Estimating Components of Population Change from Census Data for Incongruent Spatial/Temporal Units and Attributes

R. G. Cromley
A. Y. Ebenstein
D. M. Hanink

When calculating the components of population change over time, the spatial units of analysis must remain constant. However, the boundaries of these units often change from one census to the next. Another limiting factor is the absence of data values for the time period. Net migration figures might be available for a five year interval in a census but not for a twenty year interval. GIS and areal interpolation are used here to rectify boundary changes that occur from one census to the next and shift-share analysis is used to estimate the components of population change from the census data. These methods are applied to a county level study of population change in China between 1982 and 2000.

Keywords: *Population change, misaligned data, shift-share analysis, census, China.*

R. G. Cromley
Department of Geography
Campus Unit 4148, 215 Glenbrook Rd.
University of Connecticut
Storrs, CT 06269-4148
robert.cromley@uconn.edu

Avraham Y. Ebenstein
Institute for Quantitative Social Science
Harvard University

Dean M. Hanink
Department of Geography
University of Connecticut

INTRODUCTION

Census data are probably the most readily available form of demographic data for many countries but have limitations in the context of temporal studies. Many population studies involve comparisons between the population characteristics of a region for two or more dates as well as the calculation of rates of change in these characteristics over the intervening interval. In order to calculate the components of population change between two dates, however, the spatial units of analysis must remain constant in their geography; otherwise the same regions are not being compared. However, the boundaries of administrative units used to collect population data in a census are often changed during the time interval of interest. This is a common problem in the United States where census reporting units such as tracts are redesigned to reflect the growth of certain areas, and in developing countries such as Nigeria and India where new political units are added over time to recognize different ethnic groups.

A second factor limiting the calculation of population change components is the absence of the same population attributes for both dates, or the absence of a change attribute for the time period. For example, net migration figures may be available for a five year interval but if the time period of analysis were a ten year interval then those migration figures would not be appropriate. Incongruent spatial units of analysis and the associated attributes of these units pose significant problems for the analyst. In this study, GIS and areal interpolation techniques are used to rectify boundary changes over a given time period and shift-share analysis is used to estimate the components of population change. These methods are applied to a case study of population change in China between 1982 and 2000 at the county level.

THE STUDY AREA AND DATA

The distribution of China's people has long been of interest to geographers, and that interest is often intensified by that country's distribution of a new census. The era of the modern census of China began with the one taken in 1953 (Chen, 1973). The total population in 1953 was about 582 million, and for the first time urban and rural figures were reported (Orleans, 1959; Shabad, 1959). Another census was taken in 1964, reported about 705 million people in total, but many figures were not widely released. However, provincial populations were announced near the end of the Cultural Revolution in 1965, which allowed a close estimate of their numbers (Chen, 1973). The Census of 1982 first reported a total population of just over one billion and by the 2000 Census over 1.24 billion people were reported (Harvard Geospatial Library, 2008b). The Census of 1982 was also the first census for which digital databases of county level data beyond basic enumeration were available. Additional county level data became available in the 1990 and 2000 censuses.

The sources of data analyzed in this paper are China's censuses for 1982 and for 2000 as compiled by All China Marketing Research Co., a licensed affiliate of the State Statistical Bureau of China (Wei et al., 2002). The data were developed for use in a geographic information system (GIS) by the University of Michigan's China Data Center in 2005, and made available for the analysis described here by the Harvard Geospatial Library (2008a; 2008b.) Issues of data comparability over time, data accuracy and data reliability are often raised for China's censuses (Riley, 2004). The recent (1982, 1990, 2000) set, however, contains the most reliable counts of the Chinese population ever made (Veeck et al., 2007).

There are three issues of importance with respect to comparing the 1982 and 2000 censuses. One is that the number of reporting units is not equal. Excluding Taiwan, there were 2381 county units in the 1982 census and 2873 in the 2000 census. That discrepancy, of course, has affected comparison of China's population for some time, and was noted in two of the early studies described above (Roxby, 1925; Cressey, 1930). A second issue is that the county data file prepared in the GIS format for 1982 contains only seventeen variables of enumeration, and several of those are not consistent with definitions used in the 2000 Census. The third issue is the actual count in the 2000 census. The 1982 Census is considered to be especially accurate, given the care that was employed in its conduct (Ma, 1983; Donnithorne, 1984; Chengrui, 1985). One of the reasons cited for the quality of the 1982 data is that the location of the population was strongly associated with legal residence under the Hukou system of household registration. While there had been some discrepancy between de facto and de jure populations even before 1982 (Chen, 1973), the discrepancy was quite limited until economic reforms became more effective in the later 1980s. Because of a significant population of floating (non-hukou) migrants by 2000, that Census' geography would have been fairly inaccurate if registration was used in the count. Instead, the 2000 Census was designed to ensure that the actual population was counted, regardless of registration status, at place of residence

(Lavely, 2001; Chan, 2003; Chan and Wang, 2008). This final issue pertains more to the situation in China whereas the first two issues are more general in nature and the focus here.

DECOMPOSING POPULATION CHANGE

The components of absolute population change from an accounting perspective are defined as:

$$P_T - P_0 = B_{0T} - D_{0T} + IM_{0T} - OM_{0T}; \quad (1)$$

where, P_T is the population at a later time T, P_0 is the population in the initial time 0, B_{0T} is the number of births that occurred during the time period $0T$, D_{0T} is the number of deaths that occurred during the time period, IM_{0T} is the amount of in-migration that occurred during the time period, and OM_{0T} is the amount of out-migration from a region during the time period.

Births and deaths in total form the natural increase component of population change, and in-migration and out-migration in total form the net-migration component of this change. Census data are most useful in calculating overall regional population change but not the components of that change. Migration information is fairly limited in census data. In the United States, a sample of the population is asked to list where their residential location was five years previous to the date of the census, but this information would be of little utility in a study with a time interval greater than five years. Net-migration could be estimated if the natural increase component of the change was available but that information normally comes from vital statistic registries rather than from a census.

Because of the limitations of census data with respect to natural increase and net-migration information, an alternative approach to decomposing population change is proposed here based on shift/share analysis (Perloff et al., 1960), a technique normally used to decompose changes in employment for a set of regions. Although shift/share analysis was developed primarily to examine the components of employment change by industrial sectors, the technique has been used to study changes in fertility (Franklin and Plane, 2004) and migration (Ishikawa, 1992). This technique separates growth or decline unique to a region from growth and decline associated with overall national trends in total growth, and growth by segments of the population. In an employment analysis the segments of the population would correspond to different industries, whereas in this demographic analysis the segments correspond to different age-cohorts.

Absolute population change in region r (P_{rc}) is the difference between the population in region r at time T (P_{rT}) and at time zero (P_{r0}). The total population at each time is also the sum of the population in region r in each of i age-cohorts ($\Sigma_i P_{riT}$ and $\Sigma_i P_{ri0}$). Absolute population change for region r is:

$$P_{rc} = \Sigma_i P_{riT} - \Sigma_i P_{ri0}. \quad (2)$$

This can be decomposed into three components by adding arithmetic identities based on the national growth rate of each age-cohort (P_{iT}/P_{i0}) and the nation as a whole (P_T/P_0):

$$P_{rc} = \Sigma_i P_{riT} - \Sigma_i P_{ri0} (P_{iT}/P_{i0}) + \Sigma_i P_{ri0} (P_{iT}/P_{i0}) - \Sigma_i P_{ri0} (P_T/P_0) + \Sigma_i P_{ri0} (P_T/P_0) - \Sigma_i P_{ri0}, \text{ or} \quad (3)$$

$$P_{rc} = \Sigma_i (P_{riT} - P_{ri0} (P_{iT}/P_{i0})) + \Sigma_i (P_{ri0} ((P_{iT}/P_{i0}) - (P_T/P_0)) + \Sigma_i (P_{ri0} (P_T/P_0) - P_{ri0}). \quad (4)$$

The term $P_{ri0} (P_{iT}/P_{i0})$ is the size the ith age-cohort in region r at time T would have been if it had grown at the national rate for that cohort and $P_{ri0} (P_T/P_0)$ is the size the ith age-cohort in region r at time T would have been if it had grown at the overall national rate of the total population.

In this decomposition, the regional shift effect (R_r) is:

$$R_r = \Sigma_i (P_{riT} - P_{ri0} (P_{iT}/P_{i0})), \quad (5)$$

the mix or composition shift effect (M_r) is:

$$M_r = \Sigma_i (P_{ri0} ((P_{iT}/P_{i0}) - (P_T/P_0)) \quad (6)$$

and, the regional share of national growth (N_r) is:

$$N_r = \Sigma_i (P_{ri0} (P_T/P_0) - P_{ri0}). \quad (7)$$

The regional shift component is often interpreted as growth (or decline) due to the relative competitiveness (or non-competitiveness) of a region rather than change associated with national trends in sectoral growth. In the context of population change, this component is interpreted as the place-specific growth due either to net migration (a response to relative regional economic expansion or contraction) or natural increase beyond the national trends. The composition effect is interpreted as growth (or decline) due to the mix of industries in a region. If the industrial profile is dominated by industries whose employment is growing slower than the national average for all industries, the composition effect would result in a decline and if the profile is dominated by industries growing faster than the national average, this effect would result in net growth. In the context of population change, it shows the influence of the age structure in population growth. The third component is the expected share that a region has of the overall national growth or decline based on each industry growing at the total national rate. In the population context, each age-cohort is assumed to grow at the same national rate.

Each component is converted into its share of percent population change by dividing the component value by the initial regional population, P_{r0}. Converting to a rate of change, the percent change for the regional share of national growth for each unit will equal the national growth rate. As a rate, this component is constant across all regions. However, if the regional share rate and the composition effect rate are combined, then this new variable, called here the national share rate (*NS*), represents the impact of age structure on overall regional growth:

$$NS_r = \Sigma_i (P_{ri0} ((P_{iT}/P_{i0}) - 1))/ P_{r0} \quad (8)$$

The national share rate is equivalent to an indirect standardization of the regional rates. In an indirect standardization, age-specific rates for the standard population (in this case the nation as a whole) are applied to the age distribution of the region in question. If a region in the initial time period had the same age distribution as the nation as a whole, then its national share rate would equal the national growth rate. If its age structure were skewed toward slower growth cohorts, then its national share rate would be less than the national growth rate.

Each component itself can be further decomposed by age-cohort permitting a unit's total regional shift rate and national share rate to be associated with the contribution to growth made by each age-cohort. Given the limitations of the disaggregation by age for China's 1982 census, the age distribution will be limited to three age-cohorts: under fifteen years of age, fifteen to sixty-five years of age, and sixty-five and older.

CONSTRUCTING CONGRUENT SPATIAL/TEMPORAL UNITS

The problem of incongruent spatial units is endemic to spatial analysis whenever individual objects are aggregated into areal units for further analysis. Aggregation occurs for several reasons. First, when data are collected for individuals, the reporting of that data by geographical aggregation units avoids the problem of disclosing confidential information as is necessary when conducting a census. Second, there are computational advantages in the analysis of data in an aggregated form. Aggregation greatly reduces, for example, the number of observations associated with any analytical technique thus relieving the computational burden. Also, geographers traditionally study associations among attributes in regions for which zonal approximations already exist.

Because census reporting zones are not fixed and unique, individual data observations are often aggregated into different units, and many alternative geographies are possible resulting in the modifiable areal unit problem (MAUP) (Openshaw and Taylor, 1981). Data values must be then estimated for a common geography from the available information. Areal interpolation is the process for transferring attribute values from one partitioning of space to a different one (Goodchild and Lam, 1980; Lam, 1983;

Flowerdew and Green, 1989). Given two known geographies, one geography is referred to as the source zones and the other as the target zones. Normally, an attribute exists for the source zones but not for the target zones. The attribute values in the source zones are then used to estimate values for the same attribute with respect to the target zones. The subsequent spatial analysis is performed using the target zones. In the context of population change, the attribute of population is known for both temporal geographies so neither can be considered the target one. Instead both are source zones and the attribute of population is estimated for a set of target zones that is a mixture of the two temporal geographies.

Geometrically, areal interpolation requires the polygon overlay of source and target geographies using the union operation of GIS. The overlay of two geographies disaggregates the original areal units of each map layer such that the new map layer contains a new set of areal objects with a spatial resolution different from before. The attribute values are first estimated for this new layer of spatial units and these data values are re-aggregated to the spatial units of the target layer. For the temporal case, a polygon overlay of both source geographies is first conducted. To reduce the level of error associated with the interpolation process, the re-aggregation into target zones for population change should be chosen judiciously. Whenever, possible the spatial units of the target layer should be based on a nesting of units from both source layers. For example, the units in regions that are growing rapidly are often subdivided smaller units for the subsequent time period. Conversely, units in regions that face population decline are sometimes aggregated into larger units in the next time period. In both cases aggregating the smaller units into their larger counterpart eliminates estimation error for those population counts. Estimation only needs to occur whenever the units are not spatially nested over the two time counts.

Wang (2001) informally described a process such as this used to convert county population data for the major plains of China between 1982 and 1990. For estimating population change in China between 1982 and 2000, the 2381 county units in the 1982 census and 2873 counties in the 2000 census were used as the two source geographies. First, all population values for each year that were expressed as spatially intensive ratios or percentages were converted into spatially extensive counts. For example, in 1982 the only population age categories were a) under the age of 15 and b) 65 and older. The values for both groups were given as percents. The count for each age group was found by multiplying each percent by the total population for the county; the count for the remaining age group, 15 to 64, was then calculated by subtracting the sum of the other two age groups from the total population. For the 2000 census, population age categories were given as counts for five year age intervals by sex. Those counts were aggregated by sex and age into the same three age categories as the 1982 data.

Once the population values for each census were in the same categories and units, the different county geographies for the two years were overlaid in ARCGIS 9.2 using the union operation. There are 2071 counties which did not have any boundary changes during the time period. Any counties in 1982 that were completely contained within and congruent to a 2000 county were dissolved in ARCGIS 9.2 into

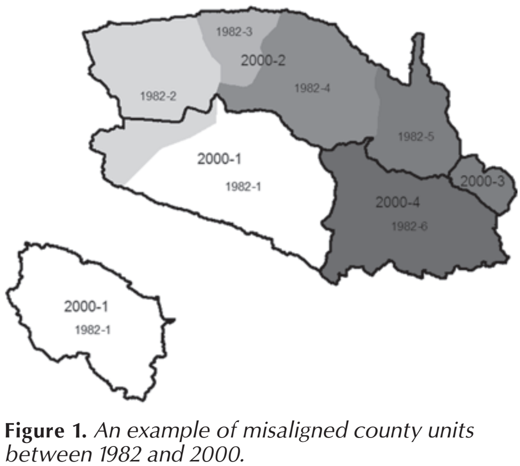

Figure 1. *An example of misaligned county units between 1982 and 2000.*

the same areal unit as the 2000 county and the 1982 population counts were aggregated to that unit. Similarly, any counties in 2000 that were completely contained within and congruent to a 1982 county were dissolved into the same area unit as the 1982 county and the 2000 population counts were aggregated to the 1982 areal unit. In these two conversions, no estimation error occurred. An actual areal interpolation estimate is only needed for counties that were not completely congruent.

As discussed above, the intersection of incongruent counties subdivides the original units into a finer set of polygons. Figure 1 displays the intersection of six counties from 1982 (labeled 1982-1 to 1982-6) and four counties from 2000 (labeled 2000-1 to 2000-4). The 1982 counties are highlighted by different grey tones and the 2000 countries are delimited by black boundaries. Counties 2000-1 and 1982-1 are fragmented counties in each census and 1982-1 lies completely within County 2000-1. Counties 1982-6 and 2000-4 are completely congruent with each other and no estimation is needed. County 1982-5 has been subdivided such that part of it is now County 2000-3 and the remainder is part of County 2000-2. County 1982-2 has also been subdivided and part if it is in County 2000-1 and part of it is in County2000-2. Counties 1982-3

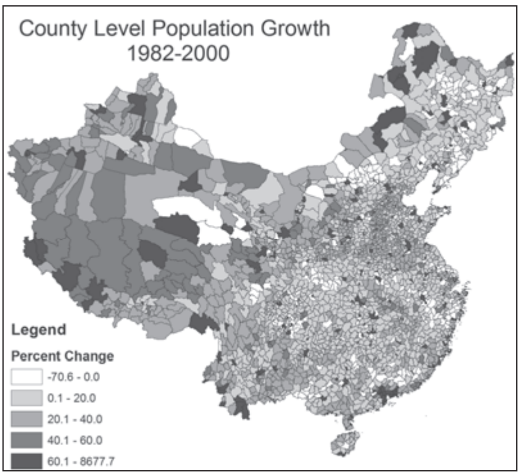

Figure 2. *China's county level population growth, 1982-2000.*

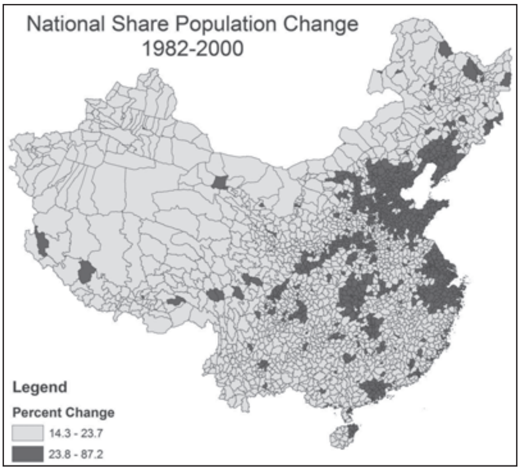

Figure 3. *The national share component of population change, 1982-2000.*

and 1982-4 are completely contained within County 2000-4. The only misalignment is the partitioning of Counties 1982-2 and 1982-5. Therefore, population estimates are necessary only for the four intersection zones associated with these two counties. Once the four unknown values are estimated, the population of County 1982-2 that lies within County 2000-1 is added to the population of County 1982-1 to form the 1982 population for County 2000-1. The populations of Counties 1982-2 and 1982-5 that lies within County 2000-2 are added to the populations of Counties 1982-3 and 1982-4 to form the 1982 population for County 2000-2. Finally the population of County 1982-5 that lies within County 2000-3 forms the 1982 population estimate for that county.

Numerous areal interpolation procedures have been developed for estimating these unknown values. All areal interpolation procedures should be pycnophylatic or volume-preserving (Tobler, 1979) in that the total estimated populations of a set of intersection zones should equal the population of the source zone that they lie within. One of the easiest procedures in terms of data requirements and calculations is areal weighting (Goodchild and Lam, 1980). Interpolated values are a weighted average of source zone values in which the weights are the proportion of the source

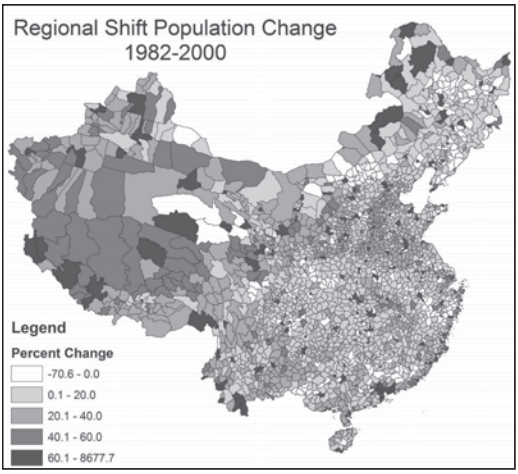

Figure 4. *The regional shift component of population change, 1982-2000.*

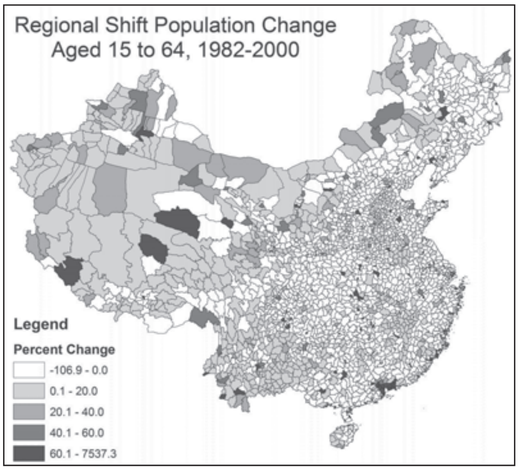

Figure 6. *The regional shift component of population change for the 15-64 age-cohort, 1982-2000.*

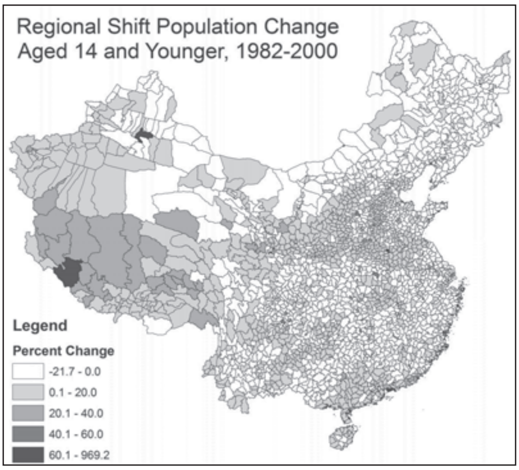

Figure 5. *The regional shift component of population change for the under 15 age-cohort, 1982-2000.*

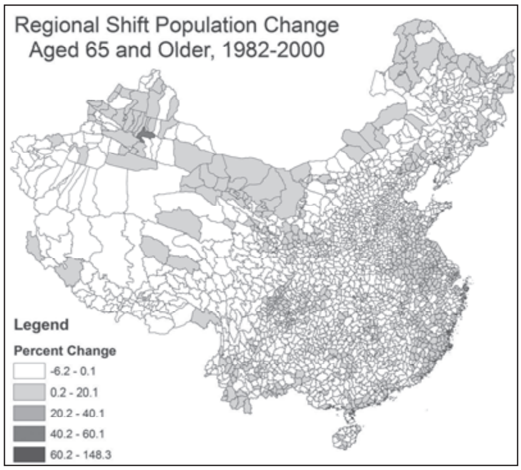

Figure 7. *The regional shift component of population change for the 65 and older age-cohort, 1982-2000.*

zone's area associated with the intersection zone. This technique makes the assumption that the population is uniformly distributed within each source zone. Dasymetric methods (see Eicher and Brewer, 2001; Fisher and Langford, 1996; Reibel and Agrawal, 2007) assume that population is more or less densely distributed according to land-use classification or some other land attribute. These methods use ancillary data such as remotely-sensed land use categories or road networks to assign population only to the areas that are more likely to contain population. More sophisticated areal interpolation techniques based on various forms of statistical analysis have also been developed (see Bracken and Martin, 1989; Flowerdew and Green, 1989; 1991; Flowerdew et al., 1994; Goodchild et al., 1993; Green, 1990; Mugglin et al., 1999); however, most test comparisons have shown that the simpler dasymetric approach (see Cockings et al., 1997; Fisher and Langford, 1995) usually produces more accurate estimates.

The dasymetric approach requires, however, that ancillary data such as land cover is available for the time periods under investigation (in our case for 1982 and 2000). Secondly, for a national level analysis, such as conducted here for all of China, such data would be voluminous. For these reasons, the areal weighting interpolation procedure is used here to proportion the population count of the involved 1982 county by the share of its area contained in the 2000 county of which it was a part. In the end, 2355 county equivalents were used for calculating population change.

CHINA'S POPULATION CHANGE (1982-2000)

Population change on a natural basis was slowed in China between 1982 and 2000, but regional change was significant as the result of internal migration (Riley, 2004). Internal migration occurred but was strongly restricted before the mid-1980s (Liang and White, 1996; Pannell, 2003). By then it was divided into two types: hukou migration occurred with permission from the authorities, and included a change in residential registration, non-hukou migration occurred without permission (Chan and Yang, 1999). The number of non-hukou, or floating, migrants increased by the late 1980s and accelerated in the 1990s as market reforms increased productivity in agriculture. That increased productivity had the migratory push effect of making many agricultural workers redundant, and the migratory pull effect of making food available outside the hukou rationing system in the growing manufacturing centres (Fang and Dewen, 2003). Migration data were collected in both the 1990 and 2000 censuses and the importance of internal movement to China's provincial population geography has been discussed by Fan (2005). The 2000 census even distinguished between hukou and non-hukou migrants, but only concerned migration within the preceding five years (Liang and Ma, 2004).

Overall, China grew from just over one billion in 1982 to approximately 1.24 billion in 2000. Two hundred and forty-six of the 2355 county equivalents experienced either zero population growth or decline (Figure 2). The greatest concentrations of these areas were in the older industrial regions of Northeast and Inner Mongolia but also in counties rimming the port areas of Shanghai and Guangdong. The higher growth rates were mostly in the west in the Tibetan and Xinjiang autonomous regions. The highest growth rates, however, were in the coastal region of Guangdong Province reflecting its growth as the major industrial center.

Among the three age groups, the under 15 group declined in size by 15.6 percent, the 15 to 64 cohort grew by 40.1 percent, and the 65 and older population grew by 79.2 percent. If each county equivalent in 1982 had the same age distribution as China as a whole, its national share growth rate would be 23.7 percent. Those regions with a higher-than-expected national share in the shift/share analysis were mainly located in the eastern part of the country (Figure 3). The distribution mirrors that of the population aged 0-14 in 1982. Those areas with a smaller share of children in 1982 had a higher national share growth rate from 1982 to 2000 given that the growth of this age cohort was below the national average.

The regional shift pattern (Figure 4) was very similar to the pattern of total growth rates because the national trends were a small component of the overall total. The regional shift differs somewhat by age cohorts. An important component of the regional shift for Tibet is associated with the under 15 age group because of the high fertility rates in this region (Figure 5). In general, the non-Han, western portion of the country had higher regional shifts for this age group because of fewer restrictions on family size. The regional shift for the sixty-five and older age group was the most uniform across the country (Figure 7). For this age group, 2349 of the 2355 county equivalents had a change between -21.7 percent and 20.1 percent.

For the 15-64 age group, 1427 of the county equivalents had a negative growth rate (Figure 6). This is most likely associated with the migration of this worker cohort from rural areas to urban/industrial counties. It is this component that accounts for the greatest share of the extraordinary growth of the Guangdong core region. Approximately 87 percent of the total population growth during this period is associated with the regional shift of the 15-64 age group. The migration of this cohort to the Guangdong core is geographically associated with a negative regional shift in the surrounding counties for the same age cohort, resulting in the very low dependency ratios in the Guangdong area and the much higher dependency ratios of the surrounding outer ring of counties surrounding it.

SUMMARY AND CONCLUSION

This paper has outlined a method using GIS operations and shift-share analysis for decomposing population change for incongruent units based on census data. Because census data normally do not specifically contain the natural increase component of population change and the migration information may not conform to the time interval of inquiry, it is difficult to decompose the population change into the standard components of natural increase and net migration. Shift-share analysis does permit a geographical analysis of population change by separating the place-specific aspects of change from broader national trends. Although the empirical focus here has been on China, the methods can be applied to the national census of any country. For some countries like the United States, however, unit boundary changes are very infrequent at the county level (or its equivalent) or above, eliminating the need for a spatial readjustment (although the shift-share analysis could still be applied). For most countries though, boundary changes do occur at the finer resolution units between censuses. Usually a national level analysis is not conducted at these scales and the investigations are focused at a more local level. In this case, more sophisticated areal interpolation techniques may be more appropriate than the areal weighting method used here. As previously discussed, at a regional or local level additional ancillary data such as remotely sensed imagery could be used to refine the population estimates.

For China, the components underline the importance of the One Child Policy in China's changing demographic structure and in its population geography as well as the impact of its economic restructuring since 1982. China's birth rate slowed dramatically between 1982 and 2000 while its death rate remained essentially stable. The regional pattern was not uniform, however, the natural rate of population growth remained much higher in peripheral areas than in eastern China, especially along the coast. The coastal area's population grew, however, and remarkably so in Guangdong, as the result of internal migration to new economic opportunities. These patterns were captured especially in the regional shift of the fifteen to sixty-four age-group in China. Chinese population policy will also need to take into account the impacts of these regional shifts. The peripheral counties surrounding the urban growth centers will have increasing dependency problems unless local economic opportunities are improved.

REFERENCES

Bracken, I., and Martin, D. (1989) The generation of spatial population distributions from census centroid data. *Environment and Planning A,* vol. 21, pp. 537-543.

Chan, K. (2003) Chinese Census 2000: New opportunities and challenges. *The China Review,* vol. 3, no. 2, pp. 1-12.

Chan, K. and Wang, M. (2008) Remapping China's regional inequalities, 1990-2006: A new assessment of *de facto* and *de jure* population data. *Eurasian Geography and Economics,* vol. 49, no. 1, pp. 21-56.

Chan, K., Liu, T. and Yang, Y. (1999). *Hukou* and non-*hukou* migrations in China: Comparisons and contrasts. *International Journal of Population Geography,* vol. 5, pp. 425-448.

Chen, C.-S. (1973) Population growth and urbanization in China, 1953-1970. *The*

Geographical Review, vol. 63, no. 1, pp. 55-72.

Chengrui, L. (1985) China's experiences in the quality control of its 1982 population census. *International Statistical Review,* vol. 53, no. 2, pp. 109-121.

Cockings, S., Fisher, P. and Langford, M. (1997) Parameterization and visualization of the errors in areal interpolation. *Geographical Analysis*, vol. 29, no. 4, pp. 314-328.

Cressey, G. (1930) The new map of China. *The Geographical Review,* vol. 20, no. 4, pp. 652-656.

Donnithorne, A. (1984) The Chinese census of 1982: Preparations, conduct and evaluation. *Hong Kong Economic Papers,* vol. 15, pp. 79-86.

Eicher, C. and Brewer, C. (2001). Dasymetric mapping and areal interpolation: implementation and evaluation. *Cartography and Geographic Information Science*, vol. 20, no. 2, pp. 125-138.

Fan, C. (2005) Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons. *The Professional Geographer,* vol. 57, no. 2, pp. 295-311.

Fang, C. and Dewen, W. (2003) Migration as marketization: What can we learn from China's 2000 census data? *The China Review,* vol. 3, no. 2, pp. 73-93.

Fisher, P. and Langford, M. (1995) Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A,* vol. 27, pp. 211-224.

Fisher, P., and Langford, M. (1996) Modeling sensitivity to accuracy in classified imagery: a study of areal interpolation by dasymetric mappring. *Professional Geographer*, vol. 48, no. 3, pp. 299-309.

Flowerdew R, and Green M 1989 Statistical methods for inference between incompatible zonal systems. In Goodchild M F and Gopal S (eds) 1989 *Accuracy of Spatial Databases.* London, Taylor and Francis Ltd: 239-248.

Flowerdew, R., and Green, M. (1994) Areal interpolation and types of data. In Fotheringham S and Rogerson P (eds) *Spatial Analysis and GIS.* London, Taylor and Francis Ltd: 121-145.

Flowerdew, R., Green, M., and Kehris, E. (1991) Using areal interpolation methods in geographic information systems. *Papers in Regional Science,* vol. 70, pp. 303-315.

Franklin, R., and Plane, D. (2003) A shift-share method for the analysis of regional fertility change: An application to the decline in childbearing in Italy, 1952-1991. *Geographical Analysis,* vol. 36, no. 1, pp. 1-20.

Goodchild, M. and Lam, N. (1980) Areal interpolation: A variant of the traditional polygon overlay problem. *Geoprocessing,* vol. 1, pp. 297-312.

Goodchild, M., Anselin, L. and Deichmann U. (1993) A framework for the areal interpolation of socioeconomic data. *Environment and Planning A,* vol. 25, pp. 383-397.

Green, M. (1990) Statistical models for areal interpolation. In Harts J, Owens H F L and Scholten H J (eds) *EGIS '90: Proceedings, First European Conference of Geographical Information Systems.* Utrecht, The Netherlands, EGIS Foundation: 392-399.

Harvard Geospatial Library. (2008a) China Historical 1982 County Population Census Data. [http://peters.hul.harvard.edu:8080/HGL/jsp/HGL.jsp].

Harvard Geospatial Library. (2008b) China Historical 2000 County Population Census Data. [http://peters.hul.harvard.edu:8080/HGL/jsp/HGL.jsp].

Ishikawa, Y. (1992) The 1970s migration turnaround in Japan revisited: A shift-share approach. *Papers in Regional Science,* vol. 71, no. 2, pp. 153-173.

Lam, N. (1983) Spatial interpolation methods: a review. *The American Cartographer,* vol. 10, pp. 129-149.

Lavely, W. (2001) First impressions from the 2000 Census of China. *Population and Development Review,* vol. 27, no. 4, pp. 755-769.

Liang, Z. and Ma, Z. (2004) China's floating population: New evidence from the 2000 census. *Population and Development Review,* vol. 30, no. 3, pp. 467-488.

Liang, Z. and White, M. (1996) Internal migration in China, 1950-1988. *Demography,* vol. 33, no. 3, pp. 375-384.

Ma, L. (1983) Preliminary results of the 1982 census in China. *The Geographical Review,* vol. 73, no. 2, pp. 198-210.

Mugglin, A., Carlin, B., Zhu, L., and Conlon, E. (1999) Bayesian areal interpolation, estimation, and smoothing: an inferential approach for geographic information systems. *Environment and Planning A*, vol. 31, pp. 1337-1352.

Openshaw, S. and Taylor, P. (1981) The modifiable areal unit problem. In Wrigley, N. and Bennett, R.J. (eds) 1981 *Quantitative Geography: A British View*. London, Routledge and Kegan Paul Ltd, pp. 60-70.

Orleans, L. (1959). The recent growth of China's urban population. *The Geographical Review,* vol. 49, no. 1, pp. 43-57.

Pannell, C. (2003) China's demographic and urban trends for the 21st century. *Eurasian Geography and Economics,* vol. 44, no. 7, pp. 479-496.

Perloff, H., Dunn, E. Jr., Lampard, E. and Muth, R. (1960). *Regions, Resources, and Economic Growth*. Baltimore: Johns Hopkins University Press.

Reibel, M., and Agrawal, A. (2007) Areal interpolation of population counts using pre-classified land cover data. *Population Policy Research Review,* vol. 26, pp. 619-633.

Riley, N. (2004) China's Population: New trends and challenges. *Population Bulletin,* vol. 59, no. 2, pp. 1-40.

Roxby, P. (1925) The distribution of population in China: Economic and political significance. *The Geographical Review,* vol. 15, no. 1, pp. 1-24.

Shabad, T. (1959) The population of China's cities. *The Geographical Review,* vol. 49, no. 1, pp. 32-42.

Tobler, W. (1979) Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association,* vol. 74, pp. 519-530.

Veeck, G., Pannell, C., Smith, C. and Huang, Y. (2007) *China's Geography: Globalization and the Dynamics of Political, Economic, and Social Change*. Lanham: Rowman & Littlefield.

Wang, F. (2001) Regional density functions and growth patterns in major plains of China, 1982-1990. *Papers in Regional Science* vol. 80, pp. 231-240.

Wei Z., Varela, O. and Hassan, M. (2002) Ownership and performance in Chinese manufacturing industry. *Journal of Multinational Financial Management,* vol. 12, pp. 61-78.