MODELING THE DISTRIBUTION OF HUMAN POPULATION WITH NIGHT-TIME SATELLITE IMAGERY AND GRIDDED POPULATION OF THE WORLD

Francesca Pozzi, CIESIN
Christopher Small, Lamont-Doherty Earth Observatory
Gregory Yetman, CIESIN
Columbia University
Palisades, NY 10964, USA
fpozzi@ciesin.columbia.edu
small@LDEO.columbia.edu
gyetman@ciesin.columbia.edu

ABSTRACT

The spatial distribution of human population is a fundamental determinant of both the societal impacts and anthropogenic drivers of global change. Version 2 of the Gridded Population of the World (GPW2) is the most recent, detailed global population dataset based solely on administrative unit data. The gridding approach is based on the assumption of uniform spatial distribution of population within each of the 127,105 administrative units, thus the spatial detail of the gridded data is directly related to the spatial resolution of the administrative data on which they are based. Night-time light imagery resolves lighted settlements as small as 2.7 km in diameter thereby providing more spatially explicit information on spatial distribution of population in areas lacking detailed census data. In this study we look at the World Stable Lights dataset as a potential means to refine the spatial detail of the population dataset. We compared the Log_{10} of population density to the nighttime light frequency for sample of regions of the world with spatially detailed administrative data and found a consistent relationship between population density and light frequency. Based on this relationship, we developed a transfer function to relate light frequency to population density and a mass-conserving algorithm that relocates fractions of populations within large administrative units to locations of lighted settlements. This partial reallocation of population into urban centers provides a more spatially explicit representation of population distribution than the original GPW2 while making minimal assumptions about factors influencing population distribution.

INTRODUCTION

The importance of spatially referenced population datasets for use in integrated global, continental and regional research on human-environment interactions has been widely recognized (e.g. Clarke and Rhind, 1992). Human population continues to grow, with numbers currently increasing each year by approximately 80 million (United Nations, 1998). Nonetheless, the environmental implications associated with population growth are not only related to the total number of people, but also to their spatial distribution. Human populations are not uniformly distributed on the Earth's surface. For instance, in 1990, 50% of the human population inhabited less than 3% of the Earth's ice-free land area (Small, 2001). Such distribution is driven by both environmental and socioeconomic factors. In terms of environmental parameters, such as climate, elevation or coastal proximity, there seems to be a certain level of adaptation to different climatic conditions that allows the expansion of the human habitat, while the physical environment might induce spatial concentration of this expanding population (Small and Cohen, 1999).

If the environment might pose limits to human expansion, usually technological, political, socioeconomic and cultural changes are the drivers of such expansion. Urbanization is probably the most evident manifestation of such changes. Almost half of the world's current population lives in cities and the number is likely to increase to over 60 percent by 2030 (United Nations, 2001). Cities have become the nodes of an interconnected economic and social structure that it is not constrained by physical or climatic boundaries. In the past few decades, nations worldwide have experienced a rapid transition from a pattern of dispersed settlements across large agricultural areas to one of dense urban settlements. Such concentration of people in spatially localized areas has significant environmental and social impacts at local, regional and global levels. In addition to increased pressure on natural resources exerted by increased population densities, rapid urbanization, especially in developing countries, outpaces the development of infrastructure and environmental regulations and often results in high levels of air and water pollution and in poor management of the urban environment (Hunter, 2001).

Bringing the development of urban areas into harmony with the natural environment is one of the basic tasks to be undertaken in achieving a sustainable urbanized world (United Nations, 1996). The availability of accurate regional and global representations of the human population distribution is therefore one of the essential requirements in studies related to urbanization, agricultural transformation, biodiversity conservation and sustainable development.

To allow cross-sectional analysis, population and physical factors need to be made available as detailed, spatially disaggregated data and reduced to comparable scales. While many environmental data are available already as spatial datasets, population data usually require some modeling, especially when the task is the development of globally or regionally consistent population distribution datasets. Considering the importance of urban areas as discussed above, information about the location and size of urban settlements can greatly improve the accuracy of modeling efforts when estimating the population distribution within a country. In this paper we present an approach that looks at the Nighttime Lights dataset as a potential means to produce an improved population distribution dataset, using the Gridded Population of the World as base reference and China and Japan as prototype case studies.

GRIDDED POPULATION OF THE WORLD

Traditionally much emphasis has been put on the generation of generic datasets of physical factors, such as vegetation, climatic variables, etc, but less effort has been put into the development of generic socioeconomic datasets. As indicated by Clark and Rhind (1992) and Deichmann (1996), the reasons are predominantly related to the nature of population data, the collection methods and the lack of direct interpolation techniques to estimate the spatial distribution of socioeconomic variables. In particular, when dealing with modeling of population distribution, a wide range of issues need to be taken into consideration, such as type and sources of population and boundary data, attribute and spatial accuracy, and modeling techniques (Deichmann, 1996).

Population data are normally collected by censuses and surveys and compiled for political or administrative units. This approach is generally limiting for cross-disciplinary studies, especially when integration with environmental data is needed. Nonetheless census data can be used to produce a globally consistent population distribution dataset, provided that national differences on frequency of data collection and resolution of administrative units are carefully taken into consideration. Typically population estimates for each administrative unit for a given year are used. Although efforts have been made to obtain accurate models of population distribution at regional scales (i.e. UNEP, 1992, Switzer and Langaas, 1994, Veldhuizen et al, 1995), the first attempt to generate a consistent global population dataset was the Gridded Population of the World (GPW), produced at the National Center for Geographic Information Analysis-NCGIA (Tobler et al, 1995).

GPW was the first raster global dataset of population totals based solely on administrative boundary data and population estimates associated with those administrative units. In the original version, two datasets were produced: i) *unsmoothed*, where the gridding algorithm assigned population in grid cells with multiple input polygons by a straight majority rule, and ii) *smoothed*, where population was distributed based on a smoothing method called pycnophylactic interpolation (Tobler, 1995) that assumes that grid cells close to administrative units with higher population density tend to contain more people than those close to low density units.

Since that first release, higher resolution population data sets have been compiled for various regions of the world. In 2000 CIESIN released an updated version of the GPW: GPW Version 2 (CIESIN, 2000). GPW2 is based on an improved median resolution of 30 km for the input data layers of administrative units and increased detail of population data. Nonetheless no effort was made to "model" population distribution and no ancillary data were used to predict population distribution or to revise the population estimates. The only assumption made is that population is uniformly distributed within each administrative unit. For a detailed description of the gridding algorithm and discussion on sources of error, see Deichmann *et al.* (2001). GPW2 data for 1990 and 1995 are available at http://sedac.ciesin.columbia.edu/plue/gpw/. These include population data for the years 1990 and 1995, both unadjusted and adjusted to match United Nations population estimates for those years. Data on land area and population density are also included.

In this study we used unadjusted population data for the year 1990 and the land area data.

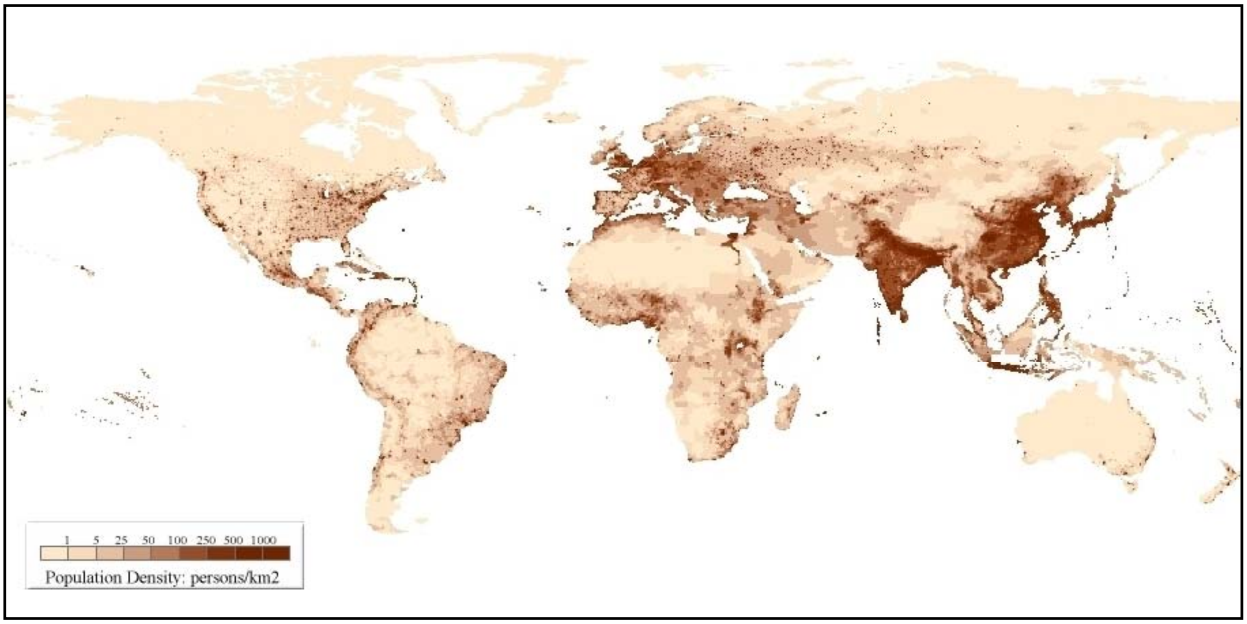

Figure 1. Gridded Population of the World: Population Density, 1990 (CIESIN, 2000)

NIGHT-TIME LIGHTS

The night-time lights dataset used in this analysis is the World Stable Light Dataset, which was produced using time series data from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) for the period 1 October 1994 to 30 April 1995 (Elvidge *et al*., 1997a, 1997b). The OLS is an oscillating scan radiometer designed for cloud imaging with two spectral bands (VIS/NIR: 0.5to 0.9 µm, and Thermal: 10.5 to 12.5 µm) and a swath of about 300 km. The visible band of the OLS is intensified at night, permitting detection of nocturnal visible-near infrared emissions from cities, towns and villages. The pixel values are measurements of the frequency with which lights were observed normalized by the total number of cloud-free observations. As detailed in Elvidge *et al*. (2001), an initial 10% detection frequency threshold was applied to extract the "stable" lights. Then manual editing and threshold filtering were applied to separate the lights into four categories: fires, lights from fishing boats, gas flares and human settlements.

The stable lights for the 1994-1995 time period have been produced for the most of the Americas, Europe, Asia and Northern Africa (Elvidge *et al.*, 1997b) and have been used for a variety of applications. For example, Sutton *et al.* (1997) examined the potential use of the stable lights data to estimate the population of urban areas around the world, Imhoff *et al.* (1997a, b) used the stable lights to estimate the extent of land areas withdrawn from agricultural production. Gallo *et al.* (1995) used the dataset in combination with vegetation data to assess the urban heat island effects. Finally Elvidge *et al.* (1997b, c) found that the area lit from from the stable lights of individual countries is highly correlated to the Gross Domestic Product.

Some issues encountered in the development of the human settlements dataset, such as the "blooming" effect, are discussed in Elvidge *et al.* (1997a, 2001) and are believed to depend on artifacts of the sensor spatial resolution and data processing. The "blooming" effect is an overestimation of the actual size of human settlements due to the large OLS pixel size, the OLS' capability to detect subpixel light source and to geolocation errors. Surface effects, such as the reflection of the lights by near shore water in the case of coastal cities can also contribute to the spread of light that can be detected by the satellite. These effects are naturally accumulated through the time series analysis. An example of thresholding techniques to reduce this effect is discussed in Imhoff *et al.* (1997a). However, because of the wide range in values for the percent frequency peaks, it is not possible to take a single threshold for sharpening the features (Elvidge *et al.*, 1997c). It is for this reason that we did not apply a threshold for this analysis. Presently, there appears to be no consistent threshold that will allow a reduction of the blooming effect without losing many small settlements that are characterized by low frequency of detection.

The dataset was originally produced at a nominal resolution of 30 arc seconds. For this analysis we aggregated the data up to 2.5 minute using bicubic re-sampling to match GPW2 cell resolution.

Figure 2. . Night-time Lights: World Stable Lights for the period 1994-1995. In white are human settlements.

ANALYSIS

To produce the new population distribution dataset, we i) examined the relationship between population density and light frequency in areas with spatially detailed census data, ii) developed a transfer function to map light frequency into population density and iii) developed a mass-conserving algorithm that relocates fractions of populations within large administrative units to locations of lighted settlements.

Transfer Function

As a first step we plotted the Log_{10} of population density against the nighttime light frequency for the entire world and for a sample of countries with spatially detailed administrative data, specifically the Northeast United States, France and the Iberian Peninsula (Spain and Portugal). As it can be seen in Figure 4, the variation in light frequency with Log_{10} of population density for the three regions shows a sigmoid increase that levels off to luminous saturation (100% frequency of occurrence) at about 1000 people/km^2. Based on this relationship, we developed the following transfer function:

$$D = \arcsin (L/100) + 1.5$$

Defined on the interval [$1.5 \le D \le 3.0$] and where

$D = Log_{10}$ (population density, persons per square kilometer)
L = Light frequency of occurrence, with $0 \le L \le 100$

This transfer function is chosen to accommodate the range of population densities over which light frequency increases from 0 to 100%. The parameters given above slightly overestimate population density for the eastern United States but is consistent with the density distributions of France and Iberia. We assume that the relationship between light usage and population density in France and Iberia is more representative of global population and that settlements in the United States are more frequently lit and less densely populated than those in other countries. By applying the transfer function as a one-sided boolean operator, we can use the light frequency to "relocate" a fraction of the population to lighted urban areas within large, low density administrative units. The one-sided boolean decision rule relocates population at densities lower than the threshold given by the transfer function but leaves population at higher densities unchanged. The result is spatial refinement in areas poorly sampled by the census data. The improvement is significant because most of these poorly sampled areas occur at the periphery of the densely inhabited region where urban settlements may directly impact nearby undeveloped environments.

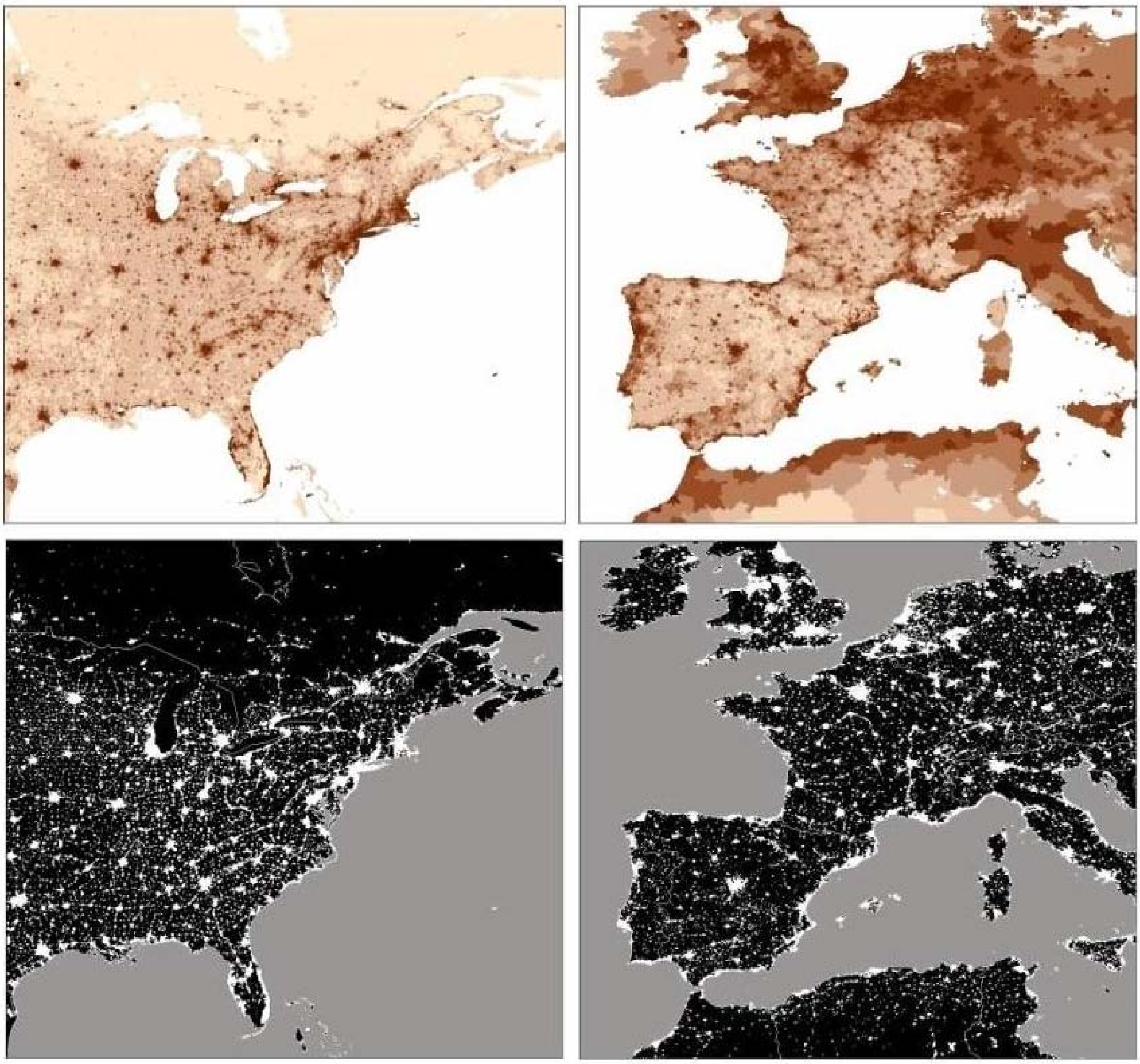

Figure 3. Detailed comparison of gridded population density and light frequency for the Eastern United States and for Europe. Spatial resolution of available census data for most of Europe is considerably lower than the resolution of the light frequency data but is comparable for France and Iberia. Spatial distributions of light frequency and population density are comparable in areas with sufficiently high census resolution. The upper right figure clearly shows the differences in census resolution across different European countries. For the population density legend, refer to Figure 1.

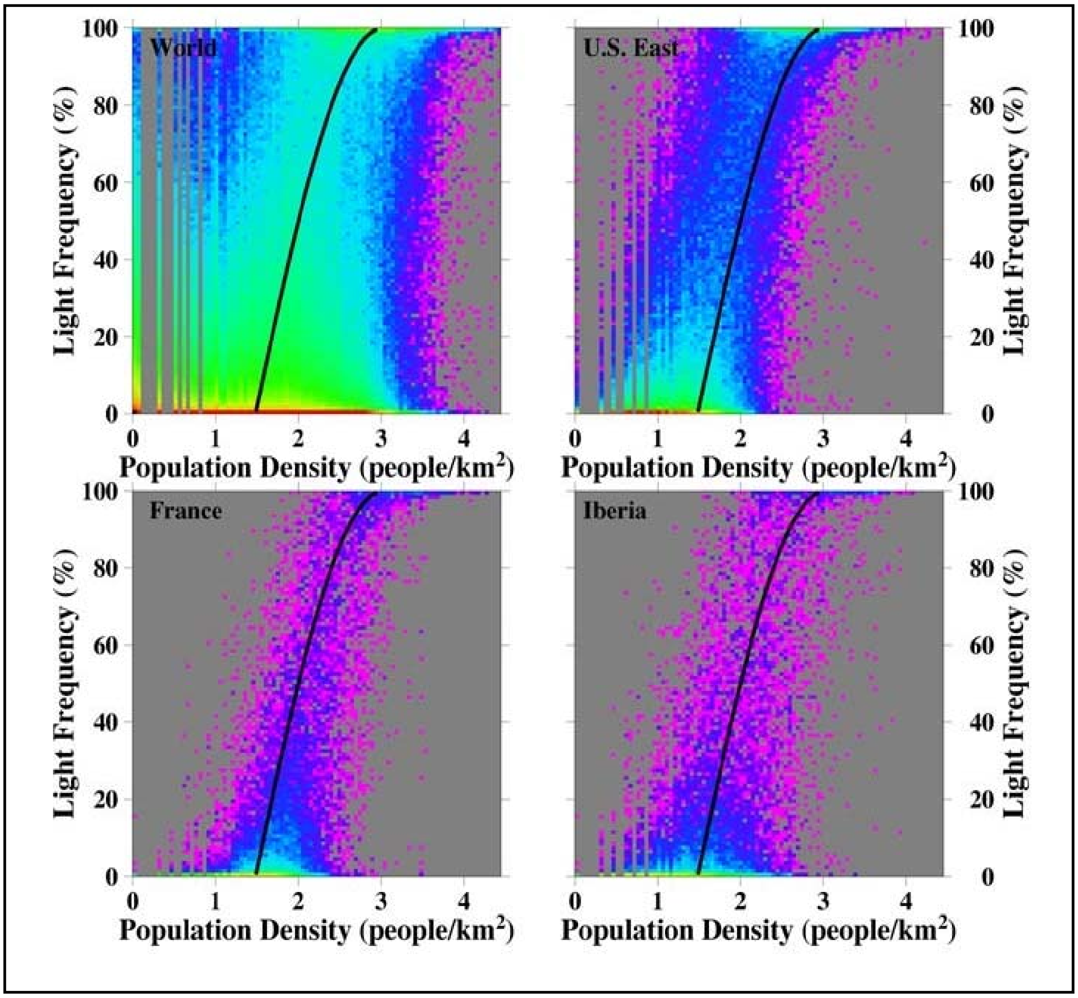

Figure 4. Bivariate distributions of population density and light frequency. Warmer colors indicate exponentially greater numbers of people. Spatial resolution of census data approaches the spatial resolution of detectable city lights in France, Spain, Portugal and the northeast United States. The superimposed curve is the transfer function used to map light frequency into population density thereby reallocating a fraction of the population in rural census districts into the areas indicated by the city lights imagery.

Mass-conserving Algorithm

The algorithm that relocates fractions of populations within large administrative units to locations of lighted settlements was based on the above transfer function, and on the decision rule that should leave the density unchanged if it is greater than or equal to the functional value at a lighted point and should increase the density to the functional value if it is less. This assumes that the points at lower densities correspond to lighted settlements in regions with coarse census districting and that points with higher densities represent settlements in areas with higher regional densities than the reference countries (Northeast United States, France, Spain and Portugal).

Case Studies. As initial case studies, countries with relatively large populations but poor spatial detail in administrative boundary data were chosen: China and Japan. The algorithm was developed as an Arc Macro Language (AML) script using ArcInfo™ with the administrative boundary and population data collected for GPW2, and the night-time lights data. The administrative boundary data for China consist of 2,445 county-level administrative units with a median resolution of 63 km and associated population estimates from the Chinese census of 1990 (Chuang *et al.*, 1990). The administrative boundary data for Japan consist of 47 prefectures with a median resolution of 90 km and population data from the Statistics Bureau of Japan (Government of Japan, 1995).

Processing Steps. A predicted surface of the Log_{10} population density was derived from the resampled night-time lights data for both countries using the transfer function described above. A boolean mask of areas where the predicted Log_{10} population densities were greater than the Log_{10} population density of GPW2 was then created. Predicted population values were then calculated as a surface for areas within this mask using the following formula:

$$P = \exp10(D) * A$$

P = Population (counts)
D = Log_{10} (population density, persons per square kilometer)
A = Land area (square kilometers, from GPW2)

The predicted population values were then totaled by administrative units and these population figures were subtracted from the administrative unit totals. In one administrative unit for China the predicted population within the mask was greater than the population of the administrative unit. Careful examination of the data revealed that the likely cause was that the administrative unit included a large suburban or industrial area outside of an urban core area that was delimited as a separate administrative unit. For this unit only the population subtracted was set to 27.4% of the population for the administrative unit. This figure is the United Nations estimate for the proportion of urban population of China in 1990 (United Nations, 1999).

The predicted population surface was then converted to vector format, where each area of contiguous predicted population was assigned the sum of the predicted population for that area. The result was joined with the administrative boundary data using a union operation. The result is a new set of administrative boundary units for each country that have been augmented with polygons derived from the night-time lights data. The population values for the large administrative units have been adjusted downwards in an amount equal to the predicted population in the new polygons contained within them. In this way the total population within each administrative unit is maintained but the spatial location has been improved by the additional polygons within the administrative units.

For the administrative unit within China where the predicted population was greater than the total for that administrative unit the population values of the night-lights derived polygons were adjusted down to equal the remainder of the total population for that administrative unit. The augmented administrative units were then processed to create grids of population estimates and land areas. This was accomplished using the same method as for GPW2, as developed by Deichmann et al. (2001). Figure 5 shows the original GPW2 Population density, the night-time lights and the reallocated population density in Japan.

Concluding Remarks

The proportion of people reassigned to the lighted cells within each administrative unit is about 15% in Japan (18,891,061 out of 123,611,995), while it is only about 4% in China (44,548,568 out of 1,130,822,942). The results of this analysis shows that in countries with relatively large populations but poor spatial detail in administrative boundary data the partial reallocation of population into urban centers provides a more spatially explicit representation of population distribution than the original GPW2. The analysis was done independently of biophysical or social parameters, thus making minimal assumptions about factors influencing population distribution. Nonetheless the type of information about the location and size of urban settlements provided by the night-time lights imagery can greatly improve the results of modeling efforts when estimating the population distribution within a country. This can be extremely useful for a global model of population distribution, especially for countries with low census resolution.

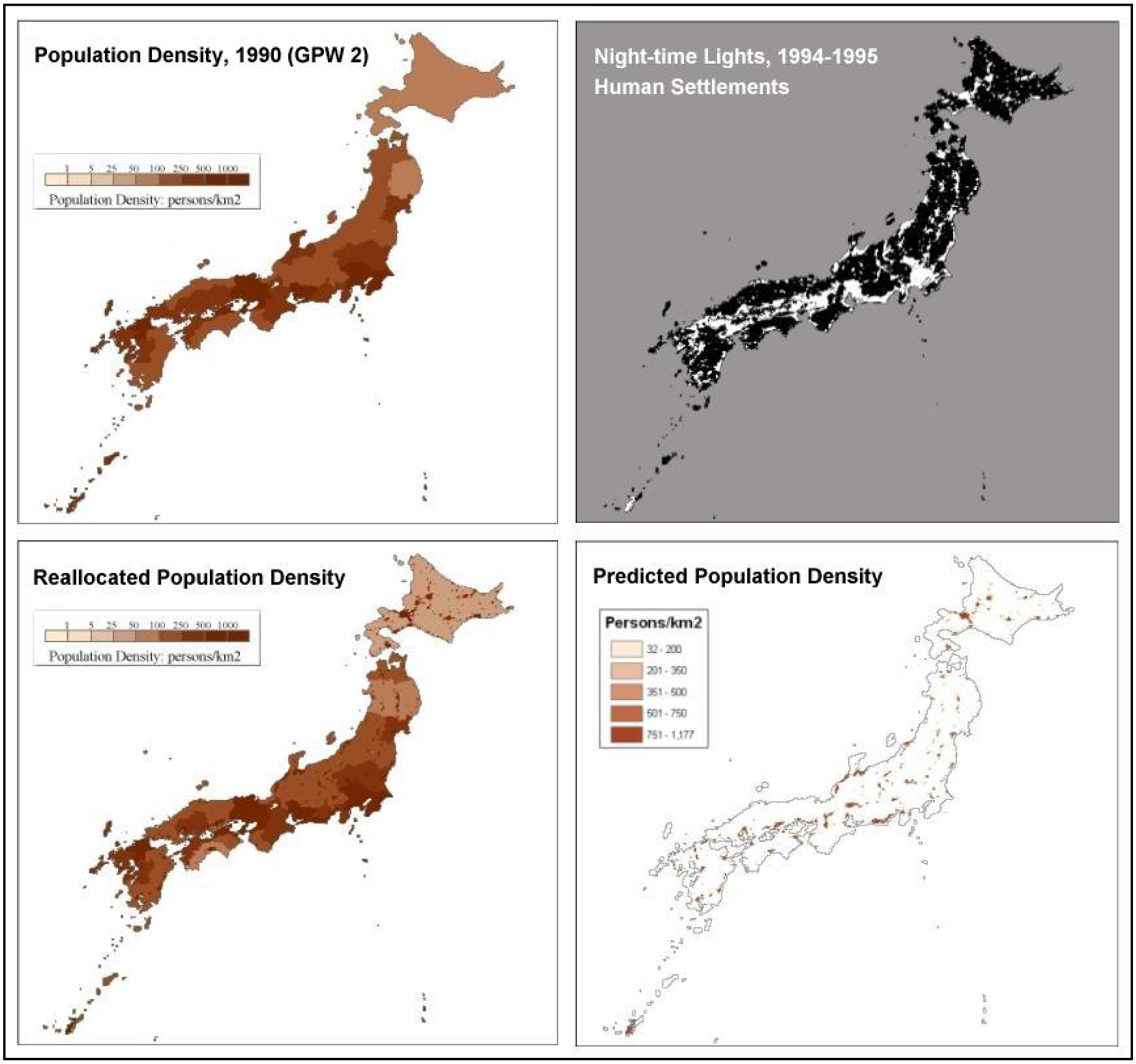

Figure 5. Population Density, Night-time Lights and Reallocated grid of population density in Japan. The lower right figure shows the Predicted population density, which represents a measure of where the population density was increased by the algorithm.

ACKNOWLEDGEMENTS

The authors acknowledge the support and resources of CIESIN's Socioeconomic Data and Applications Center (SEDAC) and the Columbia Earth Institute. The views expressed in this paper are those of the authors and are not necessarily those of CIESIN or Columbia University. The authors would also like to acknowledge Dr. Chris Elvidge for providing the World Stable Lights Datasets to CIESIN and Dr. Uwe Deichmann for his inputs on population modeling.

REFERENCES

Center for International Earth Science Information Network (CIESIN), Columbia University; International Food and Policy Research Institute (IFPRI); and World Resources Institute (WRI), (2000). Gridded Population of the World, Version 2. Palisades, NY: CIESIN, Columbia University. Available on-line at: http://sedac.ciesin.columbia.edu/plue/gpw

Clark, J.I. and Rhind, D.W. (1992). *Population Data and Global Environmental Change*, Paris, IISC/UNESCO.

Chuang, L., Xurong, Y., Lavely W. et al. (1990): China Administrative Regions GIS Data: 1:1M, County Level, available at CIESIN: ftp://ftpserver.ciesin.columbia.edu/pub/data/China/adm_bnd/CTSAR90.bnd9071/

Deichmann, U. (1996). A Review of Spatial Population Database Design and Modeling, Technical Report TR-96-3, National Center for Geographic Information and Analysis, Santa Barbara.

Deichmann, U., Balk, D. and Yetman, G. (2001). Transforming Population Data for Interdisciplinary Usages: From Census to Grid. Unpublished manuscript available on-line at: http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R. (1997a). Maping City Lights with Nighttime Data from the DMSP Operational Linescan System.*Photogrammetric Engineering and Remote Sensing, 63*(6): 727-734.

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R. and Davis, C.W. (1997b). Relation Between Satellite Observed Visible-Near Infrared Emissions, Population, Economic Activity and Electric Power Consumption, *International Journal of Remote Sensing, 18*(6): 1373-1379.

Elvidge, C.D., Baugh, K.E., Hobson, V.R., Kihn, E.A., Kroehl, H.W., Davis, E.R. and Cocero, D. (1997c). Satellite Inventory of Human Settlements Using Nocturnal Radiation emissions: a Contribution for the Global Toolchest. *Global Change Biology, 3*: 87-395.

Gallo, K.P., Tarpley, J.D., McNab, A.L. and Karl, T.R. (1995). Assessment of Urban Heat Island: a Satellite Perspective. *Atmospheric Research, 37*: 37-43.

Government of Japan (1995). 1995 Population Census of Japan. Available on-line at: http://www.stat.go.jp/

Hunter, L.M. (2001). The Environmental Implications of Population Dynamics. Report MR-1191, Rand.

Imhoff, M.L., Lawrence, W.T., Elvidge, C.D., Paul, T., Levine, E., Privalsky, M.V. and Brown, V. (1997a). Using Nighttime DMSP/OLS Images of City Lights to Estimate the Impacts of Urban Land Use on Soil Resources in the United States. *Remote Sensing of the Environment, 59*: 105-117.

Imhoff, M.L., Lawrence, W.T., Stutzer, D.C. and Elvidge, C.D. (1997b). A Technique for Using Composite DMSP/OLS "City Lights" Data to Map Urban Areas. *Remote Sensing of the Environment, 61*: 361-370.

Small, C. (2001). Global Analysis of Population Distributions and the Physical Environment. *Proceedings of the Open Meeting of the Human Dimensions of Global Environmental Change Research Community*, Brazil Academy of Science, Rio de Janeiro, Brazil, 6-8 October 2001.

Small, C. and Cohen, J.E. (1999). Continental Physiography, Climate and the Global Distribution of Human Population. *Proceedings of the Int.Symposium on Digital Earth, Chinese Academy of Sciences*, Beijing, pp. 965-971.

Sutton, P., Roberts, D., Elvidge C.D. and Baugh, K. (2001). Census from Heaven: an Estimate of Global Human Population Using Night-time Satellite Imagery. *Int. Journal of Remote Sensing, 22*(16): 3061-3076.

Sweitzer, J. and Langaas, S. (1994). Modeling Population Density in the Baltic Drainage Basin Using the Digital Chart of the World and Other Small Scale Data Sets, in Gudelis V., Povilanskas, R. and Roepstorff, R. (eds), *Coastal Conservation and Management in the Baltic Region*, Proceedings of the EUCC-WWF Conference, 2-8 May, 1994, Riga-Klapeida-Kaliningrad, pp.257-267.

Tobler, W, Deichmann, U., Gottsegen, J. and Maloy, K. (1995). The Global Demography Project, Technical Report TR-95-6, National Center for Geographic Information and Analysis, Santa Barbara.

United Nations (1996). United Nations Human Settlements Programme: The Habitat Agenda, Section IV C. Available on-line at www.unhabitat.org/unchs/english/hagenda/

United Nations (1998). The State of the World Population, 1998.

United Nations (2001). World Urbanization Prospects-1999 Revision, United Nations Department of Economic and Social Affairs, Population Division (ST/ESA/SER.A/166), Sales No. E.97.XIII.3.

United Nations Environmental Programme, UNEP (1992). *Global Atlas of Desertification*, Edward Arnolds, London.

Veldhuizen, J. van, de Velde, R. van and Woerden, J. van. (1995). Population Mapping to Support Environmental Monitoring: Some Experiences at the European Scale. *Proceedings of the Sixth European Conference on Geographical Information Systems*, EGIS Foundation, Utrecht.