The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents

Deborah Balk[1]
Francesca Pozzi[1]
Gregory Yetman[1]
Uwe Deichmann[2]
Andy Nelson[3]

[1]CIESIN, Columbia University
contact: dbalk@ciesin.columbia.edu

[2] The World Bank
[3] University of Leeds

10 September 2004

ACKNOWLEDGMENTS

The study has been undertaken by Columbia University's Earth Institute's Center for International Earth Science Information Network (CIESIN), in collaboration with the International Food and Policy Research Institute (IFPRI), the World Bank, and the International Center for Tropical Agriculture (CIAT) with primary support to CIESIN from National Aeronautics and Space Administration under Contract NAS5-03117, International Food Policy Research Center (IFPRI) under Contract 2001-X046COL, and World Conservation Monitoring Center (WCMC) for the Millennium Ecosystem Assessment under Contract CU-02225301. For comments and assistance in the preparation of this paper, we thank Adam Storeygard, Greg Booma, and Bridget Anderson. We are indebted to the many individuals for their participation in the development of the associated data products: Edwin Adkins, Sonya Ahamed, Bridget Anderson, Melanie Brickman, Greg Booma, Jordan Chamberlin, Betty Hsiu-tsiu Chang, Jessica Forrest, Christine Grimando, Robby Gutmann, Erika Helms, Glenn Hyman, Malanding Jaiteh, German Lema, Marc Levy, Rosalba Lopez, Lisa Lukang, Melissa Neuman, James "Cuz" Potter, Kate Sebastian, Chris Small, Adam Storeygard, Maarten Tromp, Katya Vasilaky, Keelia Wright, Stan Wood, and Xiaoshi Xing.

ABSTRACT

What is known about the urban world is largely derived from local knowledge. This paper showcases substantial efforts at new data integration with existing technologies to develop a new suite of global datasets on urban population and extents. These new databases far surpass past efforts to construct a systematic global database of urban areas by combining census and satellite data and methods of analysis, in an integrated geospatial framework. The resulting data allow for inquiry into analysis urban issues and population by environmental and other ecological characteristics in novel ways. This paper focuses on the methodologies employed to construct these new datasets. Summary results regarding population distribution at continent- and global-levels are also given, , as well as suggestions for future research.

INTRODUCTION

Human settlements occupy a relatively small fraction of Earth's surface area but their extent and distribution have significant impacts on their surroundings, both from an environmental and a socio-economic perspective. By 2007, it is estimated that over half of the world's population will reside in urban areas (United Nations, 2002). Despite increasing knowledge about the characteristics of urbanization, little is known about its spatial dimensions. For example, only guess work has provided prior estimates on the share of the world's inhabited land area that is urbanized, or on the classification of the world's population by city-sizes other than the very largest ones (UN, 2002; UN 2000x [Demographic yearbook]. Even when cities are tallied by their population sizes and types (such as agglomerations), little effort has gone into a systematically capturing the spatial dimension of these places.

There are also increasing demands for greater specificity in defining the impacts of agricultural change and development, particularly with regard to the likely impacts of policy, technology, and institutional changes on poverty (Wood *et al*. 1999). It is not only the growth of urban areas, but also the interconnection between urban and rural areas that is important to understand these impacts. Improved knowledge of the spatial distribution of urban and rural population is extremely important for assessing socioeconomic, demographic and environmental change in urban and rural areas.

In order to understand and study the impacts of urbanization, population and physical factors need to be made available as detailed, spatially disaggregated data and reduced to comparable scales. Although there is ample research on urban growth as separate geographic and demographic phenomenon, there is little research or dataset in which these parameters are integrated. This study proposes a new methodology to foster this integration. That methodology is the focus of this paper, along with the discussion of some analytical results and suggestions for future research.

BACKGROUND

While many environmental data are available already as spatial datasets, demographic data which tend to be collected by administrative units, require some form of spatial allocation, to convert irregularly shaped census units to globally or regionally consistent population grids. Several researchers and institutions in recent years have used new methods and data to map the global distribution of human population. The first major effort to generate a consistent global georeferenced population dataset was the Gridded Population of the World (GPW, CIESIN and CIAT, 2004), produced at the National Center for Geographic Information Analysis (NCGIA) in 1995 (Tobler *et al.*, 1997), and updated by CIESIN in 2000 (Deichmann *et al*., 2001). The input to the GPW dataset are solely administrative boundary data and population estimates associated with those administrative units. Other efforts then followed, generally incorporating satellite data and other ancillary data. In the LandScan database, for example, developed by the Oak Ridge National Laboratory, "sub-national census counts are apportioned to each grid cell based on likelihood coefficients, which are derived from proximity to roads, slope, land cover, night-time lights and other data sets, (Dobson et al, 2000). A third effort by UNEP and partners includes an "accessibility" model, whereby access to roads as a means to reallocate population (UNEP, XXXX, and CIAT, XXXX). Each of these approaches has strengths and weakness (see Dooley, 2004): GPW uses a simple areal weighting scheme for reallocation, uses the best possible census and administrative data available, but does not model population distribution. Its output resolution is 2.5 arc minutes. LandScan, conversely, does not use very high resolution population input data, but uses an extensive model to reallocate people on a 30 arc second grid. Since some of the data used to reallocate persons may be outcomes of interest (e.g., distance to roads, or land cover), LandScan must be used with caution in studies involving environmental outcomes. GPW attempts to represent decennial population counts, whereas LandScan attempts to capture ambient—or at risk—population. The Accessibility model builds on the GPW tradition, but takes into account road networks in the reallocation of population. Unlike LandScan, only roads are used, and there is no explicit effort to capture the ambient quality of the LandScan approach. Like GPW, its output resolution is also moderate (2.5 arc minutes). The study presented here forwards a new methodology to extend the initial GPW efforts, to improve output resolution, and to overcome some of the use limitations of LandScan by accounting explicitly for urban areas.

In the process of reallocating population to urban areas, it is necessary to first construct a spatial database of those areas. To accomplish that, satellite data were a necessary additional input. The night-time lights dataset (DMSP-OSL) has been increasingly used to estimate aspects of human activity at the global level. Satellite imaging of stable anthropogenic lights in fact provides an accurate, economical and straightforward way to map the global distribution of urban areas, and several studies of DMSP-OLS stable night lights have indeed shown encouraging agreement between temporally stable lighted areas and various definitions of urban extent. 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. As with the population databases, there have some relevant uses of the night-time lights data. For example, Sutton *et al.* (1997) examined the potential use of the stable lights data to revise estimates of the population of urban areas around the world; Imhoff *et al.* (1997a, 1997b) used the stable lights to estimate the extent of land areas withdrawn from agricultural production; and Elvidge *et al.* (1997b, 1997c) found that the area lit from the stable lights of individual countries is highly correlated to the Gross Domestic Product. More recent efforts include those of Schneider *et al.* (2003), which is a pilot study to map urban land cover by fusing the night-time lights dataset with GPW and a MODIS-derived land cover classification, and of Pozzi *et al.* (2003), to map global urban population by integrating GPW and the night-time lights. None of these efforts attempts to merge the lights, however, directly with city-level census data to derive population estimates of urban areas.

The method presented in this paper is based on the premise that data may be combined from several key data streams: census (or census-type) inputs on the population size (of settlements and administrative areas), with associated names; and two key pieces of geographic information, latitude and longitude of settlements, and boundaries for administrative areas and urban extents (the latter being identified using the night-time lights and ancillary geographic datasets). The resulting datasets is actually constituted by three separate products: a human settlements database, an urban extent database, and the urban-rural population grid or surface.

Thus, using GPW as a base, in 2000, CIESIN, IFPRI, the World Bank and CIAT began the multi-year effort of the Global Rural Urban Mapping Project (GRUMP). This effort aimed not only to construct an improved population grid, but also to construct a globally consistent database of urban areas.

METHODOLOGY

These three products are 1) a Human Settlements database, 2) an Urban Extent database or mask and 3) an Urban-Rural population grid or surface. Here we describe the methodology—and underlying data—used in the development of these datasets.

1.1 Human Settlements

Although there are many gazetteers listing population places, few of these contain population estimates for those named places. Similarly, databases of city population estimates rarely include geographic information, such as the latitude and longitude coordinates let alone area or other spatially explicit information about each urban area. Several collections, such as the UN *Demographic Yearbook* (UN, 200X) or the UN's *World Urbanization Prospects* include coordinates, and type of urban area, for places of 100,000 and 750,000 persons, respectively.

The GRUMP human settlements database is a global database of cities and towns of 1000 persons or more, where each settlement is spatially represented as a point, and has associated tabular information on its population and data sources. Population data were gathered primarily from official statistical offices (census data) and secondarily from other web sources, such as Gazetteer (www.gazetteer.de) and CityPop (www.citypop.de), or from specific individual databases when official statistical databases were not available. Based on the data available and applying UN growth rates, we estimated population in 1990, 1995, and 2000. In some cases, the records for cities and town included latitude and longitude coordinates. For those where coordinates were not available—that is, most places—we matched the settlement name and administrative units with the National Imagery and Mapping Agency (NIMA) database of populated places (gnswww.nima.mil/geonames/GNS/index.jsp). Although we automated the matching of places to coordinates found in the NIMA database, this process still required considerable validation with other sources, and sorting through multiple options (i.e., NIMA often provides several, slightly varying sets of coordinates to match a single place name in a given administrative unit.) Table 1 shows the distribution of data sources, while Figure 1 shows the points database in a portion of South America, by population size.

Source	Asia	Africa	Europe	North America	South America	Oceania	World	Percentage
Census	9,666	2,525	6,641	27,493	4,889	451	51,665	73.22
World Gazetteer	2,210	561	4,799	243	74	168	8,055	11.42
CityPop	1,363	319	3,364	443	304	179	5,972	8.46
Others	7	737	0	119	4,002	1	4,866	6.90
Total	13,246	4,142	14,804	28,298	9,269	799	70,558	100.00

Table 1. Distribution of sources by continent. "Census" include also data from Statistical Yearbooks, Statistics from State Departments and on-line Statistical Offices. "Others" include data from AFCities, ASCIties, the World Bank, CIA Factbooks and CELADE (Latin American and Caribbean Demographic Centre).

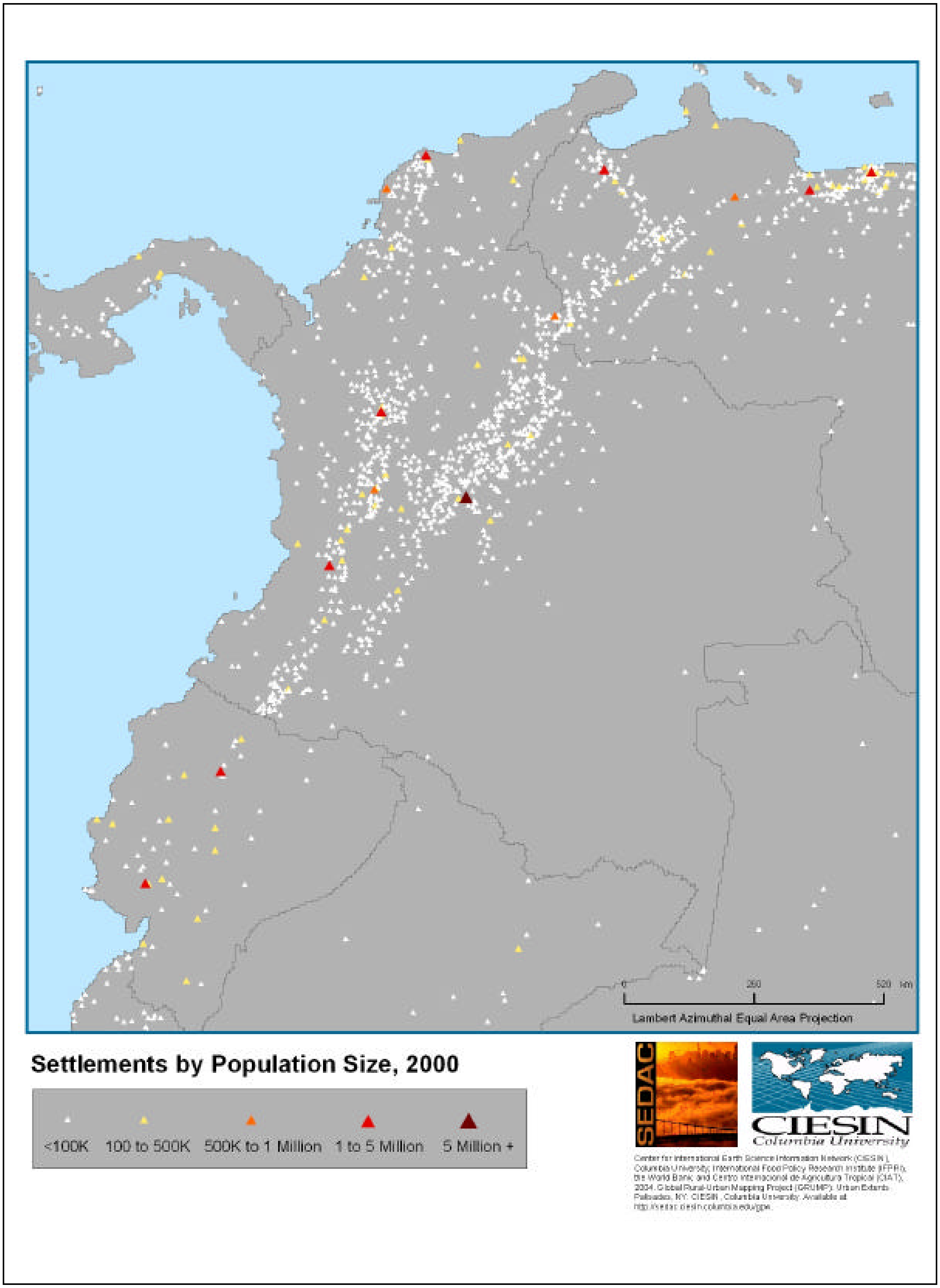

Figure 1. Point database in a portion of South America, by settlement size.

1.2 Urban Extents

While much research has been undertaken to describe the extents, landscapes, and changes of local urban areas (REF), none has been undertaken in a systematic way at the global level. Efforts to use moderate-resolution vegetation-detecting optical satellite imagery prove too costly, and inconsistent, to detect built-up areas globally (Small, 200x). While other satellite data, such as new radar data from the SRTM, holds promise for the detection of built-up areas globally, (Ngheim et al., 2001), the effort has yet to be attempted.

Thus, the GRUMP Urban Extent Mask attempts to somewhat crudely represent the extents associated to the human settlements. In particular we now describe the sources of the physical extents of the settlements and the methodology to assign population from the point database to the areal extents.

The physical extent of settlements has been derived both from raster and vector datasets, in particular:

- Night-time lights, 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, where the pixel values are measurements of the frequency with which lights were observed normalized by the total number of cloud-free observations. To delineate the physical extent of human settlements we used the World Stable Lights dataset ("cities" component).
- Digital Chart of the World (DCW)'s Populated Places: an ESRI product originally developed for the US Defense Mapping Agency (DMA) using DMA data and currently available at 1:1,000,000 scale (1993 version). The "populated places" coverage is available for most countries and contains depictions of the urbanized areas (built-up areas) of the world that are represented as polygons at 1:1,000,000 scale.
- Tactical Pilotage Charts (TPC): standard charts produced by the Australian Defense Imagery and Geospatial Organization, at a scale of 1:500,000, originally designed to provide an intermediate scale translation of cultural and terrain features for pilots/navigators flying at very low altitudes. Each chart contains information on cultural, drainage/hydrography, relief, distinctive vegetation, roads, sand ridges, power lines, and topographical features. Settlements are reported both as polygons and points. Polygons and points were digitized for a number of countries, especially where night-lights and DCW data did not adequately delimit urban areas.

All the sources of urban extent (night-lights, Landsat-derived polygons, DCW and TPCs) were combined in order to obtain the maximum possible coverage for each country. The methodology followed to combine the different sources was to use the night-time lights as baseline (due to its global coverage), and then add polygons from other sources that did not intersect any existing light. Therefore, the total number of urban polygons in each country is the number of lighted areas plus all the other polygons identifying settlements that were not already identified by the lights.

These polygons do not have attribute data, such as population, associated with them, but are characterized only by the basic geographic attributes, such as area and perimeter. To create the Urban Mask from all the different sources we developed a hierarchical process, as follows.

First, we assign population from the points to the settlement extents, based on a 3 km buffer distance. If multiple points are present, as in the case of an urban agglomeration, the sum of the population of all points is assigned to the polygon. The name of the most populous city within the buffer is also assigned to the polygons. Then, we estimate areal extents for points without polygons, based on a relationship between population size and areal extents for the points with known parameters. Using a logarithmic regression, we predict the expected geographic size of a place, given its population size. Where the number of observations is greater than 20 known relationships, we use country-specific regressions, otherwise regional regressions were used where regions were constructed according to the UN Statistics Division (UNSD) sub-continental regional codes. Based on these area values, we create circles, centered on the points. Finally, we add these newly created polygons to the existing ones to create a complete urban extent coverage for each country. Figure 2 shows the extent of urban places (as identified by the urban mask) in a portion of South America.

One of the main problems encountered when using the night-time lights dataset as a baseline to identify urban extents is the blooming effect. Due to the intrinsic characteristics of the sensor, as detailed in Elvidge *et al.* (2004), the lights tend to overestimate the real extents of urban areas. Early efforts to threshold the lights, such as by Imhoff and colleagues (1997), are inappropriate in this context because their study sample was small and concentrated in a well-light region of the United States. Our efforts were to construct a global database, thus detection is more of a concern than blooming. Recent research by Small *et al.* (forthcoming or expected in 2005?) shows that, even though a 10% threshold could reduce the blooming effect without significantly attenuating many individual small settlements for the 1994/1995 dataset, this detection frequency threshold does not provide a globally consistent basis for reconciling lighted areas to urban extent.

While blooming is noted to be a problem, it is probably much less of one for the production of a global population distribution grid because that redistribution is to go from even much coarser administrative units to these urban areas. Thus, the direction of the reallocation we argue is a vast overall improvement in the database. Furthermore, for the largest cities, where blooming is probably greatest, there tend to be better sub-urban administrative units, so that the population within the extents will show the detail of the underlying detail. Nevertheless, future work should continue to determine the possibility of reducing the lights, as appropriate, so that the blooming effect is minimized.

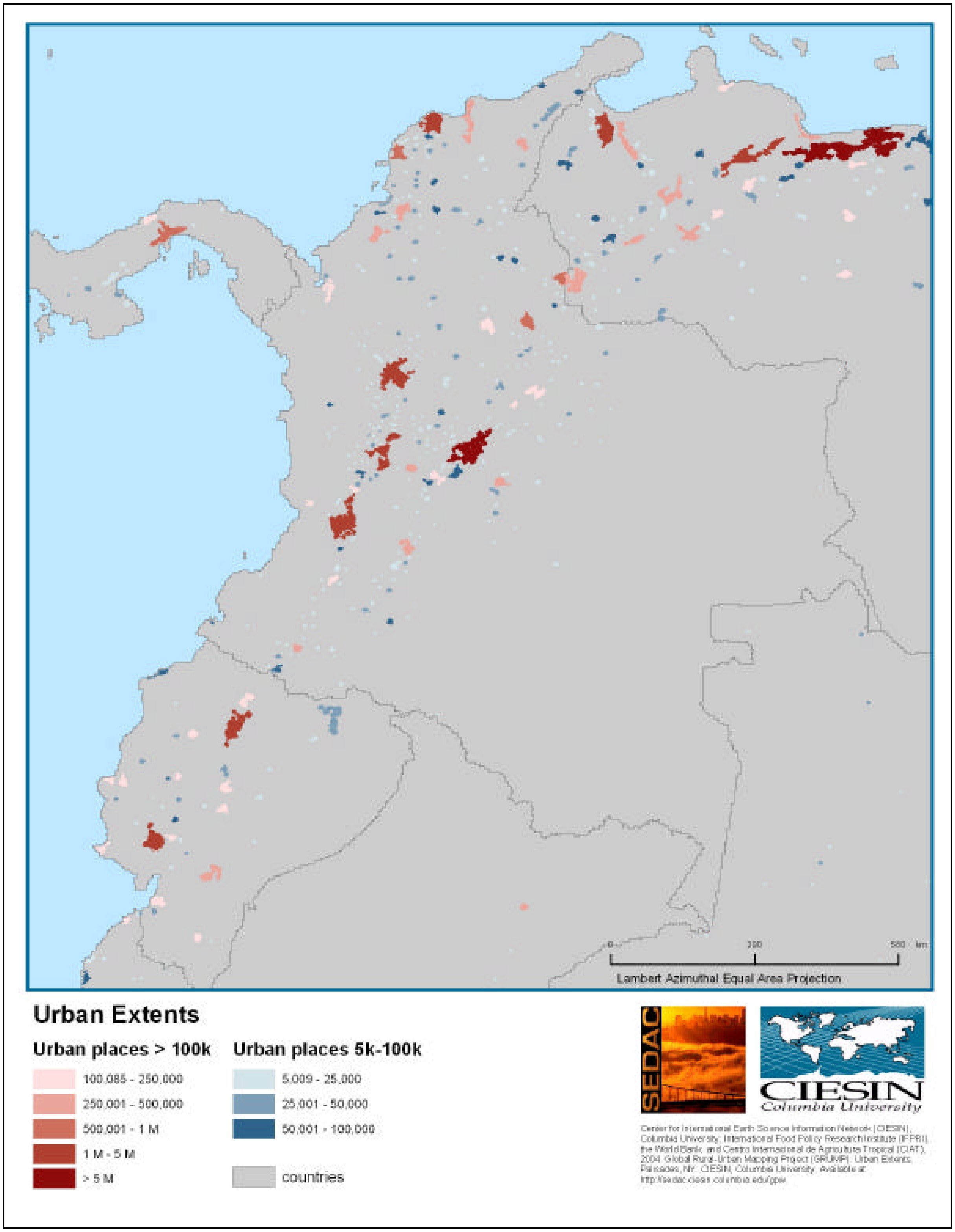

Figure 2. Urban extents in a portion of South America, by settlement size.

1.3 Urban-Rural Grid

The Urban-Rural grid is a 30-arcsecond population distribution raster dataset that was developed by combining population data from the census administrative units and from the Urban mask. To create the urban-rural surface, we developed a mass-conserving algorithm called GRUMPe (Global Rural Urban Mapping Programme) that reallocates people into urban areas, within each administrative unit. In particular we used data inputs from two vector sources: (1) Administrative polygons, containing the total population for each admin unit. (2) Urban areas, containing the urban population for each area.

These two data sets are combined using the ArcInfo command IDENTIFY. The new dataset is then passed to GRUMPe (a stand alone model written in C) that assigns population to each new polygon and labels it as rural or urban. The calculation of the reallocation of people within urban areas is based on the assumption that the total population remains constant within a given administrative unit but the rural and urban distributions change while other conditions, regarding urban and rural densities, are met. These conditions are controlled by parameters that are passed to the AML on the command line. If no parameters are specified then the AML will assign fixed values that have been empirically determined to be good first estimates.

The resulting map is shown below (Figure 3, with a close up of Cali, Colombia, showing the data before and after running GRUMPe). Note how, where urban areas are present in a given administrative unit, the density of the GPW administrative units decreases after GRUMPe because people are reallocated into their respective urban areas.

The final results from each country are recorded in an excel spreadsheet to compare the output rural and urban population totals to the UN totals. The output coverage from GRUMPe is then converted to a grid, at 30 arc-seconds resolution.

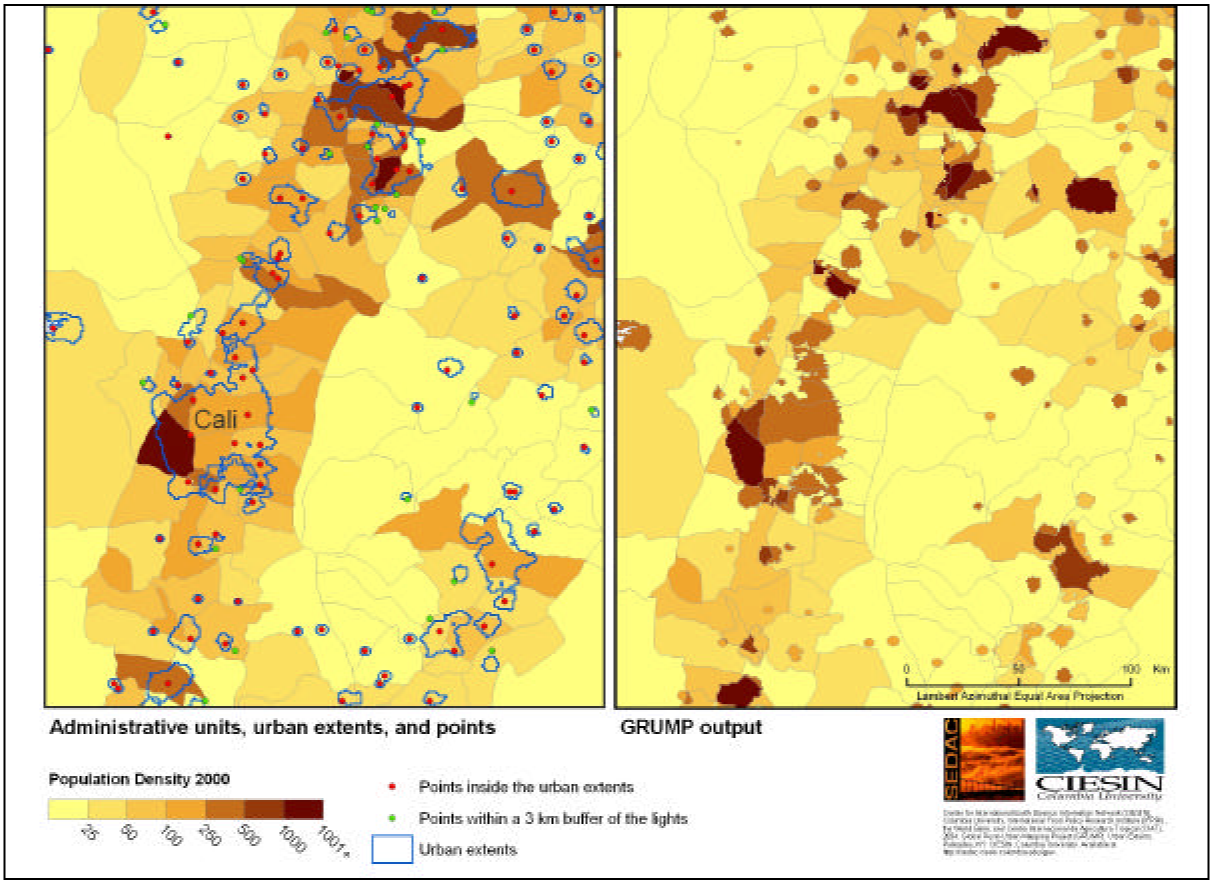

Figure 3. Close-up view of Cali, Colombia, and the surrounding areas, showing population density of the original administrative units along with the urban extents and the point settlements (l) and the population density resulting from the reallocation process (r). In the left-hand figure, the points that fall completely within the urban extents are highlighted in red, those falling within the 3km buffer in green.

DISCUSSION

The methodology presented has undoubtedly some advantages and some disadvantages compared to existing datasets, such as GPW, LandScan and other approaches.

First, compared to the CIESIN approach discussed in Pozzi *et al.* (2003), Schneider et al. (2003), the Accessibility model, and to LandScan, this method has the advantage of not predicting population density based on probability coefficients or lighted areas, but it actually uses population data for settlements from censuses. Therefore we have an independent and reliable measure of population. Further, this methodology makes use of other GIS data to identify urban areas, compensating for the small settlements in poor countries that are not detected by the night-time lights. We know that the lights dataset has two main problems: the blooming effect and the insufficient detection of small settlements that are not frequently illuminated. While there is not yet a method to improve upon these two elements at a global scale, using ancillary data to identify small settlements has proved useful in several countries in Africa. Second, although this method produces a model surface, as opposed to a more heuristic one from GPW, it allows for improvements not only in resolution but also in the positional accuracy of human population distribution.

As for GRUMPe—the mechanism through which the modelling occurred—it proved to be a good tool to refine GPW in countries where administrative data is coarse. Although the administrative data in Colombia is relatively good, the size of the units is such that the reallocation works very well. As shown in Figure 3, there are cases of relatively large administrative units with one or two cities within, and we can clearly see how the GRUMPe assigns people to the urban areas, decreasing the density of the remaining administrative unit. Then, there are cases, like Cali, where the urban areas identified by the lights expand over several administrative units. Also in this case, the reallocation process outputs a population distribution that is more consistent with the notion that people tend to be more concentrated into the urban areas rather than uniformly distributed across an administrative unit.

GRUMPe also proved to be an effective way to compensate for the blooming effect in some countries where administrative data include city boundaries. In this case the total population allocated to the light is larger than the census value for the same city, but, given the type of administrative data, the result is a structure that has a more densely populated urban core at the center of the light, surrounded by a lower-density outskirt, as shown in Figure 4.

We also have found some instances where GRUMPe does not work as effectively as in other cases. First, in countries GPW is very detailed, the reallocation process does not improve the information content about population distribution. In Malawi, for instance, there are more than 9,000 administrative units, but only about 40 urban extents, most of them very small. Therefore GRUMPe does not provide any additional information on population distribution. This is not so much of a problem, rather than noteworthy for consistency sake, since most countries do not have such high-resolution data.

Another example of the GRUMPe limitations is related to small, populated islands, like several islands in the Caribbean or in the Pacific. These islands might have one or two isolated small urban centers, but, due to the blooming effect, appear all lit. In this case, the reallocation of the population into urban areas and rural areas is not very effective, as the urban areas identified by the lights could cover the entire islands, even though the urban population is only a fraction of the total population, and the administrative data is generally good. Fortunately, inputs in the underlying administrative and population data for GPW version 3 have improved substantially for more than half of the world's island nations (Balk and Yetman, 2004).

Where the administrative data are poor in the sense that they attempt to approximate urban centers, but do so inadequately (e.g., the construction of small triangular shapes to represent urban centers in the former Soviet republics) and the lights data are moderate to poor, GRUMPe may assign too high a population density value to such a small area. In this case, its general assignment is correct, but the extent is limited both by the lights and the administrative data's shortcomings.

In sum, GRUMPe performs moderately well. Where administrative data and the lights data are good, GRUMPe does not perform very well. However, in these instances, there is less imperative for it to work well; it is only a problem in the event that it introduces error or degrades the data quality, both open questions at this point. Where administrative data are moderate or poor, and the lights (and more intensive substitutes) are moderate to good, GRUMPe performs very well. Where both the administrative data and the lights (or its substitutes) are poor GRUMPe just doesn not have much to work with. As is generally the case, there are no perfect substitutes for good data.

The main drawback of this methodology, is not GRUMPe rather, is the complex and time-consuming procedure that goes from collecting and processing the census data, to combining the city population with the spatial information about the settlements and finally to reallocating people from the administrative units into the urban centers. Some of that complexity could be reduced as institutional capacity increases in the production and distribution of urban data, as has already happened for administrative data over the past 10-15 years (see Balk and Yetman, 2004). Further gains may be made by establishing international guidelines on the definition and correspondence between metropolitan areas of different types (see Champion and Hugo, 2003). In hindsight, such guidelines would make an invaluable contribution in reducing the processing time, but also increasing the accuracy of the underlying point data, upon which both the extent mask and population surface are based.

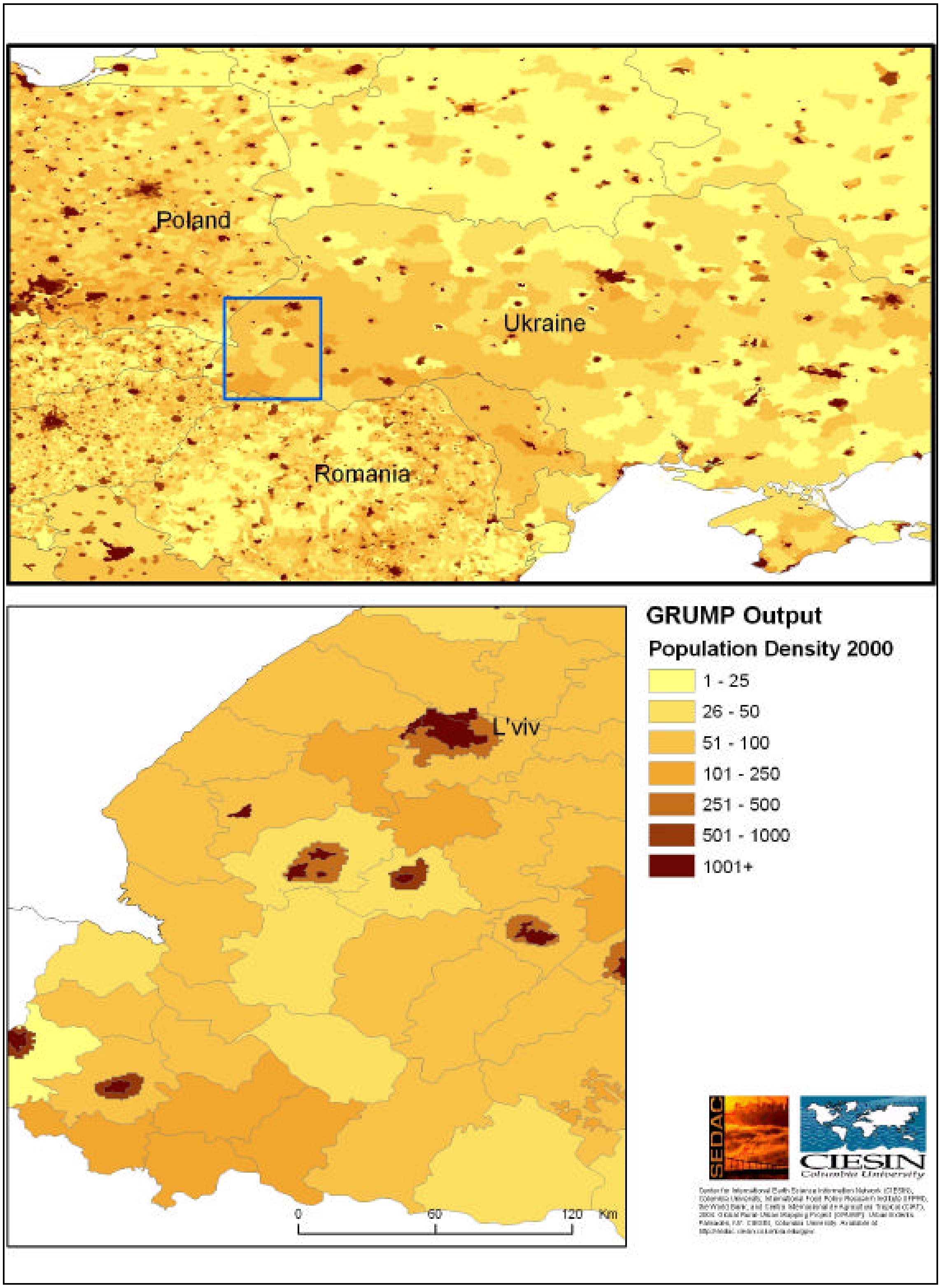

Figure 4. Grump output in Eastern Europe, and close-up view of western Ukraine, showing the effect of the grumping process where administrative data include city boundaries smaller than the urban extents derived from the lights.

RESULTS

Here we discuss some of the continent- and global-level results regarding the number and size of settlements, urban extents, and the characteristics of the three data product, in terms of in terms of input and output. Table 2 shows the input and output characteristics for the 3 data products. [Add and Describe].

Table 2. Characteristics of the three data product. Note: as the GRUMP input are the GPW administrative units, we report that number as input of the urban-rural surface grid.

Figure 5 shows a continental breakdown of the number of settlements by population size, in the database (noting the truncated display for North America). This figure speaks both to the availability of data and to the world's settlement patterns. There is much more information on smaller settlements for the more developed or urbanized regions of the world, i.e., in North and South America, Europe, and Oceania. While it is probably true that in these continents, as well as globally, there are many more small settlements than larger ones, the smaller ones require greater institutional capacity on the part of national statistical offices to track and disseminate information about. Thus, in Asia and Oceania, the GRUMP databases rely on less information about settlements below 20,000 persons. Cautious users, therefore, might want to apply a city-size threshold of 20,000 persons, if they are interested in making comparisons across countries that do not reflect data collection as much as settlement patterns.

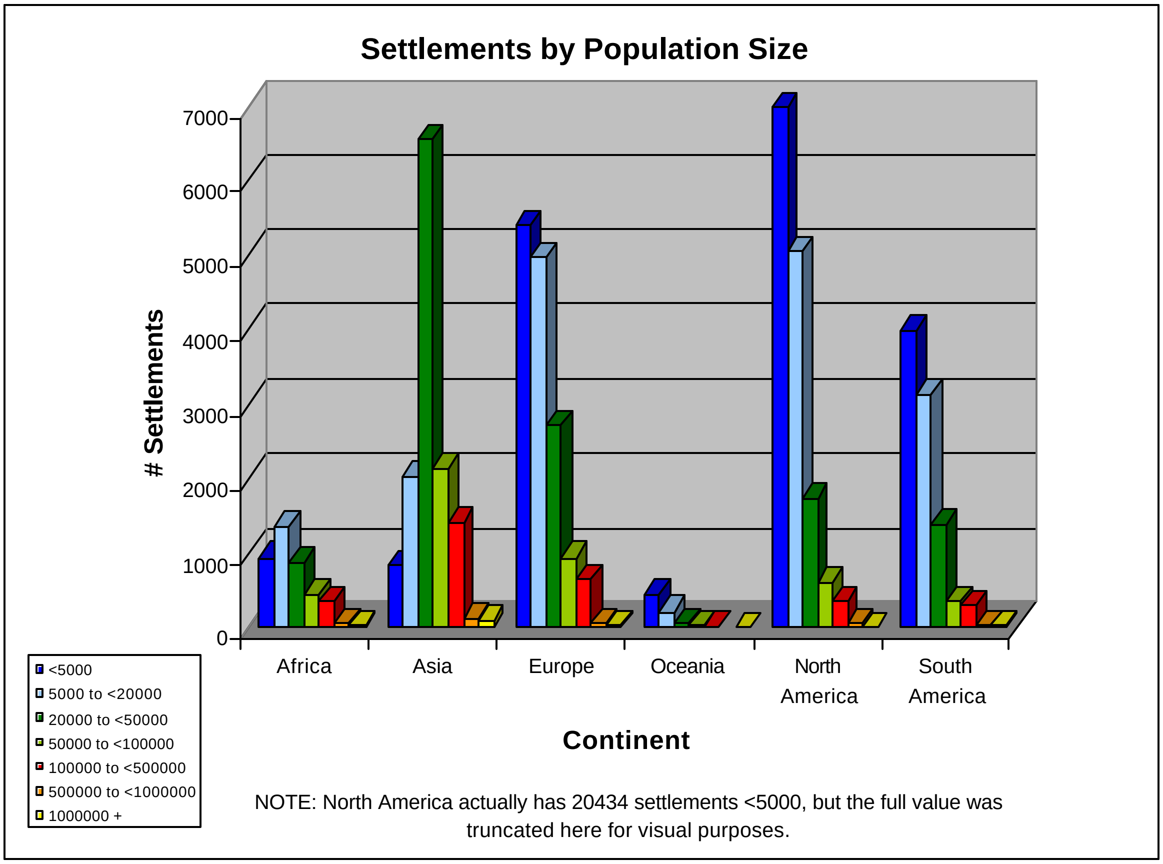

Figure 5. Settlements distribution by continent and population size.

Table 3 shows the distribution of the world's population by city size classes, with associated population densities. While there is general agreement in the population totals, and overall proportions urban, GRUMP estimates that 6.7 rather than 3.7 of the urban dwellers reside in the world's largest megacities. It estimates close to 24,000 urban areas of 5,000 persons or more, in the year 2000. And, cities that are home to 500,000 to 1 million persons are the most dense. There are three plausible, and as of yet untested, explanations for this: Perhaps this is because most countries have at least one such sized city. Alternatively, cities of this size may be more rapidly expanding—or experiencing population growth relative to the land area. Or, it may simply be that if these cities are predominantly in less light areas, that this is simply an artifact of the underlying data constraints.

Distribution of the world's population by size class of settlement, 2000						
	UN	**GRUMP**			**UN**	**GRUMP**
Size class of urban settlement (number of inhabitants)	**Total Population (000s)**		***Number of settlements***	***Population Density***	**% of Total**	**Population**
Total	6,057,000	6,048,511	24,173	***46***		
Urban area	2,862,000	2,795,840	24,173	762	47.3	46.2
10 million or more	225,000	407,327	22	2162	3.7	6.7
5 million to 10 million	169,000	271,911	40	1556	2.8	4.5
1 million to 5 million	675,000	713,650	347	1213	11.1	11.8
500,000 to 1 million	290,000	278,149	393	824	4.8	4.6
under 500,000	1,503,000	1,124,803	23,371		24.8	18.6
100,000 to 500,000		*566,353*	*2778*	702		*9.4*
50,000 to 100,000		*221,023*	*3176*	517		*3.7*
20,000 to 500,000		*228,572*	*7,291*	417		*3.8*
5,000 to 20,000		*108,855*	*10,126*	182		*1.8*
Rural area	3,195,000	3,252,671		26	52.7	53.8

Table 3. Distribution of the world's population by size class of settlements, in 2000, showing the UN numbers and the GRUMP output numbers. Note that the UN classification of under 500,000 persons is subdivided here.

Table 4 **shows the power of these new data, when integrated with other geographic data. Here the urban extent mask and the gridded population surface are overlaid with ecosystem boundaries from the Millennium Ecosystem Assessment (ref). This table, prepared for an assessment of urban systems (McGranahan et al., forthcoming) shows that coastal and island systems tend to be the most densely populated, followed by systems with water and other agricultural resources—namely, cultivated and inland water systems—but that in coastal areas, land area is disproportionately urban. Two systems—coastal and cultivated—also sustain high rural population densities. Forested and mountain ecosystems, which sustain the same total population as coastal ecosystems are much less urban, and thus sustain much lower population densities, even its urban areas.**

These data suggest that roughly 3% of the earth's surface is occupied by urban areas, the majority of which concentrated in coastal and cultivated environments. This is somewhat greater than the oft-cited suggestion of 1-2% of land area.

	Population Estimates (000s)				Population Density			Land Areas (square kilometers)			
Ecosystem	Total	Urban	Rural	% Urban	Overall	Urban	Rural	Total	Urban	Rural	% Urban
Global											
Coastal zone	1,145,552	733,655	411,897	64.0%	175	1105	70	6,537,248	664,139	5,873,109	10.2%
Cultivated	4,221,700	1,894,164	2,327,536	44.9%	119	786	70	35,474,755	2,408,775	33,065,980	6.8%
Dryland	2,150,739	964,452	1,186,287	44.8%	36	750	20	59,966,189	1,286,589	58,679,600	2.1%
Forest	1,131,528	395,622	735,906	35.0%	27	472	18	42,091,243	838,773	41,252,470	2.0%
Inland Water	1,501,769	770,019	731,750	51.3%	51	817	26	29,418,837	942,607	28,476,230	3.2%
Island	633,622	363,895	269,727	57.4%	84	1020	37	7,579,706	356,631	7,223,075	4.7%
Mountain	1,161,014	347,735	813,279	30.0%	36	634	26	32,066,433	548,373	31,518,060	1.7%
Overall	6,048,511	2,795,840	3,252,671	46.2%	46	762	26	130,637,900	3,668,694	126,969,200	2.8%

Table 4. Population estimates by ecosystem (as defined by the Millennium Ecosystem Assessment).

CONCLUSIONS

[Continue.]

One of the main objectives of this project was not only to construct an improved population grid that systematically accounts for urban centers, but also to construct a globally consistent database of those urban areas. As several studies show, there have been several attempts to map or model population distribution, but few of them account explicitly for urban areas, or attempt to merge the lights directly with city-level census data to derive population estimates of urban areas. The methodologies detailed here take a comprehensive and systematic approach to combine several data streams into estimates of urban extents and population distribution.

If we look at the three separate products of the Global Rural Urban Mapping Project (GRUMP) we can draw the following conclusions.

Data gathered from the census or census-like sources seems to provide considerably more detailed information about population distribution for settlements under 500,000 people than the UN estimates. The Population Division actually collects information for smaller cities, but does not do so systematically. The Statistics Division collects some—less systematic—information for places of 100,000 or more, but neither they nor the Pop Division attempt to collect data below 100,000 persons. We show here that nearly 20% of the world's urban population lives in cities of these sizes, thus GRUMP substantially contributes by amassing these data.

By having both population numbers and spatial extents available globally, we can … Spatial information about population distribution by settlement size (See Table 3) is useful to assess …

Thus, the GRUMP Urban Extent Mask attempts to somewhat crudely represent the extents associated to the human settlements

[Be sure to indicate the names and URLs of all three new datasets, and that they are freely available.]

REFERENCES

Balk, D. and Yetman, G. 2004. "The Global Distribution Of Population: Evaluating The Gains In Resolution Refinement", Draft Documentation for GPW v3, downloadable from http://beta.sedac.ciesin.columbia.edu/gpw.

Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT), (2004). *Gridded Population of the World (GPW), Version 3*. Palisades, NY: Columbia University. Available at http://beta.sedac.ciesin.columbia.edu/gpw.

Champion, Anothony and Graeme Hugo (eds). 2003. *New Forms of Urbanisation: Beyond the Urban-Rural Dichotomy* IUSSP [Get complete reference]

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

Dobson, J. E., E. A. Bright, P. R. Coleman, R.C. Durfee, B. A. Worley (2000), LandScan: A Global Population Database for Estimating Populations at Risk, *Photogrammetric Engineering and Remote Sensing*. 66(7): 849-857.

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.

Elvidge, C.D., Safran, J., Nelson, I.L., Tuttle, B.T., Hobson, V.R., Baugh, K.E., Dietz, J.B., Erwin, E.H., 2004, Area and position accuracy of DMSP nighttime lights data. Chapter 20 in Remote Sensing and GIS Accuracy Assessment (Lunetta R.S. and Lyon, J.G. editors), CRC Press, p. 281-292.

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.

Imhoff, M.L., W.T. Lawrence, D.C. Stutzer, and C.D. Elvidge, (1997). A technique for using composite DMSP/OLS "city lights" satellite data to map urban area, *Remote Sensing of Environment*, *61* (3), 361-370.

McGrahanan et al., 2004 (draft). "Urban Systems," Millennium Ecosystem Assessment, Trends and Conditions Assessment.

S. Nghiem, D. Balk, C. Small., U. Deichmann, A. Wannebo, R. Blom, P. Sutton, G. Yetman, R. Chen, E. Rodriguez, B. Houshmand and G. Naumann. Nov. 2001, "Global Infrastructure Mapping: The Potential for SRTM to Break New Ground", JPL White Paper, available at: http://www.ciesin.columbia.edu/whitepaperfinal.pdf.

Pozzi, F., C. Small and G. Yetman (2003). Modeling the Distribution of Human Population with Night-time Satellite Imagery and Gridded Population of the World. *Earth Observation Magazine, Vol. 12, No. 4:* 24-30

Schneider, A., M.A. Friedl, D.K. McIver, and C.E. Woodcock (2003). Mapping Urban Areas by Fusing Multiple Sources of Coarse Resolution Remotely Sensed Data. *Photogrammetric Engineering and Remote Sensing*, *69* (12):1377-1386.

Sutton, P., D. Roberts, C. Elvidge, and K. Baugh, (2001). Census from Heaven: An estimate of the global human population using night-time satellite imagery, *International Journal of Remote Sensing*, *22* (16), 3061-3076.

Tobler, Waldo, Uwe Deichmann, Jon Gottsegen and Kelly Maloy. 1997. "World Population in a Grid of Spherical Quadrilaterals," *International Journal of Population Geography*, 3:203-225.

United Nations, 2002. *World Urbanization Prospects, The 2001 Revision*. United Nations: Department of Economic and Social Affairs, Population Division, New York.

United Nations, 200x. *Demographic Yearbook*. United Nations: New York.

Woods, S., K. Sebastian, F. Nachtergaele, D. Nielson, and A. Dai, 1999 "Spatial Aspects of the Design of Targeting of Agricultural Development Strategies," Environmental and Production Technology Discussion Paper, #44, International Food Policy Research Institute, Washington, DC: May.

Woods, Robert. 2003. Urban-Rural Mortality Differentials: An Unresolved Debate. *Population and Development Review*, March 2003, *29*(1), pp. 29-46.