Seeing is Believing:

Why Demographers Need Remote Sensing Data to Understand Urbanization

Deborah Balk
Baruch College, City University of New York (CUNY)
CUNY Institute for Demographic Research

and

Mark Montgomery
Population Council and Stony Brook University

27 September 2013

Abstract Submitted to the 2014 Population Association of America Annual Meeting

Introduction

The most significant demographic trend of the 21st century arguably [1] is urbanization. This is the century when the global population becomes more urban than rural, when almost all of the population growth will take place in cities (and the cities of Asia and Africa, at that) and where the growth in the number of and population size of gigantic cities is unprecedented. Placed in a spatial context, for example to study the relationship between urbanization and climate change, it is apparent that we need information on which cities are growing and at what rates, as well as their demographic and socioeconomic composition. Urbanization is an intrinsically spatial phenomenon – cities will grow at their edges as well as vertically becoming both more expansive and dense, their populations will age, and household sizes and composition will change. These changes will have wide ranging implications – e.g., for public health, education, public infrastructure and financing, housing, ecosystems and carbon emissions. Yet, demographers are ill-prepared to study these changes. They have neither the spatial tools nor the necessary data to study, let alone predict, our urban future. With colleagues from Columbia University and the UN Population Division, we have begun to address a few of these shortcomings (Montgomery and Balk, 2011; Balk et al., 2009a; Balk et al. 2009b). The study of urbanization requires information about cities – ideally at a sub-city scale – not simply the fraction of national populations that live in urban areas, which dominated demographic thinking and data collection through major advances in the 20th century (UN, 2009). In the absence of city-specific data, subnational spatial demographic data are a must and spatial tools act as the glue to assemble the data necessary to forecast city growth (Montgomery et al, 2011).

Demographic breadth comes from censuses whereas demographic depth comes from surveys. In the past decade alone substantial improvements in the spatial information of both these data sources have made. Major survey programs such as the Demographic and Health Surveys (DHS), the World Fertility Surveys (WFS), and the Multiple Indicator Cluster Surveys (MICS) have been instrumental to our understanding of fertility and mortality – and less comprehensively broad-strokes migration – in the developing world, where censuses and vital registration systems cannot kept abreast with the data infrastructures found richer countries. Yet these surveys indicate demographic behavior at the national, or first-order administrative level (like provinces or states). Owing to the usual indication of urban and rural strata (required partly to assure accuracy of the sampling frame), we can compare urban and rural populations. But it is well understood that there is, conceptually at least, an urban-rural continuum rather than a dichotomy (Champion and Hugo, 2004; Balk 2009) and therefore something more than this approach is warranted if such survey data are to be used to study urbanization. Current rounds of the DHS and MICS data include spatial data for the survey regions.[2] The DHS, in particular, has shown great leadership by adding location information to the survey cluster. While nationally representative data such as these are not intended to be analyzed as if the clusters are representative, much can be done with the cluster data: locations can be identified in proximity to cities (sometimes even particular cities) rather than adopting the much coarser urban-rural survey classification (NRC, 2003; Dorelien et al., 2013), environmental and other contextual data – including city boundaries, when available – can be added to individual and household records (Balk et al., 2004, Dorelien et al., 2013). And such data can be used in new ways to generate population surfaces (Chin et al, 2011). To be most useful for understanding our urban future, and in particular the well-being and behavior of slum dwellers and the urban

[1] With strong competition from Aging.

[2] Regrettably, for older survey rounds, and the WFS, spatial information is much sparser (and certainly not digital).

poor, survey sampling frames may need to be revised and increasingly diligent to include these traditionally hard-to-capture populations.

Similarly, census data have been quite relevant for understanding the spatial distribution of population and, in a more limited way, city growth.[3] The time-series data on city population size used by the Population Division who produces the *World Urbanization Prospects* is collected from national statistical offices and comes primarily from censuses data, yet these data are not reported with spatial boundaries or vital rates.[4] Nevertheless, improvements in the spatial resolution of census data in the past decade have been remarkable. Most countries of the world, even poor countries, release second-order (e.g., county or district-level) administrative boundaries associated with their censuses, and many – e.g., Vietnam, Peru, Malawi, Kenya, and South Africa – produce and release data at a much finer resolution. While these data permit a wealth of inquiry on the spatial distribution of population in public health applications in particular (e.g., Tatem el., 2011; Balk et al, 2006), they are inadequate in two ways: Only the basic demographic data – population counts – are available at the finest level and spatial aggregates, even very small ones, are less analytically valuable than micro-data. Furthermore, even when fine-resolution census data are available, a corresponding list of which administrative units belong to cities is often not available. This leaves the work of urban identification to other data – which is where this paper will focus its attention. Another important development with census data is the recent availability through improved collections and data delivery tools – e.g., IPUMS and Redatam[5] – for census micro-data which facilitate the estimation of rates of urban in-migration for subnational units, for example. Such estimates are invaluable for estimating city growth. For those interested demographic estimation by ecological units beyond cities, flexible use of micro-data with user-supplied spatially boundaries (such as coastal zones) would be required. For this, additional attention would need to be paid to protecting the confidentiality of census (or survey) respondents (VanWey et al., 2005, NRC 2007, Fink et al., in progress) and improving the capacity of national statistical offices to work with a variety of geographic data.

While there have been many demographic uses of remote sensing data to study particular cities or the socio-demographic causes or consequences of rural land-use change, few studies have embraced satellite data to study urbanization. Yet, satellite data are an obvious choice because countries vary in their conceptualization and definitions of "urban". The only global urban extent data base (the Global Rural Urban Mapping Project (GRUMP)[6], uses satellite data – the '1994/95 stable city night-time lights' data set (Elvidge et al, 1997) – as a consistent measure of urban spatial extent throughout the world. The imperfections and advantages of these data have been the source of much discussion (Balk, 2009; Tatem et al., 2011), yet there is general agreement that satellite data on urban areas has much to offer. To date, no analogous urban extent time-series data exist. Yet, change in these extents over time would be invaluable to demographers. To create (and perhaps even to use) such a data set would require that demographers become more competent and comfortable in the analysis of data many types of satellite data, and form strong collaborations with experts in this area.

---

[3] The oft-cited study (Chen et al, 1998) indicating a 60-40 split of the contribution of natural increase and migration, respectively, to city growth depends exclusively on census data. That important study is woefully out of date and relies on a small number of countries. Its methods precede spatial data and methods, and conflate migration with administrative reorganization, where the latter phenomena is certainly in part the result of natural increase as well as migration.
[4] Details on how we use these UN data in a spatial framework – linking them with urban location and vital rates – in order to estimate and forecast city growth are found elsewhere (Balk et al., 2009a; Montgomery and Kim, 2008).
[5] See https://international.ipums.org/international/ and http://www.eclac.cl/redatam/default.asp, respectively.
[6] http://sedac.ciesin.columbia.edu/gpw/index.jsp

The study of urbanization poses questions for the way we demographers collect census and survey data, how we associate urban spatial extents with administrative boundaries, and our models for demographic analysis. In this paper, we demonstrate how new satellite data products, and new analysis of existing satellite data, when combined with traditional demographic data can reveal shortcomings and some strengths of those data sources, and more importantly reveal more in combination than either data set can in and of itself. The goal of this paper is to demonstrate that globally consistent urban extent time-series data can be constructed from satellites, paired with city population data (from the UN Population Division's City Database) and small areal census units and georeferenced DHS coordinates. The analysis strategy is described below.

Table 1. Datasets to be Integrated and Analyzed in the Research Plan

#	Category	Name and Central Concept	Description	Resolution: Spatial; Temporal	Source
			Urban Extents		
1	Urban Extents	Global Rural Urban Mapping Project (GRUMP): Footprints	Global data base indicating urban extent footprints based primarily on NOAA's night-time lights 1994/95 stable city lights data	1 km ; 1995	CIESIN et al., 2008
2	Urban Extents	Night-time Lights	Global data base of NOAA's night-time lights radiance callibrated lights data sets	1 km ; 1992-present	NOAA
3	Urban Extents	Visible Infrared Imager Radiometer Suite (VIIRS)	New remote sensing composite data of night lights and visible information on vegegation and other daytime-detectable features of urban and surrounding areas	500 m; 2012-present	http://npp.gsfc.nasa.gov/viirs.html
4	Urban Extents	e-Geopolis (for Africa)	Modeled urban extent based on landsat and other publicly available images combined with administrative boundary data	varies (vector-data); 2000	http://e-geopolis.eu/spip.php?rubrique69
5	Urban Extents	Landsat	Globally representative sample of developed and non-developed land covertypes in/around urban areas	30 m; 1982-present	NASA
			Demographic, Socio-economic, Vulnerability		
4	Population size	UN Cities Database	Global time series of population size of approximately 3,000 cities (of 100K+ persons) in study regions, with corresponding indicator of statistical concept used as urban definition	none; 1950-present	UN Population Division (2010)
5	Vital Rates & Population Characteristics	Demographic & Health Surveys, World Fertility Surveys, and Multiple Indicator Cluster Surveys	Demographic characteristics: fertility, mortality, age, education, water, sanitation, poverty-proxy data for all countries in Africa (most with multiple years)	coarse survey regions; 1970-present	Macro International, International Statistical Institute and UNICEF
6	Vital Rates & Population & Household Characteristics	Demographic & Health Surveys Cluster Geolocation Information	Demographic characteristics at the survey cluster location: fertility, mortality, age, education, water, sanitation, and povery-proxy data for 38 countries in Africa (many with multiple years)	survey cluster point location; c.1990-present	Macro International
7	Population and Household Characteristics	AfriPop; Gridded Population of the World, v4	Gridded and underlying vector data indicating the distribution of population based on proportional allocation of census data corresponding to administrative boundaries. Variable include population counts, age and sex	various, ranging from enumeration areas to coarse first-order administrative units; 1990-2010	http://www.afripop.org/; CIESIN et al, 2004
8	Population Characteristics	IPUMS International	Harmonized international collection of sampled census microdata covering a broad range of demographic variables (among them: core demographics, fertility, migration, and urban/rural status)	various, ranging from enumeration areas to coarse first-order administrative units; 1990-2010	Minnesota Population Center, University of Minnesota

Data

Table 1 (above) identifies the data that will used be in this analysis. As indicated above, no single demographic or satellite data source will be adequate for this study. In order to attach a time-series of urban extents to match with demographic data, first that time series must be generated. To give a sense of these layers, Figure 1 below shows change in the spatial extent in China, from 1992/93-2009 from the night-time lights. (A related stable-city light night-light product, for 1994/95 and produced only once, was used as the basis of urban extents in GRUMP.) The red extents represent 1992/93 and blue 2009. GRUMP extents are overlayed in yellow or orange hollow outlines. Several observations are obvious: (1) there has been ample change in China's urban area; (2) many small extents are still not captured by the lights data products.

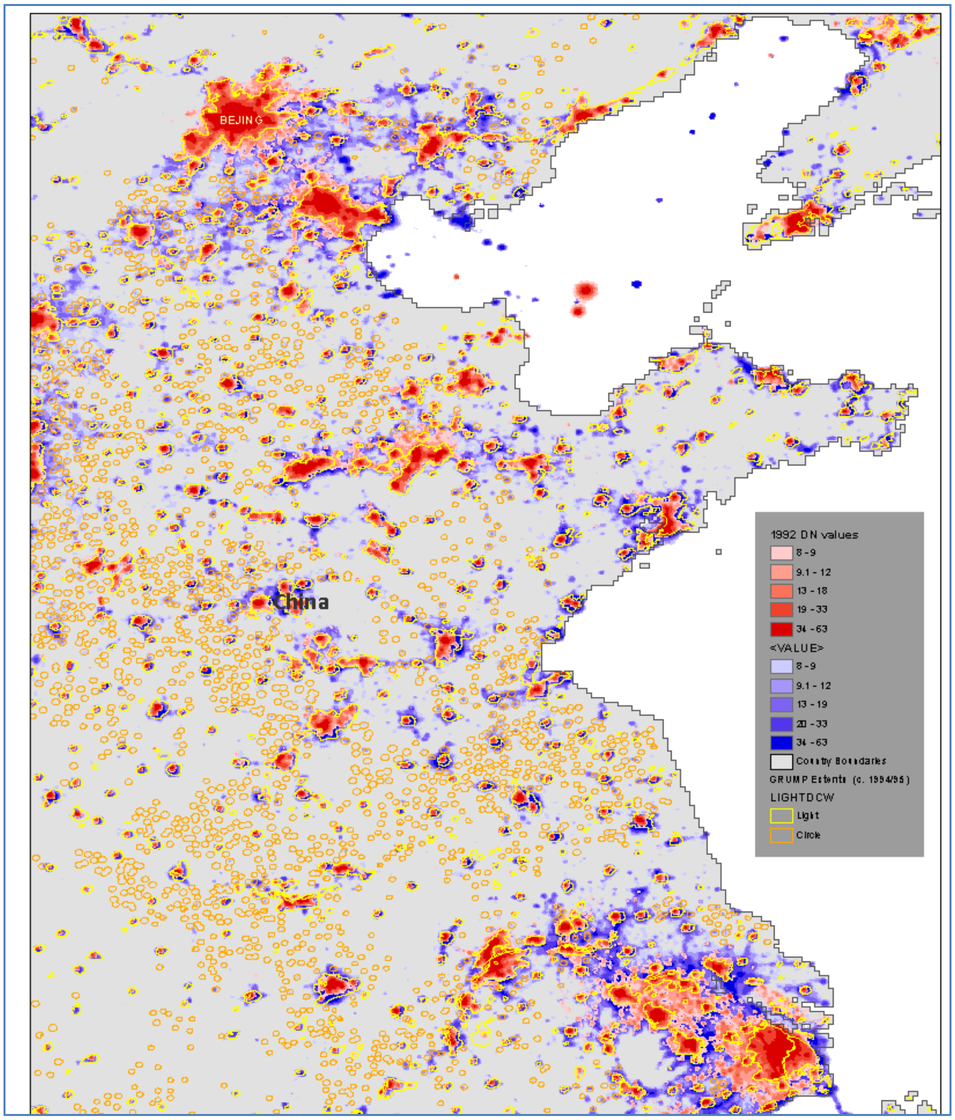


Figure 1. Night-light time series, 1992/3-2009 overlayed by GRUMP.

From prior work, we know that all cities over 100,000 persons in the UN Cities Panel database can be matched to GRUMP extents, but some of those extents – especially in the poorest countries fall below the radar of the lights. (GRUMP uses a model to estimate the extents of locations when a population settlement is known but it's area is not. See Balk 2009.) Even in China, in 2009 many small and medium-sized urban areas are not detected by the night-time lights. This problem will be exacerbated in Africa and poorer countries in Asia and the Americas. Therefore, additional satellite data sources will be used include those data from the AfriPop/AsiaPop Project. AfriPop (see Figure 2) has specialized in using satellite and other data in particular to locate small settlements – though the purpose of these data is to produce a gridded surface for use in spatial epidemiology – and they do not represent a time series. A focus on the small and medium sized cities is paramount since this is where the bulk of future growth will be and because much less is known about them, and as depicted here, they are the hardest to capture.

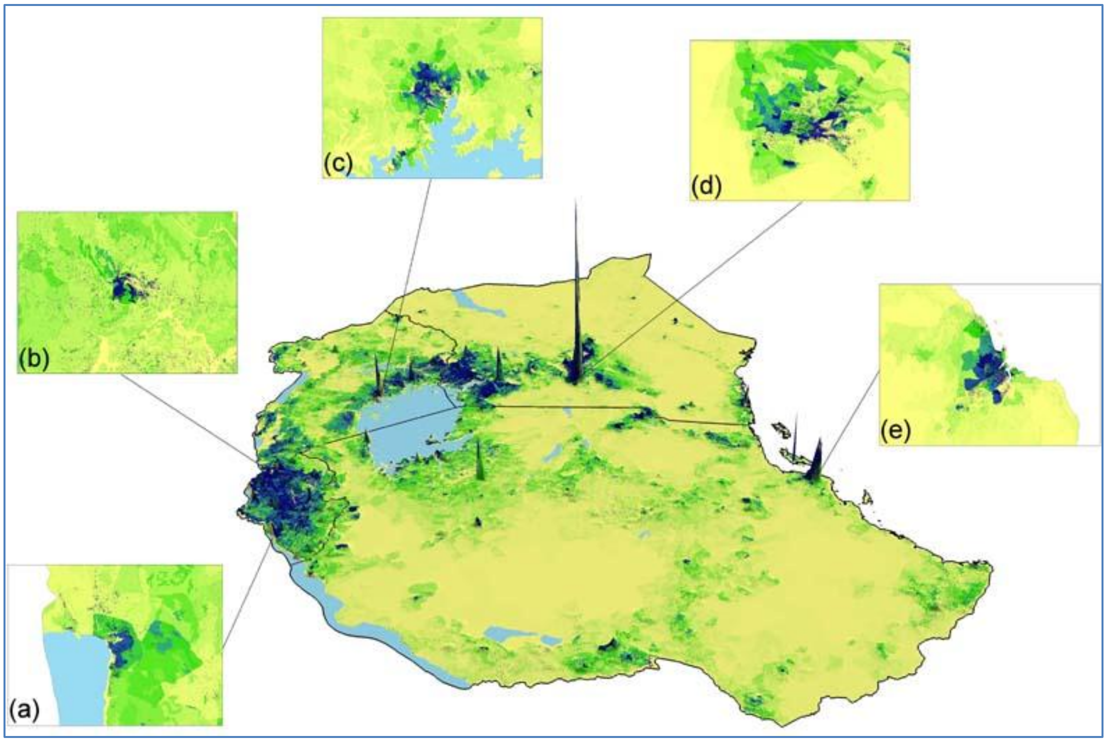


Figure 2 AfriPop Data for East Africa

An additional concern with the lights is that they bloom –have an overglow – beyond the measured area. This can be adjusted (Small et al, 2011), but to date no adjustment factor has been found to be appropriate for all places. Using new VIIRS data (see Figure 3), and new techniques with Landsat data (see Small), we will investigate the degree of blooming. By all accounts, the VIIRS data offers substantial refinement to the old night time lights data, and they are to be continued to be collected, giving them great promise for future use, but going back in time, they will still need to be associated with older projects to adjust those products. And, the lower threshold of detection is still unknown. This paper will shed light on that.

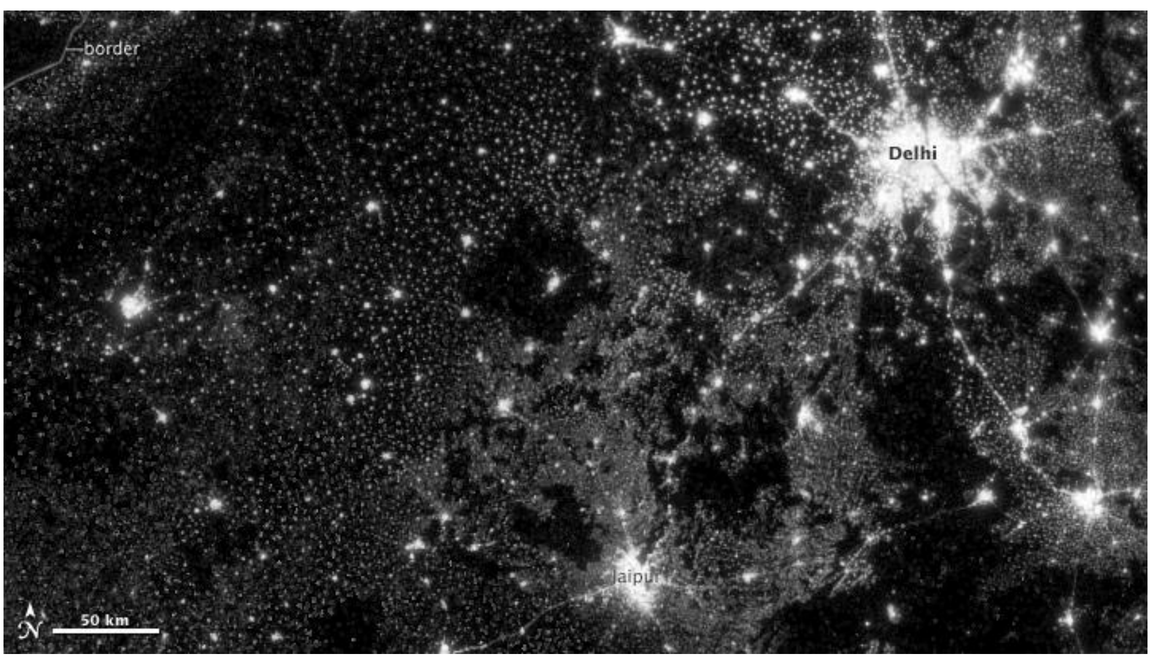


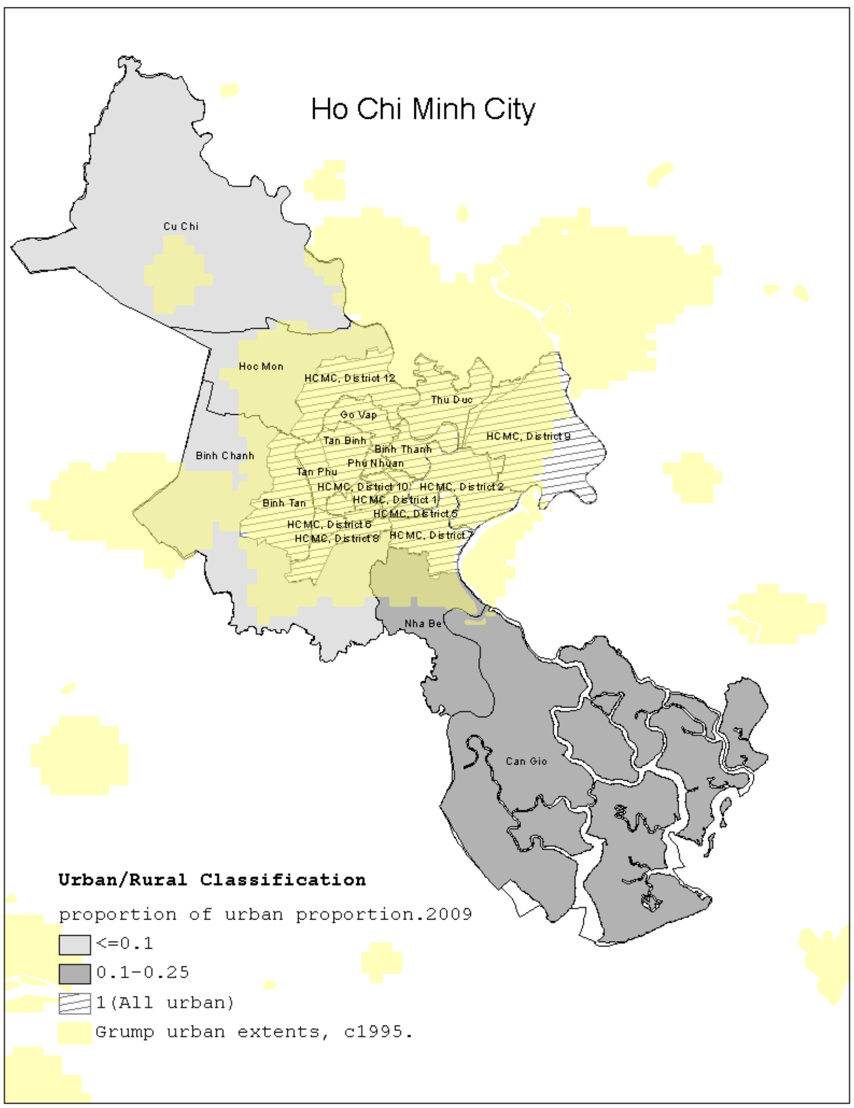
Figure 3 VIIRS Data, Delhi, India

This analysis requires substantial data integration of demographic and geographic data including satellite data, and will conform to methods used by the authors in previous work (Balk et al, 2009a, 2009b; Montgomery et al., 2011; Dorelien et al., 2013)

Analysis

The analysis plan will combine the data layers first and foremost to determine the likelihood of identifying changes in urban extents, globally, and in finding small and medium sized cities that have previously defied moderate-resolution satellite detection. It will then combine these data with three types of demographic data where available: (1) time-series from the City Database of the UN Population Division; (2) census and (3) georeferenced survey data to determine how well those data correspond and can detect places of different sizes. Georeferenced DHS data will allow us to find clustered that are classified as urban but remain undetected by such censors.

For Ho Chi Minh City (Figure 5) we see GRUMP overlayed on census classifications of urban for the year 2009 census. (The authors have access to best-available census small area unit data for the past two rounds.) Not shown are the same data for 1999 and population change. Many censuses, like Viet Nam's have additional housing characteristics and much more demographic information (including migration information) at fairly high resolution that can be used in combination with time-series satellite data (not shown) to detect areas of increased demographic concentration and expansion. (This project will not focus on Europe or North America, and therefore depopulation of urban areas will not be considered.)



By combining information in this spatial way, we will produce statistical information that can be used statistically and spatial to much more accurately indicate which cities are growing and how. This paper will focus on the description of the resulting data and the methods required to create them. It will elaborate on the implications for demographic estimation of urban trends globally as well as on city growth, composition.

References cited:

Balk, D. 2009. "More than a name: Why is Global Urban Population Mapping a GRUMPy proposition?" In P. Gamba and M. Herold, (eds.) Global Mapping of Human Settlement: Experiences, Data Sets, and Prospects, New York: Taylor and Francis, pp. 145-161.

Balk, D., M. R. Montgomery, G. McGranahan, D. Kim, V. Mara, M. Todd, T. Buettner, and A. Dorelien, 2009a. "Mapping Urban Settlements and the Risks of Climate Change in Africa, Asia, and South America," In G. Martine, et al. (eds), Population Dynamics and Climate Change, New York: United Nations Population Fund and International Institute for the Environment and Development, pp. 80-103.

Balk, D. M. R. Montgomery, G. McGranahan, and M. Todd, 2009b. "Understanding the Impacts of Climate Change: Linking Satellite and Other Spatial Data with Population Data," In G. Martine, et al. (eds), Population Dynamics and Climate Change, New York: United Nations Population Fund and International Institute for the Environment and Development, pp. 206-217.

Balk, D., U. Deichmann, G. Yetman, F. Pozzi, S.I. Hay, and A. Nelson, 2006, "Determining Global Population Distribution: Methods, Applications and Data," in Hay, S.I., Graham, A.J. and Rogers, D.J. (eds), Global mapping of infectious diseases: methods, examples and emerging applications. Advances in Parasitology, volume 62. pp. 119-156.

Balk, D., T. Pullum, A. Storeygard, F. Greenwell, M. Neuman, 2004. "A Spatial Analysis of Childhood Mortality in West Africa" Population, Space and Place 10: 175-216.

Champion, T. and G. Hugo (editors). 2004. New Forms of Urbanization: Beyond the Urban-Rural Dichotomy. Aldershot: Ashgate Publishing Company for the IUSSP.

Chen, N., P. Valente and H. Zlotnik. 1998. What do we know about recent trends in urbanization? In Richard E. Bilsborrow, editor, Migration, Urbanization and Development: New Directions and Issues, pages 59–88. New York: UNFPA.

Chin, B., L. Montana, and X. Basagaña. (2011). "Spatial modeling of geographic inequalities in child mortality across Nepal." Health & Place 17 (2011): 929-936.

Dorelien, A., D. Balk, and M. Todd, 2013. "What is Urban? Comparing a satellite view with Demographic and Health Surveys" *Population and Development Review.*

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W, Davis, E.R, 1997, "Mapping of city lights using DMSP Operational Linescan System data", Photogrammetric Engineering and Remote Sensing, 63, 727-734.
Fink, Gunther.

Montgomery, M.R. and D. Balk, 2011. "The Urban Transition in Developing Countries: Demography Meets Geography" in E. Birch and S. Wachter (eds.), Global Urbanization. (Philadelphia: University of Pennsylvania Press).

Montgomery, M.R., and D. Kim, 2008. "An econometric approach to forecasting city population growth in developing countries, in Department of Economics." 2008, Stony Brook University.

Montgomery, M. R., D. Balk, D. Kim, V. Mara, and Z. Liu (2011). "Constructing a Cities Database: Combining World Urbanization Prospects 2009 with the Global Rural–Urban Mapping Project," June 9, 2011.

National Research Council (NRC). 2007. Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed Data with Social Science Survey, Committee on the Human Dimensions of Global Change (M. Gutmann and P. Stern, eds) Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data. National Academies Press, Washington, DC.

National Research Council (NRC) 2003. Panel on Urban Population Dynamics (Mark R. Montgomery, Richard Stren, Barney Cohen and Holly E. Reed, eds), Cities Transformed: Demographic Change and Its Implications in the Developing World. National Academies Press, Washington, DC.

Tatem, A.J., Campiz, N., Gething, P.W., Snow, R.W. and Linard, C., 2011, "The effects of spatial population dataset choice on population at risk of disease estimates", Population Health Metrics, 9:4, doi: 10.1186/1478-7954-9-4.

United Nations. 2010. World Urbanization Prospects: The 2009 Revision. New York, N. Y.: Population Division, Department of Economic and Social Affairs, United Nations.

VanWey, L., R. Rindfuss, M. Gutmann, B. Entwisle, and D. Balk, 2005. "Confidentiality and Spatially Explicit Data: Concerns and Challenges," Proceedings of the National Academy of Sciences, 102 (43): 15337–15342.