A Hierarchical Regional Space Model for Contemporary China

- DELINEATING REGIONAL SYSTEMS AND CORE-PERIPHERY STRUCTURES

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

Douglas M. Messenger
Department of Anthropology and Visualization Laboratory
University of California, Davis

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

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

OBJECTIVES FOR THE REGIONAL SYSTEMS ANALYSIS

Regional systems analysis is the third step in building the model of Hierarchical Regional Space, following the steps of central place analysis and construction of the rural-urban continuum index (URC). The objective of this regional systems analysis is to delimit high-order regional systems and reveal the internal core-periphery structures of each. Here we use socioeconomic data for county-level administrative units, along with the results of the preceding two steps, to construct a Regional Systems index (RSI), which is used to determine the limits of high-order regional systems of China in 1990 and to delineate their core-periphery structures.

In this step we examine the behavior of socioeconomic variables in macroregional space. The RSI is an attempt to isolate the contribution of the core-periphery patterns that operate on the macroregional scale, including the variation embedded in the three highest levels of our central place analysis. This section introduces three procedures that we use to construct the RSI and subsequently assign counties to core-periphery zones within each macroregion:

- deurbanizing input variables that will make up the RSI;
- testing the input variables for expression of a core-periphery pattern; and
- working interactively with cross-tabulated RSI and URC values to arrive at optimal core-periphery zone assignments.

When socioeconomic data are displayed in a matrix that cross-tabulates core-periphery zones and URC classes, most of them reveal diagonal progressions from the urbanized inner core to the very rural counties in the far periphery.

PROCEDURES

Selecting and "Deurbanizing" Variables

In constructing a Regional Systems Index (RSI) to represent each county's relative position in macroregional space, we have at our disposal the several hundred socioeconomic and demographic data variables included in the ChinaA data file. As we expect, owing to the principle of spatial autocorrelation, most of these variables are intercorrelated to some degree. In selecting potential components of the RSI for a given macroregion, we eliminate from consideration those exhibiting a curvilinear or other atypical pattern through HRS, and we aim to include a broad and diverse range of social and economic attributes. In the usual case we end up with ten to twelve of the most strongly intercorrelated variables. We may illustrate with the variables used to construct the RSI for the Upper Yangzi macroregion:

Illiteracy rate of females aged 15+
Females aged 6+ with no education or only primary education
Prime-age (20-59) men as a proportion of all males
Ratio of males aged 0-14 to females aged 20-54
Men in the manufacturing sector as a percentage of the male labor force
Male commercial employees as a percentage of the male labor force
Male service employees as a percentage of the male labor force
Level of agricultural mechanization (watts per unit of cultivated land)
Rural electricity use (watt-hours per unit of cultivated land)
Collective units' product (per capita)
Purchased agricultural output (per capita)

The RSI for this macroregion, then, is based on two education variables, two age-sex variables, three occupational structure variables, two agricultural input variables, and two agricultural output variables. While aiming to represent a comparable range of phenomena in constructing the indices for each region, we have chosen specific variables on the basis of their behavior in different parts of China. It should be obvious that candidate variables that strongly express the underlying structural pattern in one macroregion might be less strong in another region—for example, irrigation may be a critical determinant of agricultural development in some macroregions while being unimportant in others due to climatological differences.

To focus our attention on spatial patterns at a macroregional scale, we excluded from the list of potential RSI components any variables that are more strongly differentiated within middle-range city systems than within high-order regional systems. Even so, we must somehow control for variation at lower levels of the systems hierarchy. It is central to the regional-systems argument that internal differentiation analogous to the core-periphery structure of macroregions also characterizes nodal systems at each lower level of the hierarchy (Skinner, 1996). As described above, the urban-rural continuum index (URC), based on the structure of settlement patterns within each county, is useful in highlighting variation below the highest levels. We use the URC to control on this variation—“deurbanizing” the data variables—in a statistical procedure analogous to detrending time-series data. In practice, we calculate the correlation of standardized variables with the Substantive Urban Index (SUI), itself comprised of the variables most strongly differentiated within city-system space, which was used to guide the construction of the URC. Based on that correlation and the URC for each county, we subtract a portion of each data value. The resulting deurbanized variables highlight the differentiation attributable to differentiation at the macroregional scale.

Spatial Relationships among Counties: Contiguity

In testing for the presence of core-periphery patterns, we need to consider the behavior of variables among contiguous counties. We expect that, moving outward across contiguous counties from a macroregional core, a variable that displays the core-periphery pattern will change monotonically. Using our GIS files of county boundaries and transportation networks, we prepare for this test by producing a table showing all pairs of contiguous counties.

A basic definition of contiguous counties would simply include all those counties which share a common boundary. A geographic information system such as Arc/Info organizes its files topologically and is readily able to determine which two polygons fall on either side of a given boundary line. This information can in turn be used to produce a list of all counties that border a given county. But this definition gives equal weight to counties which share either long or short borders, or whose borders run along either well-traveled lowlands or impassible mountain ranges. GIS procedures could be applied to exclude short or mountainous borders, but instead we have chosen to take advantage of our digital transportation network to determine which counties are effectively contiguous.

The GIS data for our transportation network (including seven classes of roads, fourteen classes of railroads, and four classes of navigable waterways) is based primarily on the work at ASIAN (Crissman, 1997). Our major efforts have focused on researching the classes of navigable waterways (relying on published sources including the Transportation Map of China) and on matching transportation nodes to our ChinaT file of cities and towns. To determine which counties were connected by direct transportation links, we performed a GIS overlay of the transportation network with the county boundaries file. This procedure intentionally excluded transportation links that ran into a county for a short distance without coming to a mapped city or town. The result was a table of five thousand pairs of counties which are effectively contiguous in terms of direct transportation links—somewhat fewer than would have been considered using the simpler definition of shared boundaries.

Spatial Progression Test

A second procedure in selecting candidate variables for the RSI is to test the spatial ordering of data values for a core-periphery pattern on a macroregional scale. An inspection of choropleth maps of these data variables shows that, as is well known, they display a high level of spatial autocorrelation. Statistical tests such as Moran's *I* could be applied to confirm the degree of correlation, but would not tell us the scale of the correlative pattern (Odland, 1988). Instead, we make an explicit statement of how we would expect variables to behave if they display a core-periphery pattern, then use our GIS to apply what we call a spatial progression test, which quantifies or scores how well each variable meets this expectation. We then select those variables with the highest scores as inputs into a provisional RSI. The provisional RSI itself is then tested in the same way, and the entire process may be repeated to achieve an index with the highest score.

This procedure begins by arranging the county data values for a candidate variable in a list by rank order. Our theoretical expectation is that the first value listed would be for an inner core county. The second value listed should be for either a county contiguous to the first core county or a county in a secondary regional core. Similarly, all subsequent values should represent either additional core counties, or counties contiguous to those counties of higher rank in the list. Following this expectation, all data values would progress monotonically and concentrically outward from each regional core.

When executing this procedure, we assign a "perfect" score to all counties whose data values appear in this expected order. Counties whose values appear out of the expected rank order—*i.e.*, counties not preceded in rank order by a contiguous county—

are flagged. In subsequent processing of the list, as additional counties are found that do fall in the expected order, we check whether these new counties are adjacent to the flagged, out-of-order counties. When this occurs, the flagged counties are assigned scores based on the ratio of the new, contiguous counties' data values to the flagged counties' values. This scoring has the effect of emphasizing major discrepancies over minor deviations from the expected pattern. In some cases a cluster of several counties with similar values appears to function together as a core; such counties would be assigned relatively high scores.

Once all counties in the list have been assigned scores in this way, we summarize the scores to quantify the performance of the data variable or provisional RSI across all counties. Low scores for a given variable may indicate that it is a poor indicator of the core-periphery structure of a particular macroregion, while low scores for a provisional RSI may indicate that the component variables express different spatial patterns. This test enables us to select for inclusion in the final RSI the best-performing variables among four to five contenders in the same category (say, agricultural inputs).

DELINEATING MACROREGIONAL BOUNDARIES AND CORE-PERIPHERY ZONES

A choropleth map of an RSI constructed for all counties in China provides a first approximation of the limits of high-order regional systems. In this procedure, we continue by assigning each county to a core-periphery zone—from the highly urbanized inner core to the remote rural periphery—in the macroregion to which it is most closely aligned. (For the few cases where our central place analysis indicates that a peripheral county is Janus-faced astride a macroregional boundary, we assign the county to zones in two or three macroregions.)

For each of the nine macroregions, we construct a region-specific RSI using the procedures described above. In constructing different indices for each region, we are able to make use of different appropriate input data variables that further clarify the internal structures of each region. By making judicious cutpoints in an ordered list of RSI values for each region, we classify counties into seven provisional core-periphery zones. A GISproduced choropleth map of these zones reveals the core-periphery structures captured by the county-based RSI. To assist in fine-tuning the CPZ assignments, we calculate for each macroregion separately a substantive index from variables that are strongly autocorrelated in both dimensions (*i.e.*, at both scales, high-order regional systems and intermediate city systems) and display mean index values in the cells of a cross-tabulated matrix of the CPZ with URC classes. Working interactively between the map and the matrix, we make a final, putatively optimal assignment of counties to core-periphery zones. These assignments consider the larger spatial context, including the spatial alignment of a county's central places and transport network in relationship to the surrounding region.

For the purpose of comparing data across China, we have also generated a ten-zone core-periphery zoning scheme based on an RSI for China Proper as a whole. This allows data for all nine macroregions to be displayed in the same matrix. With this scheme, some macroregions include only five or six zones, while others include as many as eight or nine. While this scheme obscures particular spatial details within some macroregions, it has the advantage of permitting a comparison among zones in different macroregions. For example, the innermost core zone of the Upper Yangzi region may be comparable to the second concentric zone of the Lower Yangzi region.

DEPICTION AND ANALYSIS OF REGIONAL SYSTEMS AND THEIR INTERNAL STRUCTURES

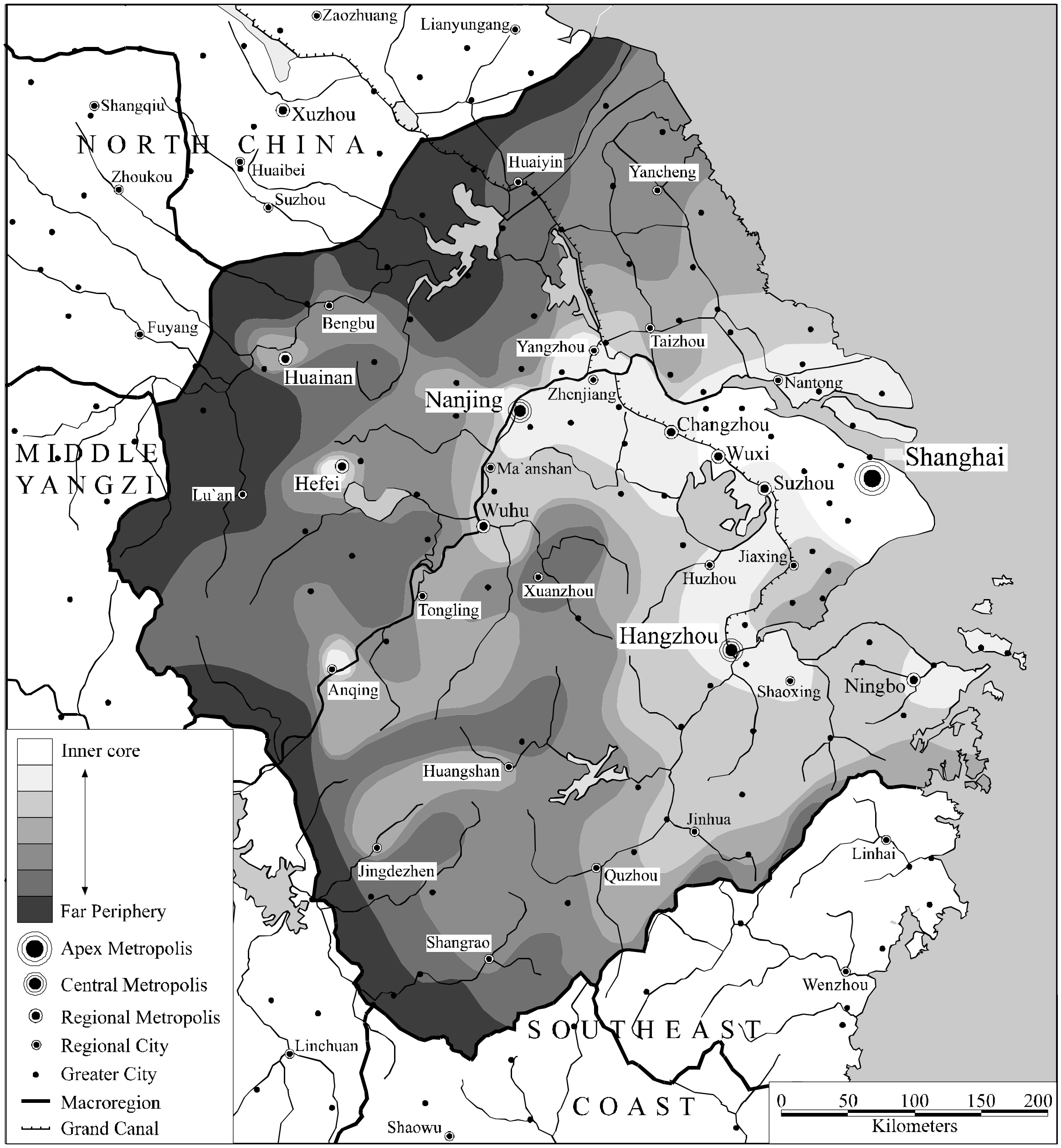
Delineation of Macroregional Boundaries and Illustrative Core-Periphery Zoning

While our county-level socioeconomic data by themselves can only be used to assign entire counties to certain core-periphery zones, we can, using our data on China's central places, transportation network, and topography, delineate in some detail the boundaries of the country's major regional systems. As stated above, we first use a choropleth map of the county-based RSI to approximate the limits of China's regional systems. We examine peripheral counties closely to determine to which region each is most closely aligned. In delineating the boundaries of regional systems, we cannot expect that regions are perfectly aligned with county boundaries; some peripheral counties must be divided among two or three systems. We draw on the results of our central place analysis, described earlier, to improve the regional systems boundaries between settlements and landscape features. Map 1 shows the boundaries derived for the nine macroregions of China Proper. The boundaries of regional city systems in selected macroregions are included on our display maps at the Geoinformatics '99 conference.

In settling on boundaries among China's regional systems, this procedure also estimates the portion of the county population that resides in each region. Each central place within a divided county is assigned to one or, for Janus-faced places, two or three regional systems. We apportion the population of these central places equally among the regions to which they are assigned. The remainder of the county's population—residents of *xiang* outside of our mapped central places—are assigned to regional systems based on the proportion of the counties' land area in each system. Subsequent analysis at the regional systems level using county-level data relies on the percentage of county residents assigned to each regional system; other raw data variables are allocated among systems proportionally.

With the boundaries of macroregional systems established in this manner, we have further approximated the boundaries of core-periphery zones within individual macroregions. Again, while the county-based core-periphery zone assignments cannot by themselves tell us how these zones fall on a scale more detailed than county boundaries, we make use of our central place, transportation, and topographical data to interpolate concentric zones for illustrative purposes. Within some of the larger of the nine macroregional systems, "internal peripheries" may delineate a nested subsystem with its own core and periphery. A GIS procedure analyzes our proposed illustrative ("smoothed") zones to maximize the degree to which they reflect the data-based county zone assignments. While we do not use the cartographic depiction of these zones for analytical purposes, they serve to illustrate the spatial pattern underlying the county-based data which may be made more apparent as more finely detailed data become available.

Map 1. China's macroregional systems in relation to provinces, showing metropolitan cities, 1990



Map 2. Core -periphery structure of the Lower Yangzi macroregion, showing high-order cities and major waterways, 1990

The results of this analysis for the Lower Yangzi macroregion are shown in Map 2, and for additional macroregions on our display maps at the Geoinformatics ‘99 conference.

Analysis of Counties

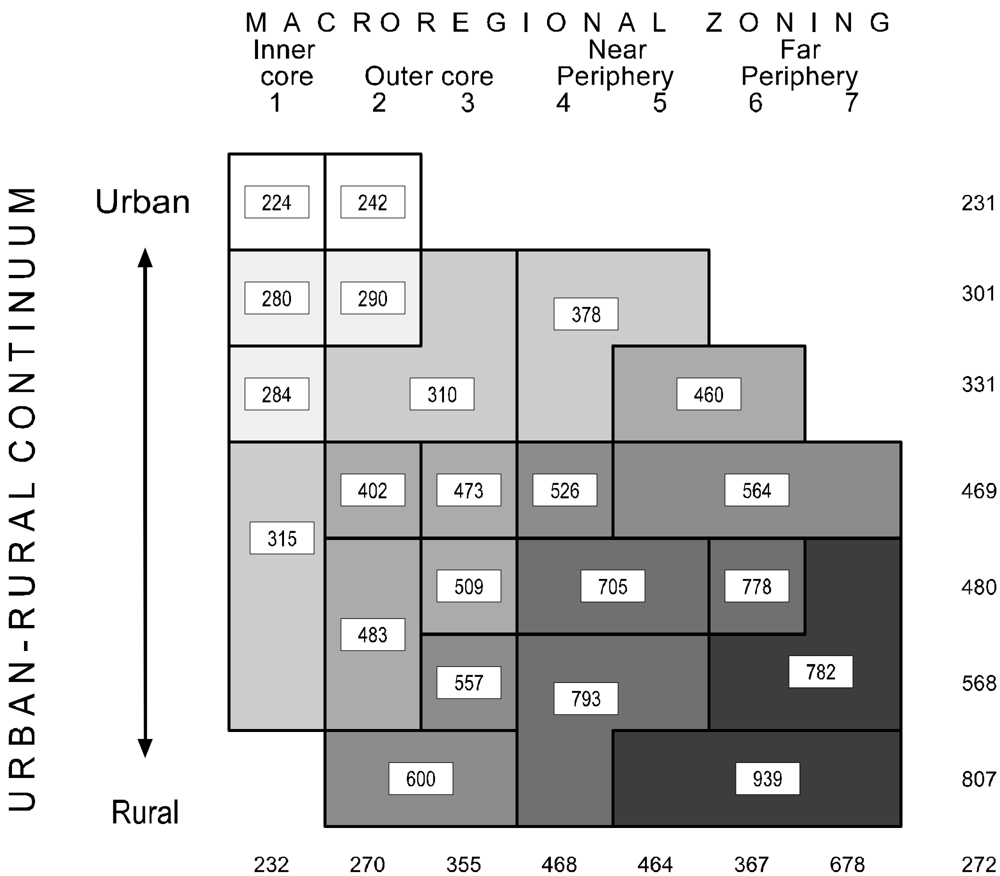
The assignment of counties to core-periphery zones and the construction of the urban-rural continuum index represent the two scales of HRS that can be completed using data for county-level administrative units. Even without the subcounty dimensions of the HRS model (described in the following paper), the first two dimensions (CPZ and URC) may be used fruitfully in the analysis of socioeconomic data collected at the county level—which, in fact, is the level at which most data becomes available to researchers. While researchers should remain aware of the areal aggregation problem, which obscures spatial patterns within county boundaries, county-level data are generally appropriate for depicting the spatial patterns which operate on a macroregional scale.

When socioeconomic data are displayed in a two-dimensional matrix that cross-tabulates core-periphery zones and URC classes, most of them reveal diagonal progressions from the urbanized inner core to the very rural counties in the far periphery. Certain attributes reveal curvilinear and other distinct patterns in HRS, some expected on theoretical grounds, others unexpected that invite explanation. We illustrate this technique with four examples of data from the Lower Yangzi macroregion. The seven core-periphery zones of this region, shaded from light to dark on Map 2, are represented on these matrices as the horizontal axis: from inner core on the left to far periphery on the right. The vertical axis of the matrices shows the urban-rural continuum broken into seven classes, from most urban at the top to most rural at the bottom.

As a first example we apply this technique to the attainment of university education by men and women. Our example shows the sex ratio rising from just over two hundred in the innermost core zones to over nine hundred in the most peripheral cells. That is, while university educated men outnumber university educated women by two to one in the upper left cell, representing Shanghai, they outnumber women by nine to one in the lower right cell, representing the most rural, peripheral counties in the Lower Yangzi

LOWER YANGZI MACROREGION, 1990

Figure 4.1 Sex Ratio* of the University Educated



*Number of men per 100 women

region. We have apportioned the data values on this matrix into seven classes, shaded here from light to dark. Starting from the upper left cell, we see increasing values (and darker shading) as we move diagonally toward the more rural, peripheral cells. As you will notice, some adjacent cells have been merged; the data values shown for these cells are the result of the application of a nonparametric smoothing algorithm.

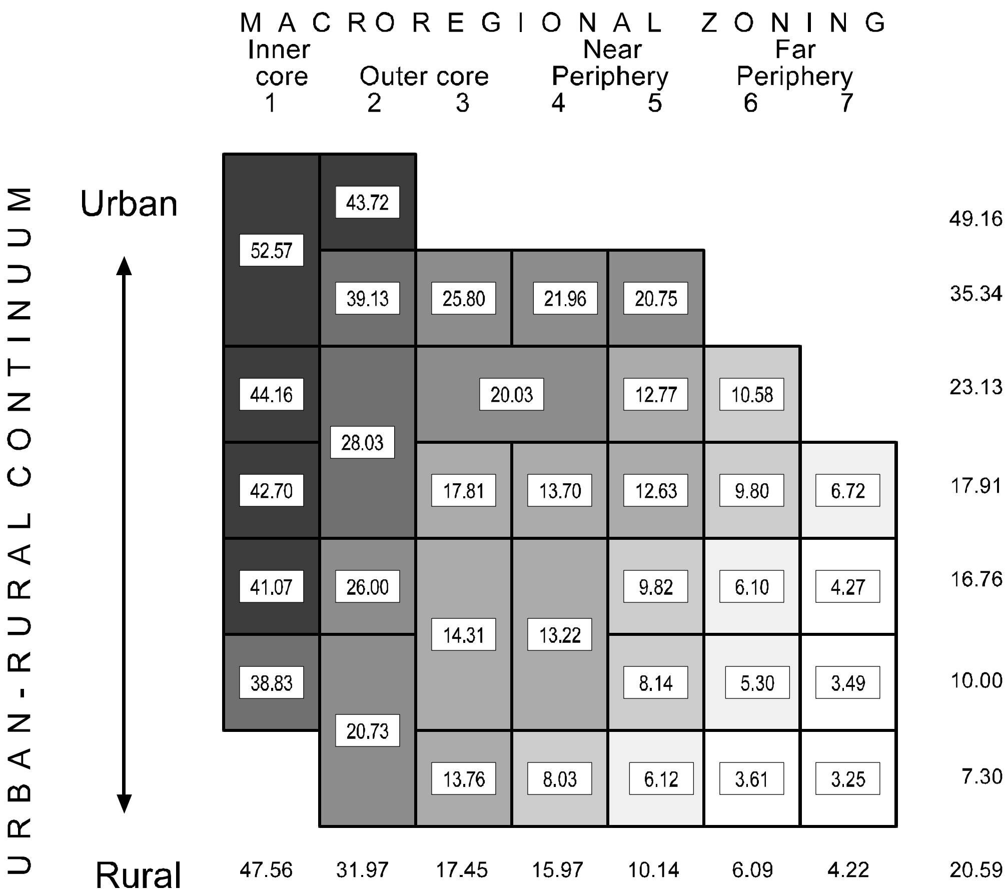
Two more examples portray occupational data. In the manufacturing sector, we show male participation in manufacturing industry decreasing along the diagonal just as the sex ratios of the university educated increased, but also reveals a relatively higher concentration of industrial employment even in less urban settings of the inner core. Another example depicts the percentage of the labor force in agriculture. Here Shanghai shows only 1.6% of the labor force in that sector, while the figure increases rapidly as we move to more rural, peripheral cells.

From occupational data for agriculture we turn to another variable related to the rural economy, rural electricity consumption. As with employment in manufacturing, this variable shows a progression from highest use in the most urban core to lowest use in the most rural, peripheral cells. Here the highest values are all concentrated in the inner core zone, even in the cells representing rather rural counties of that zone. Rural residents of equivalently urban counties in the middle zones are shown to consume from one half down to one sixth of the electricity of their counterparts in the inner core, a finding with interesting repercussions for environmental as well as social science research. (Note that this last example shows no data in the upper left corner, since there are no rural residents represented in our data for Shanghai. Likewise, all of these examples have shown a blank cell at the bottom left, indicating that no inner core counties could be classified as "most rural," and blank cells at the top right, indicating that no peripheral counties were placed in the "most urban" category.)

It is worth repeating that our practice of displaying core-periphery zones and URC classes on orthogonal axes of a data matrix does not carry the implication that our approximations of these dimensions are independent. As should be clear, the socioeconomic variables which we have used in developing these classifications tend to be correlated in space. Analyses of tabular statistical data too often give inadequate

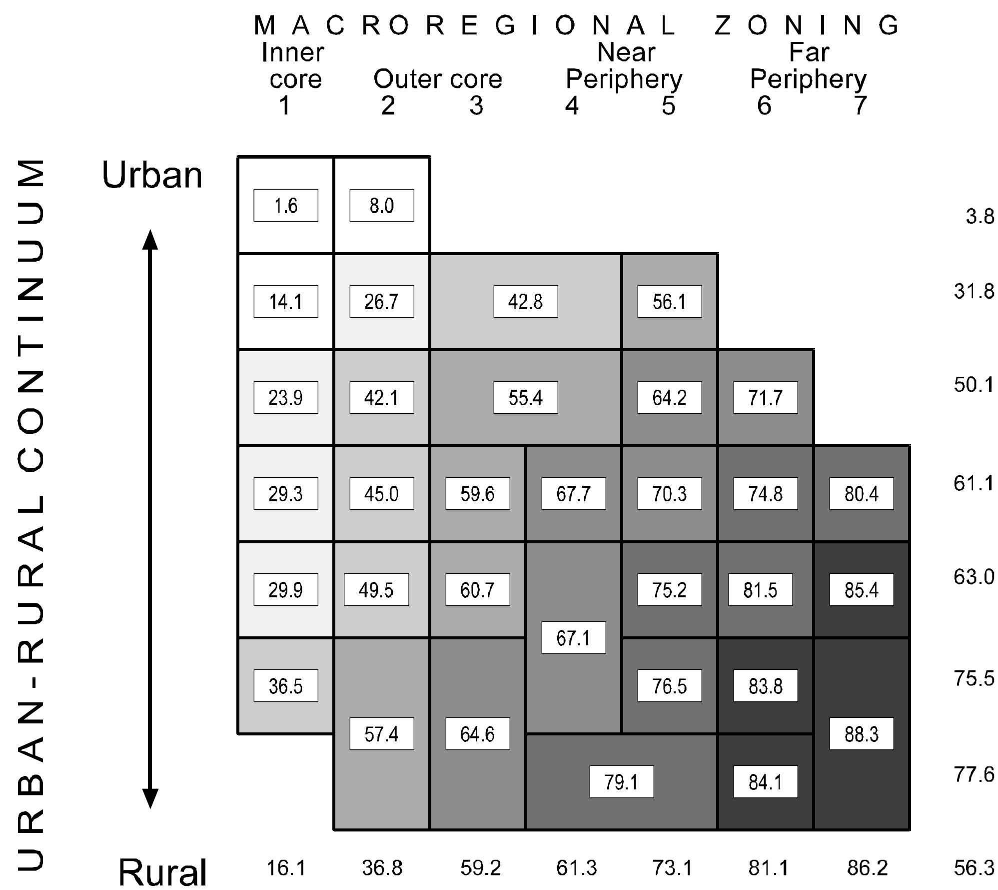
LOWER YANGZI MACROREGION, 1990

Figure 4.2 Males in Manufacturing Industry as Pct. of Male Labor Force



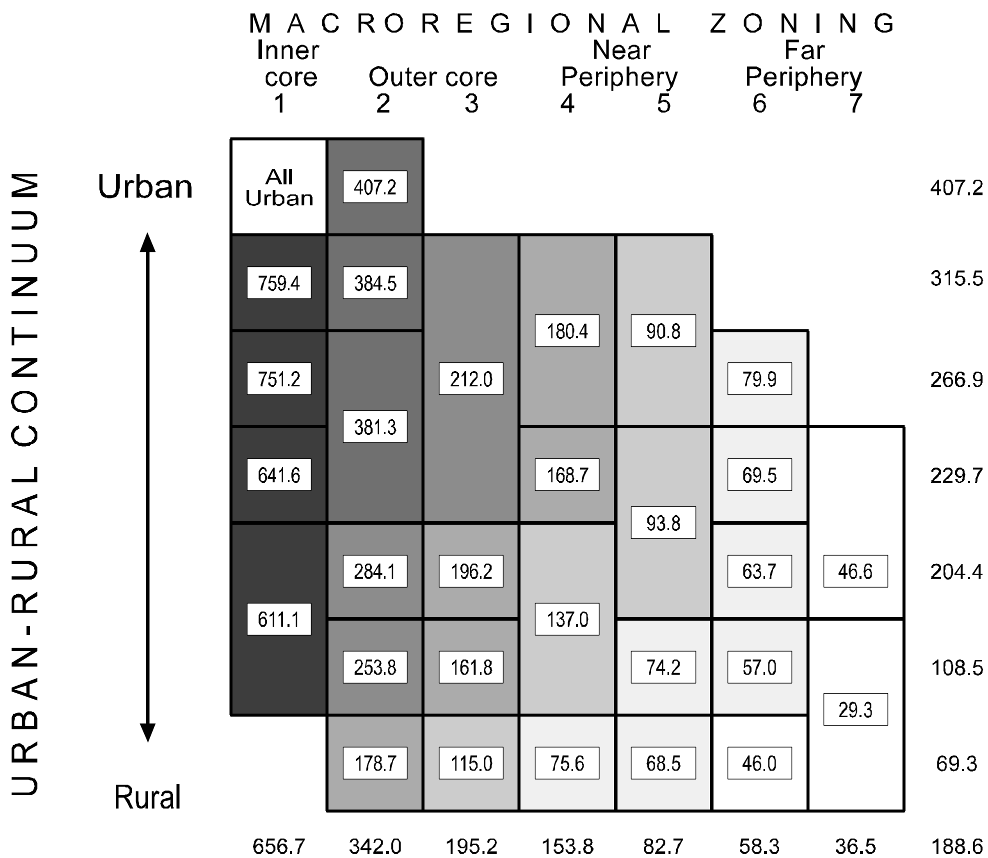
LOWER YANGZI MACROREGION, 1990

Figure 4.3 Pct. of the Labor Force in Agriculture



LOWER YANGZI MACROREGION, 1990

Figure 4.4 Electricity Use, 1000 Watt-Hours per Capita (Rural Population)



attention to such spatial correlations. Our purpose here is to draw attention to the spatial patterns operating on different scales across China. Where two dimensions of the HRS model demonstrate greater discriminatory power than a single dimension, and where three dimensions demonstrate the greatest discriminatory power, our purpose is achieved.

References

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.

Marshall, John, 1981. *The Structure of Urban Systems*. Toronto: University of Toronto Press.

Odland, John, 1988. *Spatial Autocorrelation*. Newbury Park, California: Sage (Scientific Geography Series, 9).

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.

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.