Eyes from Above: Remote Sensing and Virtual Globes in Spatial Demography

Lisa Marie Jordan

Department of Geography and Center for Demography and Population Health,
Florida State University

Florida State University
Department of Geography
Room 323 Bellamy Building
Tallahassee, FL 32306-2190

Phone: 850.644.8386,
Email: ljordan@fsu.edu
http://mailer.fsu.edu/~ljordan

Introduction

In the past few decades, research in spatial demography has revealed important patterns and disparities in fertility, mortality, and migration across space (Kitagawa and Hauser 1973; Rogers and Willekens 1986; Morrill 1993; Potter, Schmertmann et al. 2002; Weeks, Getis et al. 2004). The ability to visualize the place-based differences evident in choropleth maps has helped to attract not only demographers, but policy makers, and importantly, the general public to study both the dynamic processes and persistent clustering of events and characteristics evident in populations. This process is becoming more fine-tuned. Through remote sensing and the use of GPS, Geographic Information Systems (GIS), and in particular, Virtual Globes, we are beginning to redefine, or perhaps complicate, the way we see and understand human populations.

Though we may understand precautionary, Orwellian concerns about global surveillance systems, increasingly detailed population information and characteristics are gathered by government, business, and health providers and compiled into GIS systems that make our lives run more smoothly, from allowing faster emergency response to providing us with better communications technology and more efficient shipping and distribution of products we need. The field of demography has not been exempt from the implications of increasing spatial data collection that may affect how we think about and research population.

This paper explores the intersection of remote sensing and spatial demography by discussing the characteristics of non-gridded and gridded population data, and the differences between the measurement of residential and ambient population. The strengths and weaknesses of each type of measurement are considered as are the ways in which differing data types may be combined. Next, the implications of map making are considered, from images of the earth, to digitally produced population maps, to virtual globes. Both concerns and opportunities related to global GIS systems and virtual globes in demography are discussed below.

The Emergence of Two Types of Population Surveillance: Residential and Ambient

At the present, a small, but increasing, body of research in spatial population distribution is not being done by demographers, but by GIS specialists and scientists in remote sensing and photogrammetric engineering (Dobson, Bright et al. 2000; Bhaduri, Bright et al. 2002). I think it is important for demographers to enter debates about new techniques and measurements in studying population density and urbanization, and this article serves partly as an introduction to the world of GIScience. Of particular concern to demographers is the emergence of another type of population surveillance: one that describes ambient, rather than residential populations (Dobson, Bright et al. 2000; Bhaduri, Bright et al. 2002; Sutton, Elvidge et al. 2003).

Table 1 outlines the key differences between the two types of population data: non-gridded and gridded. The main difference in these types of spatial data is in their representation: described by polygons (vector data type) or described within a grid (raster data type). Both types of data are spatial in that they are referenced to the earth's surface. However, some types of spatial data may tend to be collected in vector form (much socioeconomic data, such as the Census Long Form, American Community Survey, economic data), whereas, other types of spatial data, typically environmental data, are collected in raster format.

Table 1. Non-gridded and Gridded Spatial Population Data

Non-gridded Spatial Population Data	**Gridded Spatial Population Data**
Vector Data: non-gridded spatial population data are represented by political or administrative boundaries (i.e. countries, states, metropolitan areas, or census tracts), which are geometrically shown as polygons. *Comparable Data*: certain types of data are collected for areas described by political and administrative boundaries, in particular, many types of socioeconomic and survey data are collected for these areas	*Raster Data*: gridded spatial population data are represented by a grid, with spherical dimensions of equal scale, such as 1 x 1 arc minute. *Comparable Data*: gridded data that survey the earth's environment are typically compiled from remote sensing sources (earth observation satellites). These data describe vegetation, land-use, and weather patterns (temperature, precipitation, storm systems).

Population research poses an interesting challenge to geographic inquiry. People are mobile, and therefore are not confined to a particular location over any length of time. For many reasons, it is important to estimate how many people either live or are working in a particular area (Bhaduri, Bright et al. 2002). This helps with planning (of all types), disaster and humanitarian relief, and routine business operation. Consequently, two types of population estimation are emerging: residential and ambient population. Table 2 outlines the key differences between these measures.

Table 2. Residential and Ambient Population: Characteristics of the Data and the Research that Employs Them

Residential Population	**Ambient Population**
Characteristics of the Data: -The population is measured at the place of residence -Daily and seasonal shifts in location are not considered because data are collected infrequently, every ten years for the decennial census -Residential population counts are often based on actual data (census), sampling techniques, or forecasts	*Characteristics of the Data*: -Population counts are distributed based on average locations (such as workplaces, stores, on roads, on railways, national parks, along coast lines, etc.) -The importance of residence may be understated -Ambient population estimates, at the present, are *inferential*, relying on residential population counts wherever they are available

Characteristics of the Research: -It is easier to attach socioeconomic and other survey information about individuals to place of residence -Residence is particularly important for studying some populations, such as children and adolescents -Residence is also important for studying neighborhood context and social settings	*Characteristics of the Research*: -Calculating ambient population is important to the creation of real-time estimation of population, particularly when responding to or anticipating emergency situations -Studying ambient population is a popular choice for describing local differences in density and urbanization for data-poor countries

Residential population data are characterized by information, which is collected based on the residence of individuals. These data are then used in spatial demography through the aggregation of individual information to spatial units of analysis, which are typically summarized within administrative or political boundaries. They are also frequently used in hierarchical modeling, where individual and neighborhood characteristics are examined simultaneously. Because geographic analyses rely on large sample sizes, surveys conducted by the Census or information collected through vital statistics is the primary source of geographic information used in demography. An alternate way of examining populations is also emerging. Ambient population captures where people *are* (on average), rather than where they reside. Though we live in homes, we are often at work, on the road, in shops, at school, or on a trip. These two interpretations of population can have very different meanings for research. At the present, residential information is essential to documenting the socioeconomic contexts in which people live and how they differ depending on place. However, in data-poor countries, ambient population can be inferred with the help of remotely sensed information.

Assessing Different Spatial Interpolation Methods

Census data can be combined with remote sensing (RS) data to generate a feasible image of how populations are spatially distributed (Chen 2002). Spatial population distributions may demonstrate residential or ambient population, depending on the interpolation method employed. A number of other studies have demonstrated the spatial interpolation of census data, to a higher resolution or smaller spatial unit of analysis, using case studies and information from remote sensing data (Yuan, Smith et al. 1997; Mesev 1998; Eicher and Brewer 2001; Mennis 2003; Tian, Yue et al. 2005). Table 3 describes three types of interpolation methods (simple, smart, and advanced), and the data sources that have been used to conduct spatial interpolation. Simple interpolation uses on geometric attributes to interpolate population estimates (Bracken and Martin 1989; Martin 1989; Martin and Bracken 1991; Flowerdew and Green 1992; Bracken 1993; Goodchild, Anselin et al. 1993), smart interpolation uses one other data source for interpolation (Monmonier and Schnell 1984; Langford, Maguire et al. 1991; Langford and Unwin 1994), such as land-use, roads, or night-time lights, and advanced interpolation uses more than two external data sources to distribute population counts across space.

Table 3 Spatial Interpolation Types and Data Sources

Type of Interpolation	Source(s) for Interpolation
Simple Interpolation:	
Gridded Population of the World (GPW) Data Set	Population Census Estimates (Tobler 1979; Tobler, Deichmann et al. 1997; Balk and Yetman 2004)
Smart Interpolation:	
Land-use:	IKONOS (Liu, Clarke et al. 2006) LANDSAT (Yuan, Smith et al. 1997; Mennis 2003; Qiu, Woller et al. 2003; Tian, Yue et al. 2005)
Roads:	TIGER (Qiu, Woller et al. 2003), vector interpolation (Reibel and Bufalino 2005)
City Lights:	DMPS-OLS (Elvidge, Baugh et al. 1997)
Advanced Interpolation:	
LandScan:	Land-use, roads, slope, elevation (Dobson, Bright et al. 2000; Bhaduri, Bright et al. 2002)
Surface Modelling of Population Distribution (SMPD):	Land-use, net primary productivity (NPP), elevation, city distribution, and transportation infrastructure distribution (Yue, Wang et al. 2005)
Gridded Rural-Urban Mapping Project (GRUMP)	Human settlement data, night-time lights (DMPS-OLS), defense population databases (Balk, Pozzi et al. 2004)

There are a number of criteria on which to judge interpolation methods: the simplicity of the technique, the quality of the estimates in reference to the Census or population sampling datasets, the usability of the method to interested parties (i.e. planners), the improvement of the technique over non-interpolated datasets, and the documentation or transparency/reproducibility of the method. Table 4 shows a classification of the types of methods, ranked from low (1) to high (5) on their simplicity, quality, usability, improvement, and documentation (SQUID). As displayed in Figure 1 with the radar diagram, of all the classifications, the roads and land-use interpolation methods rank very high for their reproducibility, simplicity, and quality of the estimates.

Road overlays are ranked as a more simple method than simple interpolation methods because they can be used to partition spatial demographic information, while maintaining vector (non-gridded) properties (Reibel and Bufalino 2005). Other schemes, such as the Gridded Rural-Urban Mapping Project (GRUMP) and LandScan are more complex and are difficult to replicate, but offer quality results that are very usable in the form of freely downloadable datasets. The Gridded Population of the World (GPW) dataset is also freely available, and provides forecasts of local population (Balk and Yetman 2004); although the quality may not be as good as LandScan (Sutton, Elvidge et al. 2003) or GRUMP (Balk, Pozzi et al. 2004). An example of gridded data from LandScan is shown in Figure 2.

Table 4 Classification of Spatial Interpolation Methods

Type	Simplicity	Quality	Usability	Improvement	Documentation
Simple Interpolation					
Gridded Population of the World (GPW)	4	2	5	2	4
Smart Interpolation					
Land-use	4	3	4	4	5
Roads	5	4	4	4	5
City Lights	3	3	2	3	3
Advanced Interpolation					
LandScan	2	4	5	4	2
Suface Modelling of Population Distribution (SMPD)	2	4	2	5	2
Gridded Rural-Urban Mapping Project (GRUMP)	2	4	5	5	4

5 – High 4 – Medium-High 3 – Medium 2 – Medium-Low 1 – Low

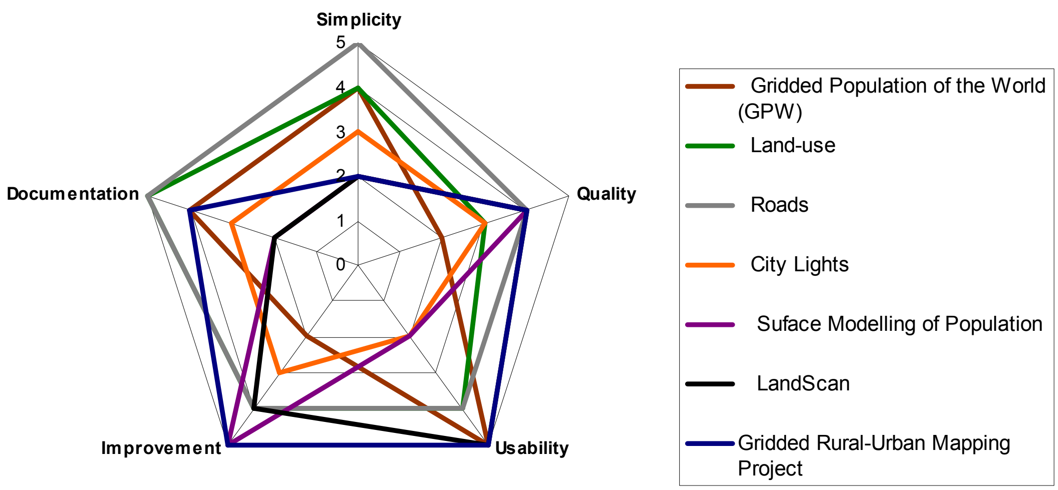

Figure 1 Radar Diagram Classifying Spatial Interpolation Methods

Among the gridded populations sets for the world that are already produced, LandScan is the best product. It is available for recent years, and subject to a number of verification tests for quality. Nevertheless, it is certainly possible to improve LandScan estimates. For example, the use of GRUMP (Balk, Pozzi et al. 2004) or DMSP-OLS data for urban area classifications (Sutton, Elvidge et al. 2003), or to rely only on land-use data for the most elegant and efficient solution, may provide better results in some cases (Tian, Yue et al. 2005). In estimating the population of China, Surface Modelling of Population Density (SMPD) may be the best available technique, particularly when attempting to forecast population values for local areas over a large amount of space (Yue, Wang et al. 2005).

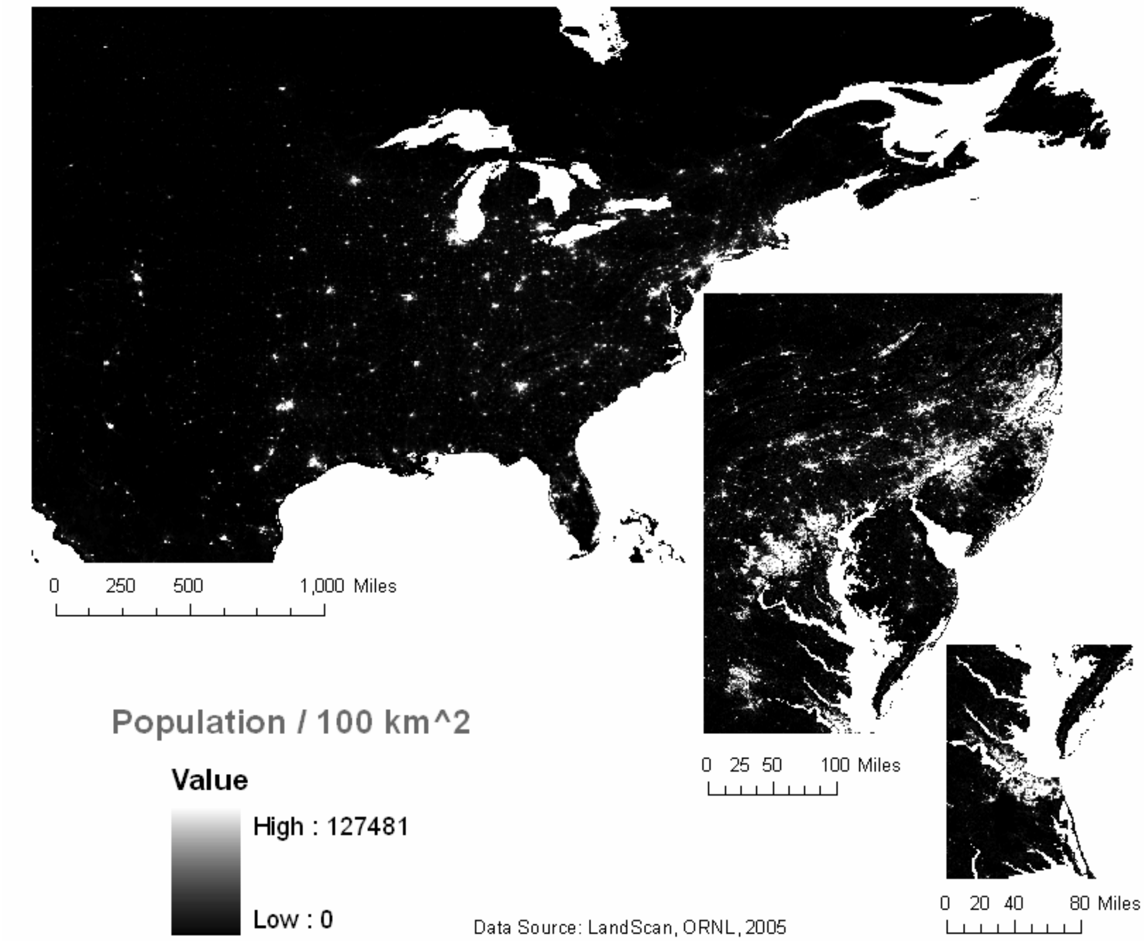

Figure 2 Sample of Gridded Population Data from LandScan, 2005

Future development of real time population estimation should consider the integration of multiple data sources as well as multiple methods of small area estimation to best predict the population for any particular area. For example, recent local surveys and knowledge should be used, whenever they are available, to improve the estimates and the ability to test spatially forecasted data. A central and updated spatial population database is a powerful start, which GRUMP begins to develop (Balk, Pozzi et al. 2004). For the moment, even high-resolution satellite images of built-up areas do not correlate well enough with population data from the census to serve as a lone proxy for population (Liu, Clarke et al. 2006); although, future research is certainly moving in this direction (Harvey 2002). In other words, we continue to rely on Census counts as an important method of validation and verification of small area estimates when creating a national or global dataset.

The Potential of Virtual Globes and Interactive GIS in Spatial Demographic Research

Dasymetric mapping, where a map of population is displayed according to another variable, such as land-use, offers a powerful visualization tool, by distinguishing where human populations are most likely to be. Virtual globes, in particular, the development of Google Earth, make spatial information simple, usable, and exciting for a large audience. The integration of dasymetric mapping and gridded global population data with virtual globe and user-friendly web-interfaces promise to make demographic information accessible to a much wider audience and research community than previously imagined. The updatable nature of web-interfaces are allowing instantaneous transmission of demographic information in order to facilitate disaster response. Two examples are outlined below: hurricane mapping and response and famine early warning systems.

Example 1: Hurricane Mapping and Response

The Florida Resources and Environmental Analysis Center (FREAC) has developed software for the U.S. Army Southern Command that receives calculations on hurricanes trajectories and wind hulls from the FSU Super Computer, estimates the population and agriculture that may be affected by the storm, and sends out automated email alerts with this information (Hodge 2007). This type of system is an essential component of early warning, disaster preparedness and response. Underlying information about the affects of the storm is spatial demographic information.

Example 2: Famine Early Warning Systems

USAID is developing a similar tool for a Famine Early Warning System (FEWS). Prototype software takes LandScan estimates and produces detailed population estimates for any user-selected area, and allows the calculation of populations affected within a certain range of a river or coastline (Watkins, Jordan et al. 2007 (in review)). Estimation techniques that were complex and slow to generate during the Indian Ocean tsunami in 2004, are now becoming simplified and accessible to non-GIS users. Figure 3 illustrates the type of information that can be retrieved to assist in humanitarian response. The polygons that represent administrative boundaries in Kenya are overlayed on LandScan population grids. The user can overlay other environmental information (e.g. river boundaries, or other gridded sources, such as land-use) in order to select the area for which they are interested in retrieving population estimates.

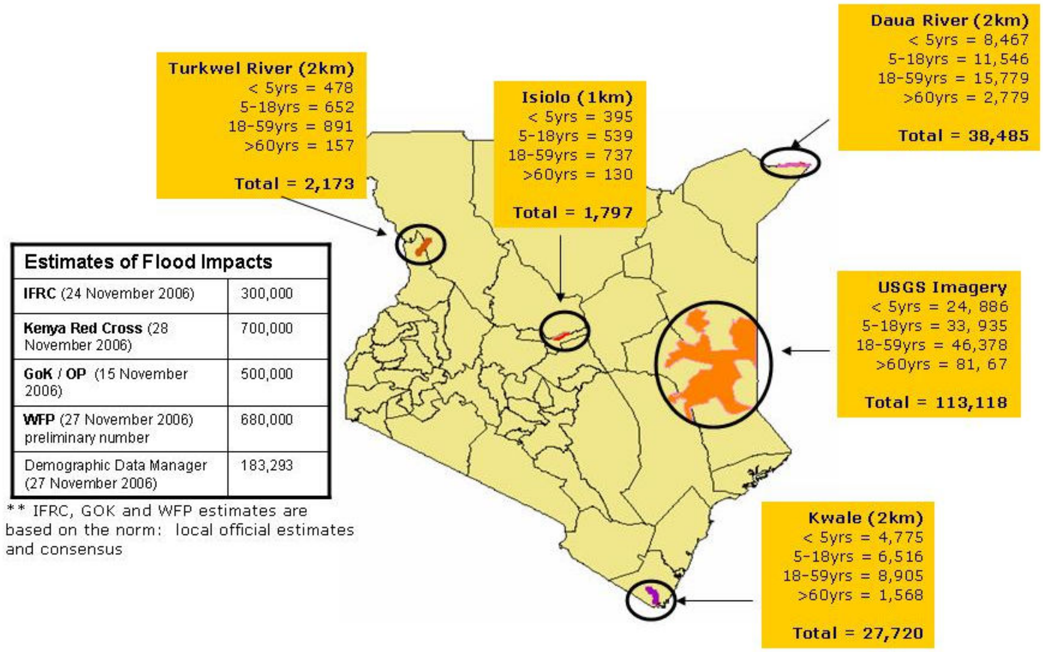

Figure 3 Application of Gridded Population Data to Estimate Flood Affected Populations

Outreach and Improving the Accessibility of Spatial Demographic Data

Virtual globes, online mapping interfaces, and open source (free, open code) GIS software are making both demographic information and remotely sensed images available to a growing audience. The potential for using virtual globes and remote sensing for improvements in disaster response and preparedness, research in public health, and studies concerning the intersection of human behavior and the environment is immense. These highly sophisticated tools are not only for experts, but a much wider community of interested users. The opportunities that these datasets present suggest that concurrent efforts to establish more equity in access to computing and bandwidth, as well as digital learning, are morally significant.

Ethical Concerns and Dimensions of Spatial Population Research

Commiserate with the excitement that virtual globes and remote sensing present to the demographic community are the concerns we should have about privacy of individuals from surveillance. Traveling across Google Earth for the first time inspires both awe, and, perhaps, a sense of concern on where this technology could lead. Layering data to truly bring in the Earth and our environment (both built and natural) into demographic relief are extremely important to demographic research; however, we are at a juncture where it is just as important to describe the depth and complexity associated with the context of place and to appreciate the missing elements from our maps just as much as the ones we include. We cannot claim to have understood human population behavior by taking pictures from space. We can, however, illuminate the intersections of human behavior across the landscape to more clearly "see" our relationships to earth.

Conclusions and Areas for Future Research

Remote sensing and virtual globes in spatial demography are not, or should not, be viewed as independent sources of information, separate from census, sampling, ethnographic, and personal documentation. Virtual globes are platforms that allow us to transverse geographic scales, from the personal to the global, by linking images, stories, videos, within a larger, global community, embedded with different built and natural environments, road and transportation networks, and industries. Such spatial methodologies and techniques add to the array of ways to research, though the pictures we paint can never fully encapsulate or completely represent the very complex and dynamic world in which we live. The more sources of information that spatial demographic research consults; the more rich and useful the datasets will be.

The differences between measuring population as residential or ambient are evidence of the current inability to capture the dynamic and changeable nature of human populations over space and time; although the challenges to small area population estimation are being reduced overall. Gridded population datasets advance spatial population research, through the continued pursuit of adequate interpolation methodologies, and bridging population information with remotely sensed imagery. Gridded datasets can be qualitatively compared with a number of criteria: simplicity of the method, quality of the data, usability of the research, improvements over non-gridded estimates, and documentation and reproducibility of the technique. Gridded population data also promise to link demographic information to a wider audience, through open source GIS and web-based mapping systems. Both opportunity and danger lie ahead in the increased use of GIS, remote sensing, and virtual globes, in relation to privacy and the abuse of information. Demographers, and the academic community at large, need to address concerns about misuse of private information and uses of remote sensing that infringe on civil liberties.

Analysis techniques with gridded population and social datasets are under development. For example, other population based datasets are being transformed to gridded, spatial representations: the Socioeconomic Data and Analysis Center (SEDAC) at Columbia University produced grids of census data, such as the number of people living in poverty, and the foreign born population (CIESIN 2007), and the G-Econ Project at Yale University has generated a gridded dataset of Gross Cell Production (G-Econ 2007). Examining the intersection of environmental characteristics and gridded social information is a logical next step. Improvements in web-based mapping software, open source GIS, and freeware GIS bring the possibility of mapping and analysis to every computer's desktop. The challenge that lies ahead is to understand humans as less static than the populations that appear on the map. Dynamic modeling and better descriptors of change are the next frontier in spatial demographic research.

References:

Balk, D., F. Pozzi, et al. (2004). "The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents." Retrieved 27 Feb, 2007, from http://sedac.ciesin.org/gpw/docs/UR_paper_webdraft1.pdf.

Balk, D. and G. Yetman. (2004). "The Global Distribution of Population: Evaluating the Gains in Resolution Refinement." Retrieved 27 Feb, 2007, from http://sedac.ciesin.org/gpw/docs/gpw3_documentation_final.pdf.

Bhaduri, B., E. Bright, et al. (2002). "LandScan: Locating people is what matters." Geoinformatices: 34-36.

Bracken, I. (1993). "An extensive surface model database for population - related information: concept and application." Environment and Planning B **20**(1): 13-27.

Bracken, I. and D. Martin (1989). "The generation of spatial population distributions from census centroid data." Environment and Planning A **21**(4): 537-543.

Chen, K. (2002). "An approach to linking remotely sensed data and areal census data." International Journal of Remote Sensing **23**(1): 37-48.

CIESIN, C. f. I. E. S. I. N. (2007). "U.S. Census Grids [online data]." Retrieved 28 Feb, 2007, from http://sedac.ciesin.columbia.edu/usgrid/.

Dobson, J. E., E. A. Bright, et al. (2000). "LandScan: A global population database for estimating populations at risk." Photogrammetric Engineering and Remote Sensing **66**(7): 849-857.

Eicher, C. L. and C. A. Brewer (2001). "Dasymetric mapping and areal interpolation: Implementation and evaluation." Cartography and Geographic Information Science **28**(2): 125-38.

Elvidge, C. D., K. E. Baugh, et al. (1997). "Mapping City Lights with Nighttime Data from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing **63**(6): 727-734.

Flowerdew, R. and M. Green (1992). "Developments in areal interpolation methods and GIS." Annals of Regional Science **6**(1): 139-51.

G-Econ. (2007). "Geographically based Economic data (G-Econ)." Retrieved 28 Feb., 2007, from http://gecon.yale.edu/.

Goodchild, M. F., L. Anselin, et al. (1993). "A framework for the areal interpolation of socioeconomic data." Environment and Planning A **25**: 393-97.

Harvey, J. (2002). "Estimating census district populations from satellite imagery: some approaches and limitations." International Journal of Remote Sensing **23**(10): 2071-2095.

Hodge, S. (2007). Florida Resources and Environmental Analysis Center (FREAC): Imaging and Online Mapping. author. Tallahasse.

Kitagawa, E. M. and P. M. Hauser (1973). Differential Mortality in the United States: A Study of Socioeconomic Epidemiology. Cambridge, Harvard UP.

Langford, M., D. J. Maguire, et al. (1991). The areal interpolation problem: Estimating population using remote sensing in a GIS framework. Handling Geographical Information: Methodology and Potential Application. I. Masser and M. Blakemore. London, Longman**:** 55-77.

Langford, M. and D. J. Unwin (1994). "Generating and mapping population density surfaces within a geographical information systems." The Cartographic Journal **31**: 21-26.

Liu, X., K. Clarke, et al. (2006). "Population density and image texture: A comparison study." Photogrammetric Engineering and Remote Sensing **72**(2): 187-196.

Martin, D. (1989). "Mapping population data from zone centroid locations." Transactions of the Institute of British Geographers **14**: 90-97.

Martin, D. and I. Bracken (1991). "Techniques for modelling popoulation-related raster databases." Environment and Planning A **23**: 1069-75.

Mennis, J. (2003). "Generating surface models of population using dasymetric mapping." The Professional Geographer **55**(1): 31-42.

Mesev, V. (1998). "The use of census data in urban image classification." Photogrammetric Engineering and Remote Sensing **64**(5): 431-438.

Monmonier, M. S. and G. A. Schnell (1984). "Land-use and land-cover data and the mapping of population density " The International Yearbook of Cartography **24**: 115-21.

Morrill, R. (1993). "Development, Diversity , and Regional Demographic Variability in the U.S." Annals of the Association of American Geographers **83**(3): 406-433.

Potter, J. E., C. P. Schmertmann, et al. (2002). "Fertility and Development: Evidence from Brazil." Demography **39**(4): 739-361.

Qiu, F., K. Woller, et al. (2003). "Modeling urban population growth from remotely sensed imagery and TIGER GIS road data." Photogrammetric Engineering and Remote Sensing **69**(9): 1031-1042.

Reibel, M. and M. E. Bufalino (2005). "Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems." Environment and Planning A **37**(1): 127-139.

Rogers, A. and F. Willekens (1986). Migration and Settlement: A Multiregional Comparative Study. Dordrecht, Reidel.

Sutton, P., C. Elvidge, et al. (2003). "Building and evaluating models to estimate ambient population density." Photogrammetric Engineering and Remote Sensing **69**(5): 545-553.

Tian, Y., T. Yue, et al. (2005). "Modeling population density using land cover data." Ecological Modelling **189**(1-2): 72-88.

Tobler, W. R. (1979). "Smooth Pycnophylactic Interpolation for Geographical Regions." Journal of the American Statistical Association **74**(367): 519-530.

Tobler, W. R., U. Deichmann, et al. (1997). "World population in a grid of spherical quadrilaterals." International Journal of Population Geography **3**(203-225).

Watkins, B., L. Jordan, et al. (2007 (in review)). "Practice Approaches to Population Estimation for Famine Early Warningn and Food Security Analysis." Disasters.

Weeks, J. R., A. Getis, et al. (2004). "The Fertility Transition in Egypt: Intraurban Patterns in Cairo." Annals of the Association of American Geographers **94**(1): 74-93.

Yuan, Y., R. M. Smith, et al. (1997). "Remodeling Census population with spatial information from Landsat TIM imagery." Computers, Environment, and Urban Systems **21**(3/4): 245-258.

Yue, T., Y. Wang, et al. (2005). "Surface modelling of human population distribution in China." Ecological Modelling **181**(4): 461-478.