A Hierarchical Regional Space Model for Contemporary China

1. CONCEPTUALIZING HRS AND CONSTRUCTING TABULAR AND SPATIAL DATAFILES

Mark Henderson
Department of Environmental Science, Policy, and Management
University of California, Berkeley

G. William Skinner
Department of Anthropology
University of California, Davis

Lawrence W. Crissman
Australian Centre, Asian Spatial Information and Analysis Network
Griffith University

Paper prepared for the Geoinformatics '99 conference
China Data Center, University of Michigan, Ann Arbor
20 June 1999

A HIERARCHICAL REGIONAL SPACE MODEL

This panel presents an overview of the concepts and methods used in operationalizing a multilevel model of spatial differentiation for the analysis of social and economic patterns in contemporary China. In this first paper we introduce the theoretical concepts behinds the model and review the spatial and statistical data prepared to construct it. The papers that follow in this panel will demonstrate how these data sets are used in central place analysis, constructing an urban-rural continuum, delineating regional systems and core-periphery structures, classifying settlement types; and positioning households in hierarchical regional space. A second panel will introduce substantive applications of this model to research in economics, anthropology, and sociology.

We refer to the spatial framework described by this model as Hierarchical Regional Space or HRS (Skinner, 1994). HRS elaborates on some fundamental elements of modern geographical thought, including central place theory, spatial autocorrelation, regional systems theory, and diffusion theory. For agrarian societies, Christaller's (1933) central place theory predicts the emergence of a hierarchy of settlements, with each level of the hierarchy providing distinctive services and attaining corresponding levels of development. Individuals on the landscape orient their economic activities to specific central places at each hierarchical level in accordance with the services provided there: to a nearby town for daily purchases, to cities for products like furniture needed with less frequency, and to metropolises for specialized services such as higher education. It follows that the hierarchical levels providing less common services require a correspondingly wider hinterland, and that for reasons of transportation efficiency the hinterlands at each level become nested. Economic activities in this hierarchy develop hand in hand with a web of social networks, making a central place analysis a useful starting point for an investigation of social patterns.

In this investigation we take advantage of another critical feature of agrarian societies, namely, that most economic, social, and demographic variables are spatially autocorrelated, *i.e.*, that they co-vary systematically through regional space. Any number of social and economic variables collected from regions across the landscape can be shown to be autocorrelated, a fact that confounds many standard statistical procedures. Our approach to this problem is to make explicit the spatial relationships among regions.

The Model of Hierarchical Regional Space

Theoretical elements:

- Central Place Theory
- Spatial Autocorrelation
- Regional Systems Theory
- Diffusion Theory

Model dimensions:

- LUH, level in the urban hierarchy
- URC, urban-rural continuum
- CPZ, core-periphery zoning
- STU, level in the urban hierarchy

County-level data are analyzed in a CPZ H URC matrix

Household-level data are analyzed in a CPZ H LUH H STU matrix

The internal structure of regions must also be considered. Anticipating regional systems theory, von Thünen (1826) describes how zones of high- to low-value economic activity fill the regions around cities. Seen as a core-periphery structure, these zones can be characterized in terms of agricultural intensity, transport efficiency, and other phenomena.

A final concept incorporated into the HRS model is diffusion theory. Introduced by Hägerstrand (1966), this theory draws attention to the diffusion of innovations along two paths. Innovations of all kinds are introduced in cities or towns and then travel out from the central places to their rural hinterlands like ripples on a pond. But in addition to this spatial path, innovations also travel down the central place hierarchy to dependent towns, where the cycle of diffusion into the hinterlands begins again. We suggest that successive diffusion waves of innovations along these paths tend to reinforce the regional systems hierarchy and core-periphery structure through time. Even as agrarian societies industrialize, elements of this structure may persist for generations.

Within this framework, specific social and economic variables express their spatial pattern most strongly at certain spatial scales. The HRS model aims to account for the contribution of each of three scales—high-order regional systems centered on metropolises, middle-range regional systems centered on cities at intermediate levels of the urban hierarchy, and local systems centered on low-order cities and market towns. The HRS model rests on three spatial dimensions which together capture variation at each level, while controlling on the others: core-periphery zoning (CPZ) and an urban-rural continuum (URC) based on county-level data, and a local systems index (LSI) derived from higher resolution data. We now introduce the statistical and spatial data files upon which this model is constructed.

STATISTICAL AND SPATIAL DATA

In the construction of the HRS model for contemporary China, we begin with three data files containing demographic and economic data for the year 1990. Tabular data for each of these files have been entered into a computer database managed with SAS (SAS Institute, Cary, North Carolina, USA) statistical software. For spatial analysis and cartographic display, two of these files are linked to a digital map file managed with Arc/Info (ESRI, Redlands, California, USA) GIS software. To facilitate our analysis, we have also developed additional GIS files describing China's physiography, hydrography, and transportation network. The Australian Centre for the Asian Spatial Information and Analysis Network (ASIAN) at Griffith University was primarily responsible for the development of the spatial data files (Crissman, 1997), which have been compiled at a nominal scale of 1:1,000,000 (for an estimated spatial accuracy of +/–650 meters).

ChinaA County-Level Data File and MC Units

The first of these data files, which we call ChinaA, is composed of county-level data from the China's 1990 population census (published volumes of tabulations for each province, most of which were published by the China Statistical Publishing House in 1992 with the title *1990 nian renkou pucha ziliao*) and other contemporaneous published sources. In this file, a record for each of China's approximately 2,800 county-level administrative units includes its 6-digit *guobiao* (national standard) identification numbers from the year 1990; codes representing the administrative status of the county; and more than one hundred raw variables giving social, demographic, occupational, agricultural, and industrial data about the county. To this file we have added several hundred additional analytical variables derived from the raw data—*e.g.,* sex ratios and percentage of land under cultivation or irrigation.

For spatial analysis and cartographic display, the ChinaA file is linked to a digital map file containing the boundaries of China's counties as they existed in 1990. Each county polygon is coded to indicate the 6-digit *guobiao* numbers from the years 1982, 1990, and 1996, with the 1990 *guobiao* number used to match that unit's record in the ChinaA file. This file is derived from the digital map product completed by ASIAN. For the present analysis we have modified a small proportion of the boundaries and *guobiao* numbers in order to replicate the status of counties at the time of the midyear 1990 census, with reference to provincial atlases, the Atlas of Cities of China, and other published sources. Modified boundaries in our file have been coded as to the source we used to make these modifications.

A significant challenge in making use of data for administrative units such as provinces or counties is the underlying assumption that all units are comparable, or, stated another way, that the socioeconomic conditions reported for a given unit are representative of the entire area within that unit's borders. It is well established that how areas are aggregated markedly affects the geographic patterns that appear when the aggregated data are mapped (*e.g.,* Monmonier, 1996: 140-144). In the case of China, it is evident that any analysis of socioeconomic data that relies on data aggregated at, for example, the provincial level obscures important geographical patterns (Denny, 1991: 207; Skinner and Yuan, 1998: 8).

The county-level data in our ChinaA data file provides substantially more detail than provincial data—the most detail that is likely to become available through the Chinese central government's statistical reporting system (Dernberger, 1994)—but we must still contend with the areal aggregation problem, particularly around major cities. In a GIS analyst's ideal world, all counties would be of identical size and shape, and each would comprise a uniform population distribution. Even the great standardizer Qin Shihuang was unable to impose that degree of order on the Chinese landscape. China's 1990 map shows some municipal boundaries drawn tightly around the built-up city area, while other municipalities extend their boundaries loosely into the surrounding countryside. The reported population densities and other statistics for the "tight" and

Statistical and Spatial Data Files

ChinaA
County-level data
2,800 counties aggregated into
2,300 MC Units

ChinaT
Central Place data
14,000 cities and towns

ChinaS
1% sample of 1990 Census
Family structure and occupation data
12 million returns grouped by household

"loose" municipalities may differ significantly even when the actual places in question are quite similar.

For certain "loose" high-order municipalities we have sufficient data to be able to disaggregate large, primarily rural periurban *qu* from the remainder of the municipalities' *shixiaqu* (districts under municipal jurisdiction). In other areas we attempt to overcome the areal aggregation problem by merging "tight" municipalities with adjacent periurban counties. In selecting districts to be detached or merged, we require clear evidence from published maps and tabular data that their urban-suburban character is comparable to the conditions seen in more typical municipalities.

Among China's 2,800 counties, about five hundred of these changes have been made to create "municipal-compatible" or MC units. (In eighty percent of these cases, the urban unit was too small to be mapped at our scale, and so the digital map file had already been prepared showing an intermediate level of aggregation, called "merged *qu*" or MQ units.) We assign these units a new identification number compatible with the 6-digit *guobiao* codes and include data for them in the ChinaA statistical tables. In the Arc/Info GIS file, we use different region subclasses to manage these aggregations of the original county polygons. All subsequent analyses with county-level units refer to these MC units.

ChinaT Central Places Data File

Our second data file, ChinaT, includes records for over 12,000 cities and towns. Every central place in China that was the central city of a *shi* (municipality) or designated as a *zhen* (town) as of 1 July 1990 is represented in this file; in addition, we have added records for *qu* (district) centers and for county seats not already in the file as *zhen* towns. The tabular data in this file are drawn from the *Zhongguo chengshi tongji nianjian* (Statistical Yearbook of China's Municipalities, published by the State Statistical Bureau) and from the 1990 census tabulations of *zhen* data. Records in this file include a 2-digit identification number added to the 6-digit county or *qu guobiao* code from 1990, codes representing the administrative status of the central place, population totals in a number of categories from the 1990 census, and, for municipalities, variables that indicate central functions—*e.g.*, types of industry, institutional capacity in finance and education, and numbers engaged in certain professions.

As with the ChinaA file, the ChinaT tabular file is linked to a GIS file for the purpose of spatial analysis and cartographic display. In this case, the GIS file consists of point locations of cities and towns. ASIAN performed the original georeferencing of these cities and towns drawing on published provincial atlases, transportation maps, and postal atlases. We have made additions and modifications with reference both to county boundaries and to the transport network, facilitating network analysis.

ChinaS Data File (1% Sample of the 1990 Census)

ChinaS, the third and largest of our three tabular data files, consists of over 12 million records comprising a one percent sample of China's 1990 population census. Access to this data file has been made possible in collaboration with the Beijing Institute of Information and Control, initiated with support from the U.S.-China Cooperative Research Program of the Luce Foundation. The one percent sample was selected from a master file listing some 1.2 billion returns ordered by identification number. The 18-digit identification number specifies each person's household, settlement, and position in the administrative hierarchy. The returns of all persons in every hundredth household are included in the ChinaS data file.

It is obviously not feasible to map the location of individual households across all of China, and thus there is no GIS counterpart to this file. However, we have been able to link ChinaS records to the settlement and county levels, corresponding to our ChinaT and ChinaA data files. Using this linkage we are able to position individual households within the HRS model, as the final paper in this panel will explain.

Physiography

Earlier analyses have demonstrated that China's high-order regional economies took shape within the major physiographical regions associated with drainage basins (Skinner 1977, 1994). But in the present analysis, we deliberately did *not* use any ecological or physiographical data in the delineation of regional systems. Only the socioeconomic data in the ChinaA, ChinaT, and ChinaS data files already described here were used for that purpose. An interesting outcome of our work is to demonstrate the extent to which contemporary macroregional systems resemble the historical and physiographical regions.

The one instance where we do need to consider physiography is in the construction of our transport network, which is used as background information in the central place analysis. To represent the terrain we made use of the global 1 km digital elevation model, GTOPO-30, produced by the U.S. Geological Survey and other agencies. We used GIS techniques with this data set to estimate the increased lengths of arcs that climb over mountain passes compared with those which remain along relatively flat valley floors.

Transport Network and Hydrography

Construction of GIS data on China's roads, railways, and waterways was initiated by ASIAN under a license from the Institute of Geography and the Science Press, Chinese Academy of Sciences, publishers of the Land Use Map of China (Crissman, 1997; Skinner, Henderson, and Messenger, 1998). Roads and railways shown on that map series and in other sources were digitized as arcs and coded according to several classification systems (*e.g.*, paved or unpaved; national, provincial, or local control; electrified and double-tracked; etc.). Rivers and lakes were digitized as arcs and polygons. Recognizing the importance of water transport in China, we assigned codes to waterways based on their navigability as shown on several published sources including the Atlas of China, the Transportation Map of China, and provincial atlases.

Arcs from separate GIS files of roads, rails, and waterways were selected for inclusion in the transport network. Superhighways (none of which were in service as of our analysis date of 1990) and non-navigable waterways were excluded. To accommodate transportation across major lakes and in near-coastal waters, two additional transportation components was developed. A 10 km mesh was inserted in place of all lakes and wide rivers which had at least one shoreline arc coded as navigable. Similarly for near-coastal waters, a 20 km mesh was inserted into a 50 km buffer zone along China's coastline.

To this collection of transport arcs, nodes were added where arcs most closely approached each of the 12,000-plus points representing cities and towns in the ChinaT file. Each of these nodes was coded with the 8-digit *guobiao* code for the city or town. Additional uncoded nodes were added at the boundaries of UTM zones so that each arc falls into only one UTM zone. The resulting GIS file consists of some 65,000 arcs.

To facilitate the calculation of transport distances, arcs were coded with their planimetric length in an equidistant projection and with an estimate of their length over the topographic surface of the earth. The Universal Transverse Mercator (UTM) coordinate system was selected to provide the equidistant projection because it produces a scale error of not more than 0.1% within each UTM zone; eleven UTM zones (43-53) were used to cover all of China. As mentioned above, arcs were "draped" over the terrain represented by the 1 km digital elevation model to calculate the surface lengths of each segment of the transport network.

By weighting transport arcs according to their classifications, and considering the calculated equidistant surface length of each arc, we are able to determine the relative transport costs among any cities and towns in China. The resulting transport network digital map is critical to our central place analysis and finds other uses in modeling transportation and delimiting metropolitan systems.

CONCLUSION

The statistical and spatial data files we have assembled represent a high-resolution economic and demographic profile of China in 1990. Our use of geographic information systems is key to maintaining different aggregations of data, particularly in overcoming the spatial aggregation problem seen in tabular analyses of county data, and in identifying the spatial connections among counties and central places across our transport network. These data files will be a rich source for further investigation of social patterns in China.

Our immediate objective in assembling these data files is, of course, to support the construction of the Hierarchical Regional Space model. As we hope will become apparent in the following papers, that model provides a useful framework for analyzing this collection of data on 2,800 counties in ChinaA, 14,000 central places in ChinaT, 65,000 transport links in our transport network, and 12 million households in ChinaS. Construction of the HRS model begins with the central place analysis, making use of ChinaT and the transport network. Subsequent steps use ChinaA to delineate core-periphery structures and ChinaS to situate households within the full model. In each case, the conclusions we draw about social patterns in China rest upon the statistical and spatial data in these files.

References

Christaller, Walter, 1933. *Die zentralen Orte in Süddeuschland.* Jena: Gustav Fisher. Translated by Carlisle W. Baskin, 1966, as *Central Places in Southern Germany*. Englewood Cliffs, NJ: Prentice-Hall.

Crissman, Lawrence W., 1997. Australian Centre of the Asian Spatial Information and Analysis Network [catalog], Nathan (Brisbane), Australia: Griffith University. Available online: http://www.asian.gu.edu.au/, accessed 1 December 1998.

Denny, David L., 1991. "Provincial economic differences diminished in the decade of reform." In U.S Congress Joint Economic Committee, *China's economic dilemmas in the 1990s: The problems of reform, modernization, and interdependence*, volume 1. Washington, D.C.: Government Printing Office, pp.186-208.

Dernberger, Robert F., 1997. "Provincial Yearbooks' County-level Data Files, personal comments and evaluation." China in Time and Space project, University of Washington. Available online: http://citas.csde.washington.edu/data/provyrbk/yrbkdoc.htm, accessed 1 December 1998.

Hägerstrand, Torsten, 1966. "Aspects of the Spatial Structures of Social Communication and the Diffusion of Information." Papers of the Regional Science Association 15, p.27.

Monmonier, Mark, 1996. *How to Lie with Maps*, second edition. Chicago: University of Chicago Press.

Skinner, G. William, 1977. "Cities and the Hierarchy of Local Systems," in G. William Skinner (ed.), *The City in Late Imperial China*, Stanford: Stanford University Press, pp.275-351.

————, 1994. "Differential Development in Lingnan," in Thomas P. Lyons and Victor Nee, (ed.s), *The Economic Transformation of South China: Reform and Development in the Post-Mao Era.* Cornell East Asia Series no.70.

Skinner, G. William, and Yuan, Jianhua, 1998. "Spatial Differentiation in Reproductive Behavior in the Lower Yangzi Macroregion of China, 1966-1990." Paper presented at the Social Science History Association, Chicago, 21 November.

Skinner, G. William, Mark Henderson, and Douglas Messenger, 1998. Geographic Information Systems metadata, Davis: University of California. Available online: http://qing.ucdavis.edu/, accessed 1 December 1998.

Von Thünen, Johann Heinrich, 1826. *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.* Berlin. Translated by Carla M. Wartenberg, 1966, as *Isolated State.* Oxford: Pergamon.