled a GIS coverage of county boundaries for China as of 1990. We manage the GIS coverage with Arc/Info GIS software, and separately use SAS statistical software to maintain a tabular database of statistical information about each county. Each county record in our database is identified with a 6-digit identification number based on the *guobiao* code used by the Chinese government. We use that ID number to link the tabular database to the polygons in the GIS that represent each county.

A nearly universal problem when analyzing politically defined units like counties is that these units are not defined consistently from place to place. In China, some rural counties (*xian*) include significant urban areas, while in other places, equivalent urban areas have been separated out into their own county-level unit, a municipality or *shi*. We would compute quite different statistics for these two kinds of cases even if the actual situation on the ground were identical.

Dr. Skinner's solution to this version of the 'Modifiable Areal Unit Problem' is to recombine some county-level units to achieve a more uniform definition of a county. We manage these aggregations as Arc/Info regions, which allows us to keep all of the underlying boundary and polygon data in a single Arc/Info coverage. Each county can be made up of multiple polygons, but linked to a single database record for that county.

We are also interested in comparing our 1990 data with data from other years. In particular, we have data from 1982 and 1985; thanks to Dr. Crissman's work, we also have boundary data for 1996. Again, our approach is to store all of the underlying boundary and polygon information in a single Arc/Info coverage, and define each years' counties as Regions.

When we want to compare data across years, though, we must ensure that we are comparing units that did not change from one year to the next. Unfortunately, boundary changes are not that uncommon in China over the past two decades. Unless we can reconstruct county-level statistics from subcounty data, the best we can do is to create 'cross-time comparison units' comprising the 'least common denominator' area for the two time periods.

Cost Surfaces

Again in collaboration with Dr. Crissman, we have access to GIS files of roads, railways, and rivers and lakes. At Davis, we researched which rivers and lakes are considered navigable, and added them to the roads and railways to form a transportation network. We are also using a GIS file that provides the latitude and longitude of some 12,000 cities and towns in China. These are identified by an 8-digit *guobiao* code. Each town is linked to a node on the transportation network, which allows us to calculate the transportation distance between any two towns.

And finally we have made use of a 1km digital elevation model of China, available from the U.S. Geological Survey. Using that file we adjusted the lengths of each segment in the transportation network to account for the different costs of going up over mountains or flat along river valleys. In conjunction with data about the relative efficiency of the different modes of transportation represented in our network this lets us calculate not only the *distance* between any two towns, but the *relative cost* of that travel.

By combining our transport network and the digital elevation model into a single raster file, we produce a transportation cost surface. That surface indicates the relative cost of travelling across each 1km raster cell. From this data, we can compute the relative travel costs not just among our 12,000 cities and towns, but among any two points in China. This sort of information can be of use in estimating two other types of boundaries: township boundaries and the boundaries of urban systems.

Township Boundaries

There are a few sources of statistical data aggregated at the township level in China, but it is difficult to find maps of township boundaries that can be used in a GIS. Again I should note that at UC Davis, Dr. Skinner, geographer Zhu Hongxing, and GIS analyst Douglas Messenger have just completed digitizing an accurate map of township boundaries for Shandong province; none of the methods I'm describing for approximating township boundaries are actually being used in the UC Davis project now.

But if you aren't so fortunate as to have accurate township boundaries, what can you do? One approach is to digitize the point locations of the township seats and then generate Thiessen polygons around these points. A study that came out a year or so ago did this for Tibet using Dr. Crissman's *zhen* point data. This is similar to the technique commonly used in the United States and Britain for displaying census data before census tract and block boundaries were readily available.

But Thiessen polygons present some disadvantages. Uncontrolled, they arbitrarily cross county boundaries or other natural boundaries. Thiessen polygons assign locations to the township seat that is closest as the crow flies. If we accept the premise that township boundaries are often related to local topography and transport patterns, it may be possible to improve on this by using the cost surface model. Using a cost-allocation algorithm, we can assign locations to the township seat that has the lowest transport cost. Since township boundaries never cross county boundaries, we can use county boundaries as barriers to this algorithm. Obviously, higher resolution and quality of the cost surface model will improve the results.

Boundaries of Urban Systems

As I said, we are not approximating township boundaries in the project at UC Davis, but we have made an effort to delineate the boundaries of macroregions and regional city systems—not administrative units but functional economic units. The first step in this task is a central place analysis, in which all 12,000 cities and towns are assigned to a level in the urban hierarchy. The transportation network provides the context for these assignments.

Based on the configuration of settlements at each level in the urban hierarchy, we assign each locale in our database to a position along an Urban-Rural Continuum.

Theoretically, we expect the Urban-Rural Continuum to look something like the concentric circles around each regional city.

Just as we expect each regional city system to display somewhat concentric patterns of social and economic development, the macroregions as a whole also display core-periphery patterns. Using county-level statistical data, we have approximated the boundaries of seven Core-Periphery Zones in each of China's nine Macroregions.

The GIS and statistical procedures that we used to arrive at the Core-Periphery Zone boundaries are documented elsewhere. To conclude this presentation, I'll provide a few examples of analysis using these Core-Periphery Zones and the Urban-Rural Continuum as two dimension of the Hierarchical Regional Space model

Reported fertility in HRS

China's population planning program, and especially its One Child Policy, are of great interest to many researchers. With the HRS model we can ask how reproduction varies in space across China. In fact, Dr. Skinner and his collaborator in Beijing, Dr. Yuan Jianhua, have been asking that for several years now and have come up with some very interesting results. Here I'll show a simple example. Leaving aside the question of whether all births are actually reported to the Chinese authorities, we used 1990 census data to reconstruct how many women aged 30 or more have had three or more children.

First we differentiate this statistic according to the Urban-Rural Continuum. We see that this measure of fertility varies sharply across 9 classes of the Urban-Rural Continuum, ranging from less than 20% in the most urban cities (like Shanghai) to over 50% in the most rural settings.

Next we differentiate this statistic by Core-Periphery Zones. We see that this measure of fertility varies sharply across 7 zones, ranging from less than 25% in the inner core (again, around Shanghai) to over 60% in the most peripheral two zones.

A particular strength of the HRS model is that we can cross-tabulate these two dimensions to produce a matrix of 9x7 or 56 cells. When this statistic is computed for each cell, we can further differentiate this (or any) statistic by its spatial context.

Urban Land Use in HRS

My own research is concerned with land use change, particularly the conversion of forests and agricultural lands to urban uses. One source of data on urban land use is this type of satellite image, which captures the nighttime lights of developed urban areas. Researchers at NOAA, UC Santa Barbara, and the University of Denver have used such images to develop an algorithm for estimating urban populations based on the amount of light and the socioeconomic level of a given country. The results are pretty good for the United States and for some small countries. But for China, their algorithm produces an estimate of 1.6 billion people—over 300 million more than the commonly accepted estimates.

Those of us who work on China, though, realize that using a single nationwide statistic to compute China's population in this algorithm is unrealistic. We know that there is a lot of variation across China, and the HRS model captures a lot of that variation. For example, we may be able to use core-periphery zones to improve on these projections.

While I haven't yet computed the population statistics, I can report that light emissions do vary systematically by core-periphery zones. For the Lower Yangzi macroregion, we see about 20% urban land area in the inner core, 10% in the second tier, and dropping off sharply to less than 1% thereafter. Combined with socioeconomic statistics for each zone, we should be able to further calibrate models that use these satellite images.

Notes: Maps and charts prepared for this presentation will be available via the project's web site, http://qing.ucdavis.edu. I gratefully acknowledge the advice and data provided Dr. Skinner, GIS analyst Douglas Messenger, and statistician Wei Wang at UC Davis; Dr. Lawrence W. Crissman at Griffith University, Australia; and Dr. Paul Sutton at UC Santa Barbara.