BUILDING A GLOBAL DEMOGRAPHIC MAP DATABASE FOR DISASTER RELIEF AND CRISIS MANAGEMENT

Joshua Comenetz
Population Division
U.S. Census Bureau
4600 Silver Hill Road
Washington, DC 20233, USA
Joshua.Comenetz@census.gov

Abstract

Timely provision of maps tied to subnational demographic information is critical for improving the efficiency of humanitarian relief efforts during and after natural disasters, conflicts, and other crises. Gathering maps and data after an event is time-consuming, especially when an event affects multiple countries. Despite the demand, a global database of administrative boundary maps linked to up-to-date subnational population data is not currently available.

In response to this need, the U.S. National Research Council published a report on estimating populations at risk from disasters in 2007. The report recommended that the U.S. Census Bureau increase its capacity to provide global, subnational demographic data and maps. The Census Bureau is currently working to create an international database that will contain demographic data and administrative boundary maps for use in disaster planning and humanitarian relief.

The Populations at Risk initiative consists of three components: assembly of an archive of digital base maps containing subnational administrative boundaries for every country and territory in the world, expansion of the Census Bureau's database of subnational demographic data, and development of methods for enhancing the currency and spatial resolution of census-based maps by use of satellite image processing. The base map archive currently incorporates boundary files in ArcGIS shapefile format that match the geographic structure found in each national census at the lowest available administrative level: first (state/province), second (county/district), third (municipality/township), or occasionally fourth or lower (e.g., village). These maps will be linked to demographic data derived primarily from the census reports produced by the statistical agencies of each country and territory, covering age, sex, and other demographic and housing variables such as housing type, literacy rate, and language spoken. Creation of a global database poses several challenges, among them the different interpretations of national boundaries found on neighboring countries' maps and differing national definitions of important variables such as literacy.

Maps based on satellite images will be used to improve data resolution and update available census results. Image processing methods, similar to those used to produce the U.S. National Land Cover Data of 2001 (a national land cover map and digital database), have been developed

---

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, *or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.*

to map human-created impervious surfaces (hard surfaces such as roads and roofs), leading to maps that show the percentage of built-up area at pixel scale. These raster maps, aggregated to the level of administrative units found in census data, are used to calculate the relationship between built-up area and population density. The results of the preceding calculation make it possible to create estimated population density maps at pixel level, and to project census-derived population densities forward to the present by comparing satellite images obtained at the time of the census and more recently. Image analysis will also aid in evaluating the quality of census data, by revealing regions where published census results differ significantly from image-based estimates of population density.

The Populations at Risk effort draws on the results and products of long-term and ongoing Census Bureau projects related to international population estimation and mapping. These include production of an annual demographic base for the Oak Ridge National Laboratory LandScan global gridded population database, production of the online International Data Base of national-level population estimates, creation of subnational demographic data and maps to assist in charting the international, subnational spread of HIV infection for the U.S. President's Emergency Plan for AIDS Relief, and investigations into the relationship between built-up areas and population distribution in Haiti and Pakistan. The latter studies developed the methods described above for connecting satellite images to census data on population density.

Maps and databases produced through the Populations at Risk initiative will contribute to better planning for natural and humanitarian crises. The ability to provide maps of population distribution and other demographic characteristics, at short notice, for small geographic areas worldwide, can aid in making disaster relief more efficient and cost-effective.

Introduction

The U.S. Census Bureau has a long history of producing annual population estimates by age and sex for every country in the world at the first (province/state) and in some cases second or lower administrative levels. Through the Populations at Risk (Pop@Risk) initiative, the Census Bureau plans to improve access to current international geo-demographic data with the goal of contributing to response to natural disasters and human-induced crises worldwide.

Over the past decade, digital demographic data, boundary maps in GIS formats, and satellite images have become increasingly available. Advances in data accuracy and spatial precision can be achieved by incorporating census data and digital base maps at lower administrative levels, application of improved subnational demographic projection methods, and through use of satellite images to map populated places. Image data can also assist in developing demographic estimates for countries that have no recent census data or that restrict public access to demographic data. This would be of particular value in developing countries where migration and other population changes mean that census results become outdated rapidly. Provided via an internet portal, the Pop@Risk database could be used to improve the efficiency of response to climatic disasters, famines, and other natural or human-induced crises through provision of demographic estimates that are more geographically and temporally accurate than has previously been possible.

Goals

The primary purpose of the international Pop@Risk project is to produce and provide geospatial data and maps that can be used to help save lives during and after natural or human-induced disasters. This Census Bureau effort responds to the National Research Council's (NRC's) call for better data and maps in support of disaster response. A recent NRC report included the following recommendations (NRC 2007, pp. 152-55):

> Develop a template of minimum acceptable population and other geospatial data sets that are required by disaster responders. The data sets should be updated frequently (at least mid-decade if not more frequently) and include digital census enumeration units and other census maps in digital form.
>
> Support should be given to test the accuracy of estimates of size and distribution of populations based on remotely sensed imagery, particularly in rural and urban areas of countries with spatially, demographically, and temporally inadequate census data.
>
> The U.S. Census Bureau should be given greater responsibility for understanding populations at risk and should be funded to…support the U.S. government in international disaster response and humanitarian assistance activities. The U.S. Census Bureau should also have an active research program in using and developing these tools and methods, including remotely sensed imagery and field surveys.

Timely provision of good subnational demographic information can improve the efficiency of humanitarian relief efforts during and after disasters. The Pop@Risk database will contribute to the solution of problems ranging from basic questions—"How many tents should we load on the plane?"—to longer-term issues such as provision of age-appropriate health care to evacuees. Better targeting of services may improve both the quality of response and allocation of scarce financial resources.

Background

For many years, the Population Division of the U.S. Census Bureau has been a source of international demographic estimates at the national and subnational levels. Annual national estimates and projections (data for entire countries) are included in the International Data Base (U.S. Census Bureau 2009). Subnational estimates can be linked to digital maps, allowing for the creation of international demographic maps. Some of these data and maps serve as a base for the LandScan international gridded population database (www.ornl.gov/sci/landscan), created by the Oak Ridge National Laboratory.

Data

The Census Bureau's digital map archive covers every country and territory in the world. Figure 1 shows the smallest administrative unit (ADM) currently held in ArcGIS shapefile format for each country. The goal is to obtain both data and maps at the lowest (smallest) available ADM level. ADM levels are defined as follows:

- ADM0 – Countries and territories.
- ADM1 – States, provinces, or equivalent.
- ADM2 – Districts, counties, or equivalent.
- ADM3 – Municipalities, townships, or equivalent.

- ADM4 and below – Villages or other small units.

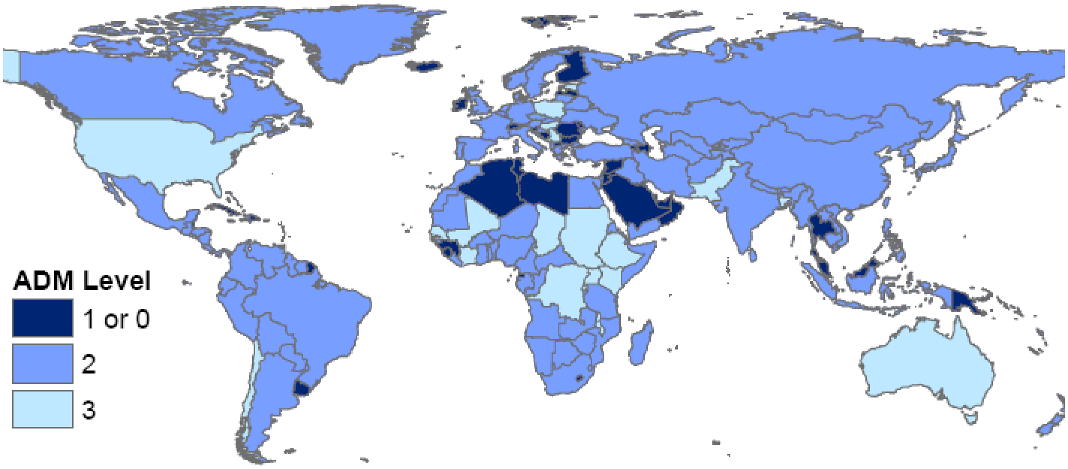

Figure 1. Smallest administrative level held in the U.S. Census Bureau digital map archive. There are no ADM3s in some U.S. states.

Most of the world is covered at the ADM2 level, but even at ADM2 there are gaps in the Middle East/North Africa and other areas. On the other hand, ADM3 maps have been obtained for more than twenty countries. Separate maps are obtained for each ADM level wherever possible. While smaller ADMs can usually be aggregated to form higher-level ADM units, this is not always the case (examples include South Africa and New Zealand). Also, boundary revisions can occur at any time, making it difficult to map change over time (Comenetz and Caviedes 2002). Data sources include the websites of national statistical agencies, international organizations such as the United Nations and World Health Organization, the Department of State and other U.S. government agencies, online collections maintained by academic and other non-governmental organizations, and digitization of paper maps by Census Bureau staff.

The Census Bureau's subnational demographic holdings similarly include population data (total population, and age/sex breakdowns typically in 5- or 10-year bands) for the entire world. In more than 90% of countries or territories, covering over 95% of the world's population, these data are derived from national censuses. Census data are essential to creation and maintenance of a Pop@Risk database because censuses usually cover all of a country's territory, allow for "continuity of statistics from census to census," and provide "detail…about population sub-groups in local areas" (NRC 2002, p. 7). Where national census data are unavailable or outdated, official estimates or other sources are used—for example, in countries affected by war (Afghanistan, Angola) or where no census has been conducted in more than a decade (Andorra, Greenland).

Figure 2 contains excerpts from the existing demographic archive. Columns on the left identify geographic areas down to ADM3. The "Geo Match" column provides the unique code used to join population data with shapefile polygons. Total, male, and female population estimates are each broken down by five-year age bands (sample columns shown; full range is 0-4 years, 5-9, 10-14,…75-79, 80 and over).

FIPS	COUNTRY	ADM1 NAME	ADM2 NAME	ADM3 NAME	GEO MATCH	Name Change/ Alternate Name	CONTI NENT	IDB YEAR	TOTAL	TOT 0-4	TOT 80+
ET	ETHIOPIA						Africa	2007	76511887	12372169	227988
ET48	ETHIOPIA	DIRE DAWA				DIRE DAWA PROVISIONAL ADM1	Africa		360351	58270	1074
ET48	ETHIOPIA	DIRE DAWA	DIRE DAWA		003ETH015001001		Africa		235858	38139	703
ET48	ETHIOPIA	DIRE DAWA	GURGURA		003ETH015001002		Africa		124493	20131	371
ET49	ETHIOPIA	GAMBELA HIZBOCH				GAMBELLA	Africa		260197	42074	775
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1				Africa		64679	10459	193
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1	GAMBELA	003ETH012001004		Africa		37827	6117	113
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 1	ITANG	003ETH012001003		Africa		26852	4342	80
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2				Africa		51791	8375	154
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	ABOBO	*003ETH012002005*		Africa		19992	3233	60
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	GOG	003ETH012002009		Africa		21651	3501	65
ET49	ETHIOPIA	GAMBELA HIZBOCH	ZONE 2	JOR	003ETH012002010		Africa		10148	1641	30

Figure 2. Excerpt from demographic data archive showing 2007 data for parts of Ethiopia.

Starting in 2008, the Census Bureau began to plan for expansion of the data archive to include more variables. Important variables—components of the "minimally essential data set" (NRC 2007, 89-96) for populations at risk projects—vary according to type of disaster and recovery stage but always include total population disaggregated by age and sex. Table 1 outlines other variables that are potentially useful in disaster response and that are actually available for many countries.

Type of Event	Some Key Variables
A. War and genocide	• Baseline populations of administrative areas – total population, age, sex • Cultural variables – ethnicity, language • Household composition for estimating housing needs
B. Disease	• Populations most vulnerable to disease – children, elderly, lower income • Health status, e.g., HIV infection increases vulnerability to other diseases
C. Short-term disasters (e.g., hurricanes, earthquakes, tsunamis)	• Baseline populations of small administrative areas • Vulnerable populations – elderly, lower income, migrants (especially those lacking knowledge of national language) • Housing characteristics – e.g., wall and roof materials, source of drinking water, age of housing
D. Longer-term disasters (e.g., climate variation, sea level rise, famine)	• Population density, urban and rural population • Growth projection by coastal administrative unit • Ethnicity • Occupation • Elevation above sea level (McGranahan *et al.* 2007)

Table 1. Geo-demographic variables of value for a Populations at Risk database.

Data Problems

To provide geo-demographic information that is as accurate and current as possible, both archives—demographic data and digital maps—require revision and updating as new census results, demographic estimates, and administrative boundary maps are released by national governments. The following complicate this process:

- Demographic data and maps are often produced by different agencies or released separately (it is typically more difficult to obtain detailed administrative maps than numerical data). If only one of these two elements is available, it is not possible to *map* population. Also, governments may release data and maps at different ADM levels. For example, ADM2 data cannot be mapped if the best-available shapefile contains only ADM1 boundaries.

- Countries around the world choose to split, combine, or completely redraw administrative areas. For example, Ethiopia (Figure 3) and South Africa redrew provincial and lower-level boundaries in the early 1990s; India and Nigeria periodically create new states; and new counties are occasionally created in the United States.

- Censuses are relatively infrequent—usually a decade or more apart. Most population estimates are based on censuses, and therefore calculated using the ADMs in force at the time that the most recent census was conducted. If boundaries are redrawn after a census, it can be difficult to produce accurate estimates of the populations of new ADMs unless the new ADMs are defined in terms of pre-existing smaller units. For example, if a country redraws ADM1 boundaries but keeps ADM2 boundaries constant, one can calculate the new ADM1 populations; otherwise, this may not be possible.

- Data and maps made available by governments can be withdrawn or revised at any time. Frequent online searches must be conducted, and personal contacts maintained with agencies that produce data, to ensure that all new materials are added to the Census Bureau archive as soon as possible after release or revision.

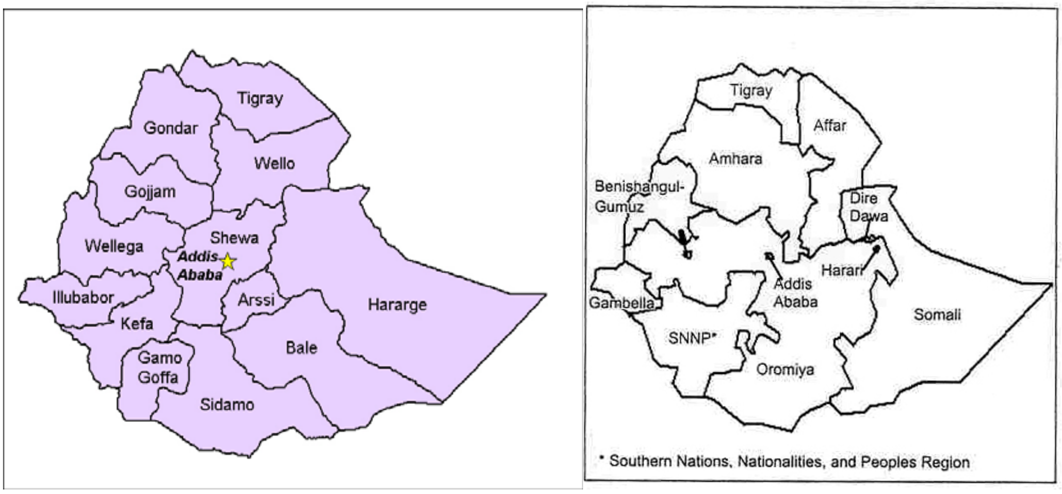

Figure 3. Ethiopian provincial boundaries in effect at time of 1984 census (left) and 1994 census (right). The newer boundaries were drawn along ethnic lines.

The above issues mean that revision and update is an ongoing process. All available data and maps are collected, but data and maps with matching geography frequently must be obtained at different times or from different sources. An online Pop@Risk database of demographic maps will therefore reflect the lowest common denominator in both archives—that is, the smallest and most recently delineated ADMs for which both data and digital boundaries are available.

Remote Sensing Methods

The goals of satellite imagery based population mapping are to:

- Map population distribution at finer scales than is possible with census data (Tatem *et al.* 2007). This is especially useful where countries do not gather or release data for smaller administrative units. Current Census Bureau projects are distributing population to pixels (square areas) of 1.3 hectare or smaller.

- Provide information about the geographic distribution of population where no census has been taken recently, or where census data are no longer reliable. Examples include Somalia, eastern Ethiopia, and Lebanon. Image-based mapping also has the potential to highlight possible quality problems with national censuses by revealing geographic regions where there are large differences between intercensal population change as measured by censuses and by image analysis that looks at variation in the built environment.

- Improve the quality of intercensal population estimates. Relatively few countries take censuses more frequently than once per decade (UN 2009). Image mapping could help validate demographic projections produced by standard mathematical methods by revealing, for example, rapid growth in urban areas. Image mapping is unlikely to be of as much use in areas of population decline, because the built environment will remain; however, this will not pose a serious problem in the medium term because populations are still growing in most countries.

To map population distribution at pixel scale, the Census Bureau is using a combination of demographic data from national censuses and impervious surface (hard land covers such as roads and roofs) maps derived from satellite imagery. The methodology used to classify impervious land cover features was largely adapted from methods developed for the Multi-Resolution Land Characteristics Consortium's National Land Cover Data (Yang *et al*. 2003, Homer *et al.* 2007) by the U.S. Geological Survey's Center for Earth Resources Observation and Science (USGS EROS). USGS EROS used a classification and regression tree approach, which is increasingly common in image classification. In this approach, multiple small training sites derived from high-resolution imagery (e.g., Quickbird or IKONOS, with a pixel size of 4m or less) are used to train a regression tree that is then used to estimate impervious surfaces over larger areas using lower resolution imagery such as Landsat or ASTER (pixel size of 28.5m or 15m, respectively).

Use of a mix of imagery types is an efficient and cost-effective way to map entire countries because it takes advantage of the benefits of each kind of imagery. High-resolution imagery has the advantage that it shows the built environment clearly. It is therefore possible to classify each pixel as impervious or pervious, and to omit non-human impervious surfaces such as bare rock from the "impervious" category. However, each scene (individual image) covers a comparatively small area—meaning that processing enough images for a national map becomes time-consuming—and it is also difficult or impossible to obtain complete coverage for entire countries. Landsat and ASTER imagery provide more complete coverage of the earth and require fewer scenes to create a national map, so processing time is reduced.

The key question is how well image-based maps of impervious surface will correlate with population distribution. Comprehensive testing of the procedures requires:

- Census data, linked to boundary maps of the smallest available ADMs in several countries. The procedure must be tested in regions with different environmental characteristics (e.g., desert, humid tropical, mountainous) and urban planning and building traditions, because varying environments and construction styles are likely to produce different patterns of impervious land cover.

- National impervious surface maps of the same countries, derived from satellite images obtained as near to the census dates as possible.

- Information on the location of urban areas in each country. The population-impervious surface relationship will vary not only among world regions but also within countries, because in poorer countries, for example, rural roads are less likely to be paved than city roads.

With the above in hand for a country, the next step is to aggregate Landsat pixels to the smallest available ADM boundaries. It is then possible to chart the correlation between impervious area and census-derived population at the administrative unit level. The goal is to calibrate image maps using census data, making it possible to estimate population in areas where images but not census data are available. Among the questions to be answered are:

- How well does impervious area correlate with census-derived population counts?

- How does the impervious area-population relationship vary regionally (among major world environmental and cultural regions)?

- Is the impervious area-population relationship consistent within countries, or within urban and rural areas of each country?

- To what extent does the impervious area-population relationship vary over time? Such changes could occur through increasing urban sprawl or changes in building methods.

If there is regional consistency in the impervious area-population relationship, then it will be possible to apply the results from countries with census data to neighbors with none—for example, results from Ethiopia or Kenya could be used to map population in Somalia. Temporal consistency in the impervious area-population relationship will determine whether imagery can be used to map population change between censuses, though it seems unlikely that the relationship could change radically in the space of a decade or two.

Example: Haiti

The methods described above were used in a recently completed study of Haiti. The illustrations in Figure 4 illustrate the process. First, data from the 2003 census were mapped at ADM3 level (the 560 *sections communales*) (Figure 4, upper left, with darker colors indicating greater population). Next, satellite imagery was used to map human-created impervious surface by ADM3 (Figure 4, upper right, with darker colors indicating more impervious area). To obtain pixel-level population estimates, people were only allowed to be assigned to the areas shown in green in Figure 4, lower left: the remaining areas contained steep slopes, water, or were otherwise assumed uninhabitable. The final population distribution map is shown in Figure 4, lower right, with rural or sparsely-populated areas in brown and denser or urban population in orange or yellow.

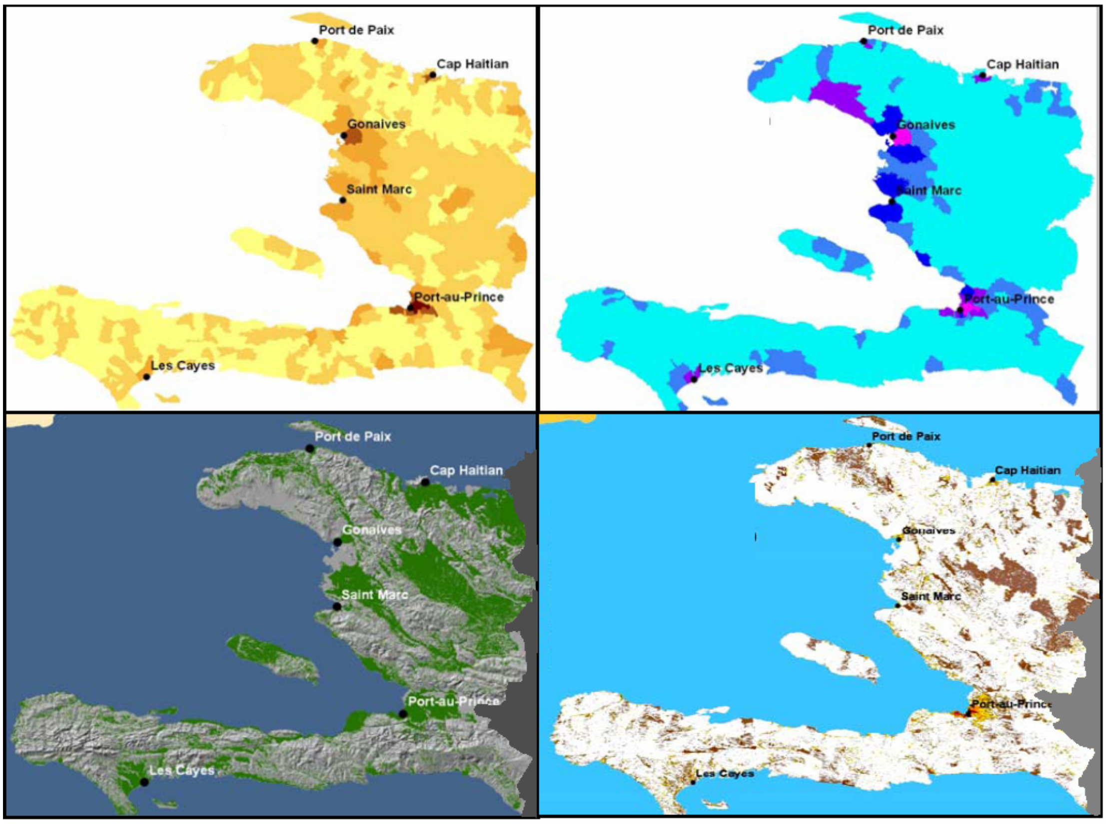

Figure 4. Mapping population in Haiti, based on image processing and census data.

Conclusion: Next steps

As noted above, the remote sensing methods must be tested in several diverse regions in order to refine them for general use. Following completion of the Haiti project, the Census Bureau began the next step in this process: a study of Pakistan, a country with different demographic patterns and physical characteristics that is also much larger in both population and land area.

Acknowledgements

The efforts of Population Division staff—Theresa Andrews, Derek Azar, Ryan Engstrom, Jordan Graesser, Robert Leddy, and Nancy Schechtman—were essential to completion of this paper.

References

Comenetz, J., and Caviedes, C., 2002. Climate variability, political crises, and historical population displacements in Ethiopia. *Global Environmental Change B: Environmental Hazards* 4, pp. 113-127.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J., 2007. Completion of the 2001 National Land Cover

Database for the Conterminous United States. *Photogrammetric Engineering and Remote Sensing* 73, pp. 337-341.

McGranahan, G., Balk, D., and Anderson, B., 2007. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. *Environment and Urbanization* 19, pp. 17-37.

NRC (National Research Council), 2002. *Down to Earth: Geographical Information for Sustainable Development in Africa.* Washington, DC: National Academies Press.

NRC (National Research Council), 2007. *Tools and Methods for Estimating Populations at Risk.* Washington, DC: National Academies Press.

Tatem, A.J., Noor, A.M., von Hagen, C., Di Gregorio, A., and Hay, S.I., 2007. High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa. *PLoS ONE* 2(12): e1298 [online]. Available at: www.plosone.org/article/info:doi/10.1371/ journal.pone.0001298 [accessed July 2009].

UN (United Nations), 2009. *Census dates for all countries* [online] (updated June 30, 2009). Available at: unstats.un.org/unsd/demographic/sources/census/censusdates.htm [accessed July 2009].

U.S. Census Bureau, 2009. *International Data Base* [online] (June 2009 release). Available at: www.census.gov/ipc/www/idb [accessed July 2009].

Yang, L., Huang, C., Homer, C.G., Wylie, B. and Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. *Canadian Journal of Remote Sensing* 29, pp. 230-240.