Know Your Boundaries:
Using spatial information to reduce uncertainty in urban population estimates

Deborah Balk*
Mark Montgomery
Thomas Buettner
Valentina Mara
Donghwan Kim
Megan Todd

Draft date: 29 September 2009

Paper to be presented at the International Union for the Scientific Study of Population Conference, Marrakech, Morocco, 27 Sept – 2 Oct, 2009.

* Corresponding Author: Deborah Balk, CUNY Institute for Demographic Research (CIDR) and Baruch College, City University of New York, email: deborah.balk@baruch.cuny.edu

Institutional affiliations: Deborah Balk, Baruch College, City University of New York (CUNY) and CUNY Institute for Demographic Research (CIDR); Mark Montgomery, Population Council and Stony Brook University; Thomas Buettner, United Nations Population Division; Valentina Mara, Center for International Earth Science Information Network (CIESIN), Columbia University; Donghwan Kim, Stony Brook University; and Megan Todd, (CIDR).

The work was funded in part by the U.S. National Institutes of Child Health and Development award R21 HD054846 to the City University of New York, the Population Council, and Columbia University.

I. Overview

Future population growth will be markedly different from the past: It will take place predominantly in urban areas, and mostly in the cities of Asia and Africa (United Nations, 2006). It is likely to take place in smaller and medium-size cities (Montgomery, 2008; United Nations, 1998). It will face new challenges from a changing environment (NRC, 2003; McGranahan et al 2007, 2005). Although current estimates and projections could establish these basic findings, they are by no means beyond criticism. The major demographic shortcomings of urban population estimates and projections have been articulated (NRC, 2003; Bocquier 2005).[1] But it is the lack of spatial information that has not yet been recognized as a fundamental omission from the record. Although the spatial dimension is of obvious importance to planners and policymakers, information on spatial boundaries is not systematically reported by national statistical offices, and therefore are not included in the data records held by the United Nations (UN) Population Division, which compiles country-level information into country, continental and global estimates and projections of urban population. Using the UN urban population estimates as the baseline, this paper compares them with those derived from the Global Rural Urban Mapping Project (GRUMP; Balk et al. 2005) in an explicit spatial framework.

Understanding future urban population change depends critically on a solid understanding of the estimated size (and composition) of the population as well as the factors that cause growth. But while the importance of a spatially explicit analysis of demographic and other dynamics is undisputed, the empirical basis for it is untested especially in developing countries. Weak statistical systems in many developing countries are not able to generate empirical data that are updated regularly, are detailed, and are spatially explicit. While most countries now undertake a decennial census, projections of future population suffer from the fact that about two-thirds of the world's population lives in countries that do not have sufficient vital registration system to record births and deaths (Mahapatra et al. 2007). While the availability of vital registration data has not much changed over the last four decades, nationally representative surveys are increasingly filling the gap (Boerma and Stansfield 2007). These surveys generate important data streams, but are limited in their spatial resolution. Much of what is established about the demographic dynamics in developing countries, therefore, rests on evidence generated by surveys and indirect estimation techniques used by demographers (Hill et al. 2007) and speaks only to national level estimates. Expanding the estimation into the sub-national and urban settlement dimension remains a formidable challenge.

Yet there is reason to be optimistic. Other scientific disciplines, such as geography, economics and the physical sciences have not only generated more theory on the subject of urban dynamics, but have also unveiled new sources of empirical evidence. For example, by adding location to city population size, it is possible to treat isolated cities differently from cities in such close proximity that they seem poised to fuse into larger agglomerations, and to differentiate coastal cities, which have considerably higher population densities (McGranahan et al., 2005), from those inland. New data streams and computing capacities and tools also would allow the consideration of natural barriers that can hinder or redirect city growth.

Spatial correspondence of boundaries and population would help to define cities. National definitions of "urban" vary a great deal, as very different concepts are employed to define what qualifies an area or settlement as urban. In preparing estimates and projections of city populations, the United Nations relies on data supplied by national sources that incorporate the definitions of urban areas adopted by national authorities. Much heterogeneity remains across, and even within, countries in the criteria applied to establish the geographic boundaries of these cities (NRC, 2003). Not able to reconcile the many different concepts used by national statistical offices, the UN Population Division decided to accept the heterogeneity. The results of this analysis are published every two years in *World Urbanization Prospects*, the most recent edition of

[1] The United Nations Population Division, who produces these estimates and projections, are among those who are critical of the current methods.

which is United Nations (2008). These authoritative volumes present population estimates and forecasts for regular five-year intervals, from 1950-2025, for urban agglomerations of 750,000 persons and above, and for all national capitals irrespective of size.

Even strenuous efforts on the part of the Population Division have not succeeded in producing strict comparability in the units for which city populations are recorded, and far more heterogeneity remains in the UN's final figures than is commonly realized. The Population Division often cannot reconcile the data reported to the UN with its preferred agglomeration concept, and for these cases the *World Urbanization Prospects* estimates represent the size of the city proper or metropolitan area instead. Another limitation is that the Population Division publishes city-specific estimates and projections only for cities with a population of 750,000 and above at the base year; smaller cities are taken into account but their population estimates are not published. This treatment of smaller cities is unfortunate, given the United Nation's forecast that over the next few decades, roughly half of the urban population of developing countries will be found in cities of 500,000 and fewer inhabitants (United Nations, 1998). In *World Urbanization Prospects,* the Population Division takes care to spell out the limitations of its work, and emphasizes how often the estimation and projection units depart from the ideal agglomeration concept, but as the UN figures make their way into multiple international population and development databases, these warnings and caveats are generally left behind.

This paper draws upon a database compiled by a team of investigators – at the City University of New York, the Population Council, Columbia University's Socioeconomic Data and Applications Center (SEDAC) and the United Nations Population Division – to review and revise the methods of urban population projections using new spatial and demographic data for urban areas. This paper will focus only on a small part of the project: namely, how accurate is the baseline – the estimates upon which population forecasts are made – and what are the bounds of uncertainty in spatial terms. It will use spatial data to analyze the UN's City Database. We evaluate the population estimates along with the system of classification used by the UN (city proper, metropolitan area or agglomeration) against external data and discuss the implications for current estimates. We conclude with recommendation for future data collection and methodological development.

2. New Spatial Data to Accompany Population Estimates

This project has compiled data from two large data collections: the UN Population Division's City Database (which forms the basis of the estimates in the *World Urbanization Prospects*) and SEDAC's GRUMP. The former has time series data, the latter does not. The latter is spatial the former is not.

The City Database contains both administrative and demographic information on a total of 3,822 cities, of which 524 have populations of 750,000 or more, compiled from 1950 to the present. About two-thirds of all the cities in the database are in the less developed world, with close to half of the cities in Asia. There are only seven cities per country in Africa, on average, compared to 34 cities per country in Asia. Over the 1950 to 2000 period, the average city in Asia has had 560,000 inhabitants versus 307,000 in Africa. In the less developed regions, the database contains (on average) observations for about 28 years. The smaller cities in the UN's city database (those under 750,000 population) have not generally been subjected to the rigorous scrutiny applied by the Population Division to larger cities. Entries for these small cities have been maintained in the database, but changes in city definitions over time are common and relatively few of the discrepancies have been resolved. Apart from national capitals, the cities and towns under 100,000 population appear nowhere in the Population Division and Statistical Office's published databases, and yet these urban places are generally numerous in most countries and, taken collectively, account for a generous portion of the total urban population (perhaps as much as 20% according to GRUMP).

The GRUMP Database is a global collection of population estimates for human settlements rendered as point locations, urban extents, and a population surface (i.e., a global grid of population) derived from the points, extents, and administrative boundary data (Balk, 2009; Balk et al. 2005). The collection includes population estimates for points as far back as 1950, but for most places only estimates around 1990 and 2000 are available. GRUMP stands apart from its predecessors most notably by using remotely-sensed satellite imagery, in particular NOAA's night-time lights data (Elvidge et al. 1997), as a measure of the extent of each urban area. There are some shortcomings with these data for this purpose (Balk et al., 2005)[2] but they provide the only systematic global assessment of urban areas. A primary objective of the GRUMP project was to associate each urban extent (as derived from the lights) with a place name, name(s) of administrative area(s), and a population estimate. Like the UN's City Database, GRUMP is probably weakest in Africa where the data inputs are weakest; this affects night-time lights data, human settlements data, and administrative boundary data. (The magnitude of this deficit is unknown but cross-validation with the UN's City Database should help in further assessing this shortcoming.)

Using the most recent version of the UN's City Database (United Nations 2008), the research team has extended the reach of the UN data to include hundreds of additional observations on small cities and towns, which were collected in the course of the 2008 update of SEDAC's Global Rural-Urban Mapping Project (CIESIN et al ,2008). The GRUMP project has been organized from the outset on an explicitly spatial basis, and the spatial resources made available through GRUMP allow each urban settlement in the combined UN–GRUMP dataset to be located spatially. Each settlement is attached to latitude and longitude coordinates approximating its presumed[3] geographic center (or "centroid"), then these points are overlaid with estimates of the spatial extent of the urban agglomeration from GRUMP and the administrative boundaries associated with those urban areas. The addition of small cities and towns fills a long-standing gap in the UN's collection of urban materials, and the spatial frame into which these data are now fitted is entirely new.

To understand the nature of the combined UN–GRUMP dataset, consider the case of Banjul, capital of the Gambia, depicted in Figure 1 in a 1993 snapshot of these data. The administrative units into which the region near Banjul is formally divided (the administrative boundaries are indicated in dark lines) include Banjul proper, as well as neighboring areas (the Kinifing Municipal Council and the Western, Kombo North administrative units). Across these units, we can see the locations of other small cities and large towns, depicted as dots, each with its population. In the United Nations records, a population count is available for Banjul city proper (whose population was reckoned at 78,583 persons in 1993) and another count is available for the larger urban agglomeration (in which resided some 271,000 persons). GRUMP supplies an alternative night-time lights estimate of the spatial extent and yet another population estimate of the urban agglomeration. In Figure 1, the spatial extent of the agglomeration is depicted in yellow shading. Using the population data for the administrative areas in which lights are observed, and employing assumptions about population density within the lighted areas, GRUMP provides an alternative estimate of the population of the Banjul urban agglomeration, putting the total at nearly 385,000 persons in 1995 (the year closest to the 1993 estimates shown elsewhere in the map, and the year corresponding to census dates)[4].

---

[2] Two noteworthy issues (detailed in Balk et al., 2005) are (1) the night-time lights may overestimate the areal extent of urban localities; and (2) satellite "footprints" still need to be associated with names and population estimates (Balk, 2009). GRUMP uses one method, but surely there are many that could have been implemented.

[3] By the data collectors. In reality, much variation and ambiguity exists in the rendering spatial features that cover an area – such as a city – as a point location. Even when no ambiguity exists on the part of the data collectors, that information is rarely passed along to downstream data users.

[4] In 2000, the estimate for the Banjul urban agglomeration from GRUMP, depicted in yellow, was just over 495,000 persons. Figure 1 supplies the 2000 estimates, whereas Table 1 supplies GRUMP's estimates for 1990, 1995 and 2000.

In Figure 2 we depict the time-series dimension of the dataset, using only the UN population records for Banjul. The population of this city has generally been reported to the UN in terms of the urban agglomeration, although for one report in the early 1960s the unit is not now known. Other cities in the dataset exhibit similar features, with not uncommon changes in the reporting unit and population counts that are irregularly spaced in time.

Another example is the city of Beijing. Similar to other developing countries, China has made frequent changes to its administrative boundaries and accompanying urban definitions, a practice that has sown confusion among the experts struggling to understand this country's urban trends (Chan and Hu, 2003). Even at a single point in time, it can be difficult to grasp how boundaries are implicated in Chinese city definitions. To show how spatial data help clarify matters, we present in Table 2 the basic population counts for the administrative units that make up Beijing province, and accompany these conventional tabular data with the maps shown in Figure 3. These tables are inputs to both the UN City Database and to GRUMP.

In 2000, according to the urban definitions most recently adopted by the Chinese government, the *de facto* population of Beijing was reported as 11.5 million persons. But to whom, precisely, does the label "Beijing resident" apply? The 11.5 million total (shown as subtotal B in Table 2) was derived by adding the populations of the administrative districts in the city proper (the 8.5 million people of subtotal A, who live in the districts depicted in beige) to the full populations of neighboring "city districts" (another 3.0 million persons, in the areas surrounding the city proper, whose outer boundaries are depicted in brown). However, these city districts include substantial number of rural residents. If we were to depart from the official definition and count only the urban residents of the city districts toward the total, this redefinition would reduce the population of Beijing from 11.5 million to about 9.9 million (subtotal C of the table). This is exactly what was done in the UN City Database and in GRUMP. Though these two independent estimates concur, they disagree with the estimate supplied by the Chinese government.

Going further afield, the entity of Beijing might be defined to include those urban localities in the outlying counties of the province (counties are depicted in blue), on the theory that the smaller cities participate in networks of transport, communication, and services that link them to Beijing proper and to other parts of the city. These adjustments would produce an estimated total population for Beijing of 10.5 million persons (subtotal E of Table 2). Yet in the absence of spatial data, it is impossible to know where these population centers are found: the lighter blue shading indicates these localities identified by GRUMP. It is noteworthy that one of these areas (in Daxing Xian) is contiguous to the Beijing agglomeration.

Clearly, it would be difficult even to weigh the merits of these alternative definitions without reference to maps and other spatially-coded data. These are just two of many comparisons that can be made to illustrate the challenges associated with estimating urban population.

3. Analysis

This paper will amass the illustrative examples shown here into a systematic analysis that compares the UN City Database with GRUMP and administrative data for cities in Africa, Asia, and Latin America. Administrative data are particularly useful because they have received the bulk of investment in the past 50 years, but administrative areas may radically differ conceptually from cities. Much greater scrutiny will be needed to show how the urban agglomeration concept compares to administrative boundaries and GRUMP urban extents. Similarly, "city proper" definitions may indeed correspond closely with the smallest administrative unit associated with a place, or not at all (if, for example, a "city proper" is a much smaller geographic area than the administrative unit to which it is assigned). The analysis does not presume that one set of estimates (those in the City Database or GRUMP) is more accurate than another: rather it will look for correspondence and divergence between the two estimates, along with population estimates of associated administrative areas.

One main difference between the UN City Database and the GRUMP data is the total number of cities: For Asia, Africa and Latin America only, GRUMP has more than six times as many cities as the UN collection (Table 3). The bulk of the additional cities are those with a population of less than 100,000 which the UN does not formally include in its collection. But even among places with more than 100,000 persons, GRUMP identifies one-third more localities. The evaluation herein uses the UN City Database as our base, but it is worth noting that that record is incomplete, especially where smaller cities are concerned. Using India as an example, the UN database contains about 270 cities (indicated in red), but there were a total of about 2,900 cities as identified by GRUMP (indicated in blue), as shown in Figure 4. At present, it is unclear how these GRUMP-only urban areas are treated by the UN's estimation of India's urban population. Surely this varies on a country-to-country basis. How does this basic distribution translate into differences in population estimation? Table 4 suggests that using GRUMP estimates yields at least 300 million more urban dwellers in the year 2000 than UN estimates. Although the bulk of GRUMP's additional cities are found in the small cities, the difference in population is greatest (in total and proportion) for the large cities. The implications of these differences have not been fully explored, in part because the UN has no counterpart data for the locations that are in GRUMP but not its City Database. The rest of this analysis is based only locations where we have matched the UN and GRUMP localities.

We difference the population estimates (for year 2000, other IUSSP paper by Balk and colleagues) of the UN City Database and those in GRUMP. We use two GRUMP data streams for comparison here. (We have not yet similarly quantified the administrative data.) (1) The first is the raw GRUMP point settlement input. We call this GRUMP Sum of Points, herein. These point data contain latitude, longitude and population values. Unlike administrative boundary data and their associated population characteristics, which tend to be collected and distributed by national statistical offices (NSO), point data come from a much wider variety of sources, including gazetteers and secondary sources of data that typically undergo far less scrutiny than NSOs. These data are essential because they contain place name and coordinates; the satellite data upon which the GRUMP extents are based do not have names, so it is important to have a transparent method for assigning names to the urban footprints. The population values of these points are summed to render an initial estimate of population for each urban extent. (2) The second data stream is GRUMP's modeled population of urban extents (herein, "model" or "modeled population") which uses the sum of points for each administrative areas as well as total population values for the administrative areas in which each extent falls, to insure that the point totals do not exceed the administrative totals. Furthermore, because the footprints are based on night-time lights data, which represent an agglomeration concept, an additional algorithm is used to redistribute the population from the administrative area to the urban one. (See Balk, 2009 for details.)

We identify the places where the UN and GRUMP estimates match, and where (and how) they diverge. *A priori*, we anticipated greatest agreement on the largest cities – since they tend to have the finest detail in terms of inputs and have already been the subject of considerable scrutiny – but these localities may also have disproportionately large satellite footprints, which may lead to GRUMP agglomerations larger than those found in the City Database. We further expected to find less agreement where smaller cities are concerned, and anticipated that small and medium cites would tend to have much lower population estimates in the UN City Database than in GRUMP. Overall, Figure 5 shows how well the UN City Population sizes match with GRUMP Sum of Points and GRUMP Model. The majority of UN Cities have population estimates for the year 2000 that are within 10% of the GRUMP estimates when the Sum of Points measure is used, but far fewer cities agree that closely when the GRUMP model population is used (less than 30%). Figure 6 shows the distribution of these differences by city size. The most notable discrepancy – population differences of more than 100% – is found with the largest cities, not the smallest ones, per our priors. These differences are particularly exacerbated when comparing the UN city populations with the GRUMP modeled population. This is perhaps because GRUMP extents represent a generous footprint for urban areas, corresponding to agglomerations of larger areas.

Might some of this difference be due to the statistical concept used to define the population of each city? In Figure 7, we look at statistical concept by city size. The top panel of Figure 7 shows UN cities that are the only UN city included in a given GRUMP extent. The bottom panel includes only UN cities for which more than one UN city fall within a given GRUMP extent. (One such example is Buenos Aires and La Plata, Argentina, as seen in Figure 8, discussed in greater detail below). Statistical concepts vary both by city size and whether there is more than one UN city that belongs to the same urban extent. Among UN cities that are the only UN city in a GRUMP extent, it is clear that city proper is the most widely used classification for smaller cities ant that it is infrequently used for the largest cities. Large cities, in contrast, have a greater share of UN cities with unknown definitions. Among these cities, the use of the agglomeration concept increases with city size, as one might expect. The patterns are less striking in the lower panel of Figure 7. Among the many-to-one matched cities – those UN cities that share a GRUMP extent with at least one other UN city – use of the city proper classification also declines as city size increases, but it accounts for more than half of classifications regardless of city size. These cities are on average larger than the one-to-one matches, and the parts of the urban area that comprise it may make for a more complex arrangement of city types. (Further investigation is warranted.)

In the example of Greater Buenos Aries, Argentina (Figure 8), the UN Cities Database identifies two locations – Buenos Aires and La Plata. The UN reports that both use the agglomeration statistical concept, yet not the single agglomeration that GRUMP suggestions. The majority of UN Cities that share a GRUMP extent do not share statistical concepts. (Future work will describe this in more details.) Future estimates of population growth may benefit from knowledge of city proximity.

Because the basic description varies on many fronts, we describe these differences with OLS regression to control for various data quality characteristics and classification type. In Table 5, we predict the differences between the year 2000 estimate of population for UN city and the GRUMP estimates (both sum of points and modeled estimates), controlling for statistical concept, city size, GRUMP data quality, and continent. (For the time-being, this analysis is restricted to one-to-one matches of the UN cities to GRUMP extents.) First, the R-square values indicate that more of the difference can be explained in the sum of points estimation (somewhat more than 25%) as opposed to the modeled estimates. Second, in both models, differences are greatest when the statistical concept of metropolitan area is used. Third, city size matters: the larger the city, the more likely that the UN estimate of population will differ from GRUMP's. Fourth, the greater the GRUMP data quality, the smaller the difference between GRUMP and UN estimates. Data quality is measured in two ways: the number of points per extent measures the number of settlements that were found in each urban extent footprint. Populous cities tend to have more settlements, as do places with more advanced statistical infrastructure. As these regressions already control for city size, this effect is interpreted as a data quality measure. The other measure of data quality is the number of administrative units per urban extent. Again, more populous and geographically large cities tend to have more sub-city administrative units, but even within city sizes classifications, this tends to vary by the investments countries make into the spatial data infrastructure. But the finer the spatial data, the better the GRUMP model estimation will be. Finer resolution data lead to a closer correspondence between the GRUMP estimates and the UN's. This suggests that improvements in data quality may reduce the uncertainty in population estimation.[5]

4. Concluding Remarks

It has been long understood that spatial data have a role to play in determining urbanization. For

[5] It is possible that there is high correspondence between these estimates and that both are wrong. That possibility will be considered in future work.

example, in Chen et al. (1998) reclassification is lumped with migration in comparison with fertility. But they are sorely under-utilized in practice. This analysis – though preliminary and incomplete -- attempts to demonstrate that spatial tools are viable and central to the demographic understanding of urbanization, and do so in a way that clarifies the long-standing efforts at the United Nations, and in national statistical offices, along the way.

When data holdings are strong at the national statistical offices, the UN and others benefit. The question really is what about when they are not? When there is close correspondence on population estimation, it is still fair game to debate the best way to measure and quantify urban population and the associated land area that people occupy (should city parks be included as urban?). But when correspondence is distant, and it is so because of data quality differences, along with modeling assumptions, it is important to offer a range of population estimates rather than rely on a single set. The future world is an urban one, thus best to improve our ability to describe it in a transparent way for researchers and planners alike.

References

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*. London: Taylor and Francis.

Balk, D., F. Pozzi, G. Yetman, U. Deichmann, and A. Nelson (2005). “The distribution of people and the dimension of place: Methodologies to improve the global estimation of urban extents.” International Society for Photogrammetry and Remote Sensing Proceedings of the Urban Remote Sensing Conference, Tempe, AZ, March 2005.

Bocquier, Philippe. (2005) “World Urbanization Prospects: An alternative to the UN model of projection compatible with mobility transition theory.” *Demographic Research* 12(9).

Boerma, JT and Stansfield, SK (2007), “Health statistics now: are we making the right investments?” *Lancet*. March 3.

Champion, Tony and Graeme Hugo, editors. *New Forms of Urbanization: Beyond the Urban–Rural Dichotomy*. Ashgate, Aldershot, 2004.

Chen, Nancy, Paolo Valente, and Hania 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. United Nations Population Fund (UNFPA), New York.

Center for International Earth Science Information Network (CIESIN), Columbia University; the International Food Policy Research Institute; the World Bank; and the Centro Internacional de Agricultura Tropical (CIAT). 2008 *Global Urban-Rural Mapping Project* (GRUMP). Data available in beta version at http://sedac.ciesin.columbia.edu/gpw/.Socioeconomic Data and Applications Center (SEDAC), Columbia University, 2008.

Elvidge, Christopher D., K. E. Baugh, E. A. Kihn, H. W. Kroehl and E. R. Davis, 1997. Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogrammetric Engineering and Remote Sensing*, 63(6): 727-734.

Hill K, Lopez AD, Shibuya K, Jha P;(2007), “Interim measures for meeting needs for health sector data: births, deaths, and causes of death,” *Lancet*. 2007 Nov 16.

McGrahanan, G.., P. Marcotullio, X. Bai, D. Balk, T. Braga, I. Douglas, T. Elmqvist, W. Rees, D. Satterthwaite, J. Songsore, and H. Zlotnik (2005). “Urban Systems”, Chapter 27 in Conditions and Trends Assessment of the Millennium Ecosystem Assessment, Chicago: Island Press.

McGranahan, G., D. Balk, and B. Anderson (2007), “The Rising Risks of Climate Change and Human Settlements in Low Elevation Coastal Zones,” *Environment and Urbanization*, Vol. 19, No. 1, pp. 17-37.

Mahapatra, G., Shibuya, K., Lopez, AD., Coullare, F., Notzon, F., Rao C., and Szreter, S. (2007), “Civil registration systems and vital statistics: successes and missed opportunities,” *Lancet*. 2007 Oct 26.

Montgomery, Mark R. (2008) “The Urban Transformation of the Developing World,” *Science* 319: 761-764

National Research Council, (2003), *Cities Transformed: Demographic Change and its Implication in the Developing World*, Mark R Montgomery, Richard Stren, Barney Cohen and Holly E Reed (editors). National Academy Press, Washington DC.

United Nations, 2008. *World Urbanization Prospects: The 2007 Revision,* New York: United Nations.

United Nations, 1998. *World Urbanization Prospects: The 1996 Revision*. United Nations, Department of Economic and Social Affairs, Population Division, New York.

Figure 1. Combined UN City Database – GRUMP data for Banjul, the Gambia, c. 1993.

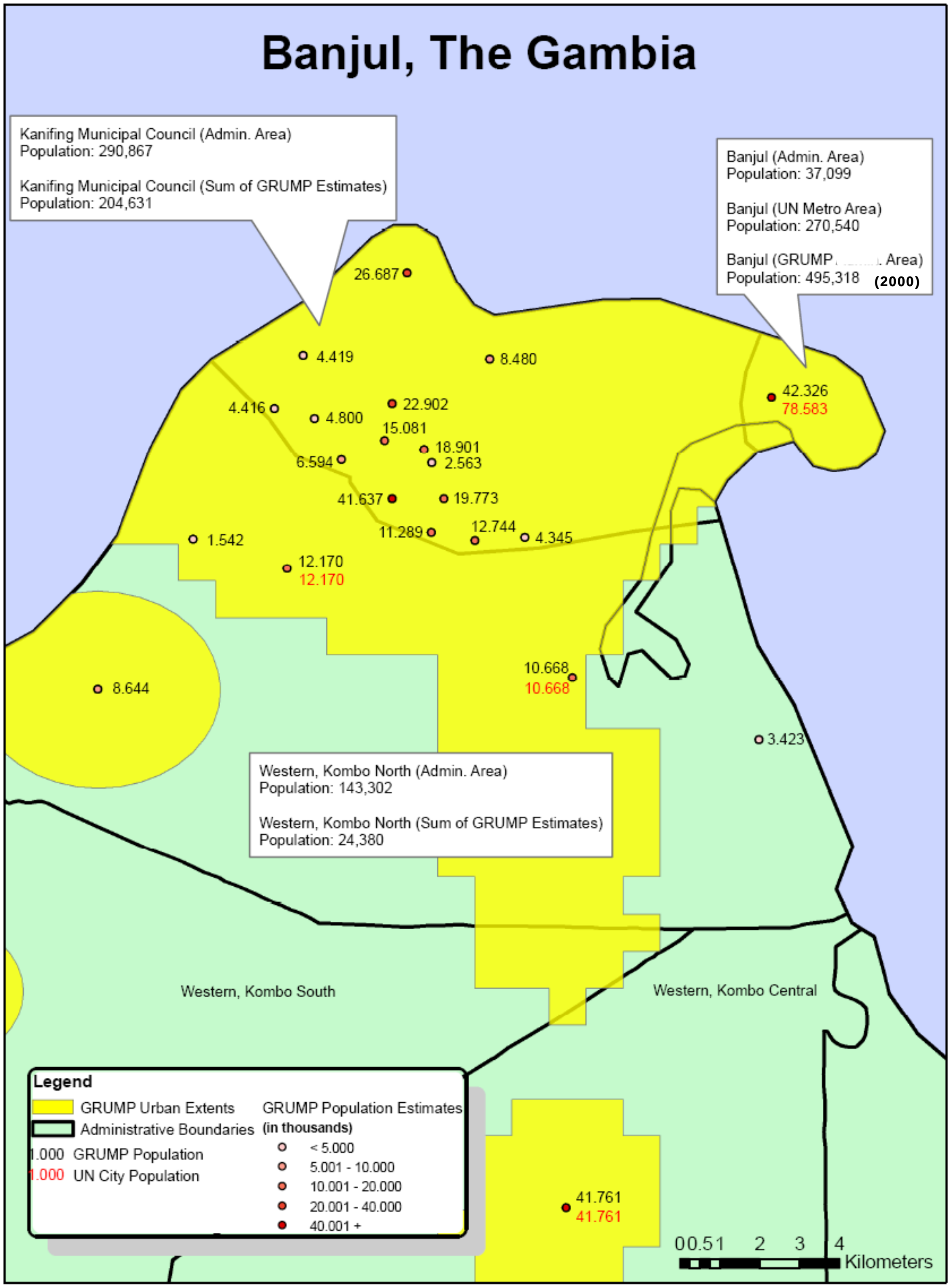


Figure 2. Classification used by the UN's City Database for Banjul, the Gambia, 1950-2000.

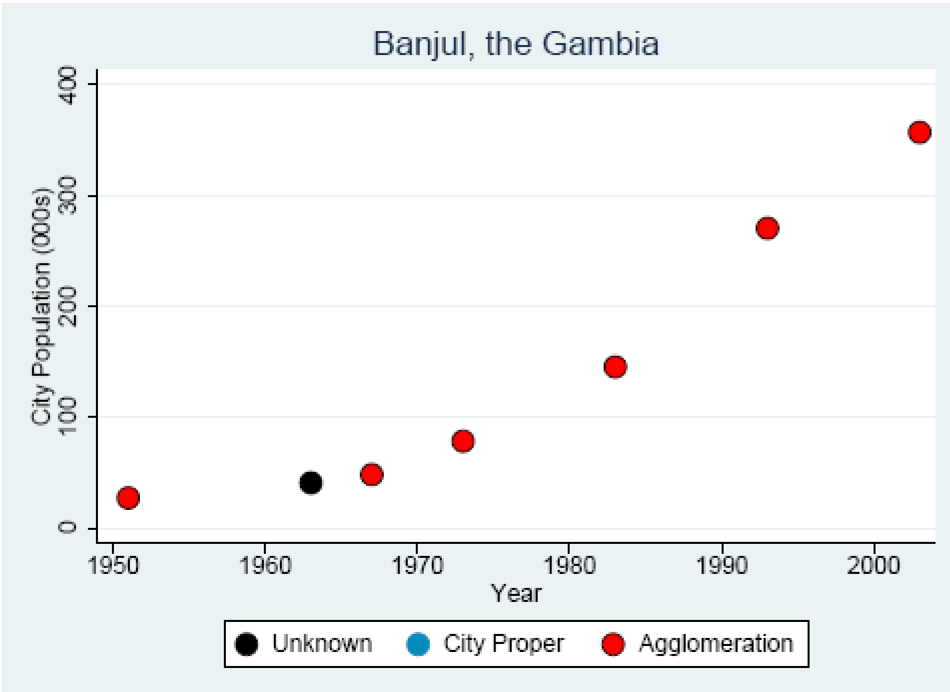


Table 1. Various Population Estimates of Banjul, the Gambia (1993, unless otherwise noted in italics)						
Estimated correspond to points (p) or agglomerations (a)						
	UN		**GRUMP**		**ADMIN**	
Banjul Metro Area	270,540	p*	*384,608*	a**		
BANJUL, Banjul					37,099	a
Banjul			42,326	p		
SUBTOTAL			42,326			
KANIFING MUNICIPAL COUNCIL, Kanifing Municipal Council					290,867	a
Abuko			4,345	p		
Bakau			26,687	p		
Bakoteh			6,594	p		
Bununka Kunda			41,637	p		
Dippa Kunda			15,081	p		
Eboetown			2,563	p		
Faji Kunda			12,744	p		
Kololi			4,416	p		
Kotu			4,419	p		
Latri Kunda			22,902	p		
Latri Kunda Sab			11,289	p		
Manjai Kunda			4,800	p		
Old Jeshwang			8,480	p		
Serre Kunda			18,901	p		
Talinding Kunja			19,773	p		
SUBTOTAL			204,631			
WESTERN, Kombo North					143,302	a
Bijilo			1,542			
Lamin	10,668		10,668			
Sukuta	12,170		12,170			
SUBTOTAL			24,380			
SUM OF GRUMP POINTS (1993)			271,337			
Sum extrapolated to 2000			*410,754*			
Sum interpolated to 1990			*239,502*			

* WUP indicates "agglomeration" as the statistical concept for Banjul.
** Population estimate for 1995 (estimate for 1990 was 306K persons and for 2000 was 495K persons).

Figure 3. Administrative boundaries and GRUMP urban extents, Beijing, China, 2000.

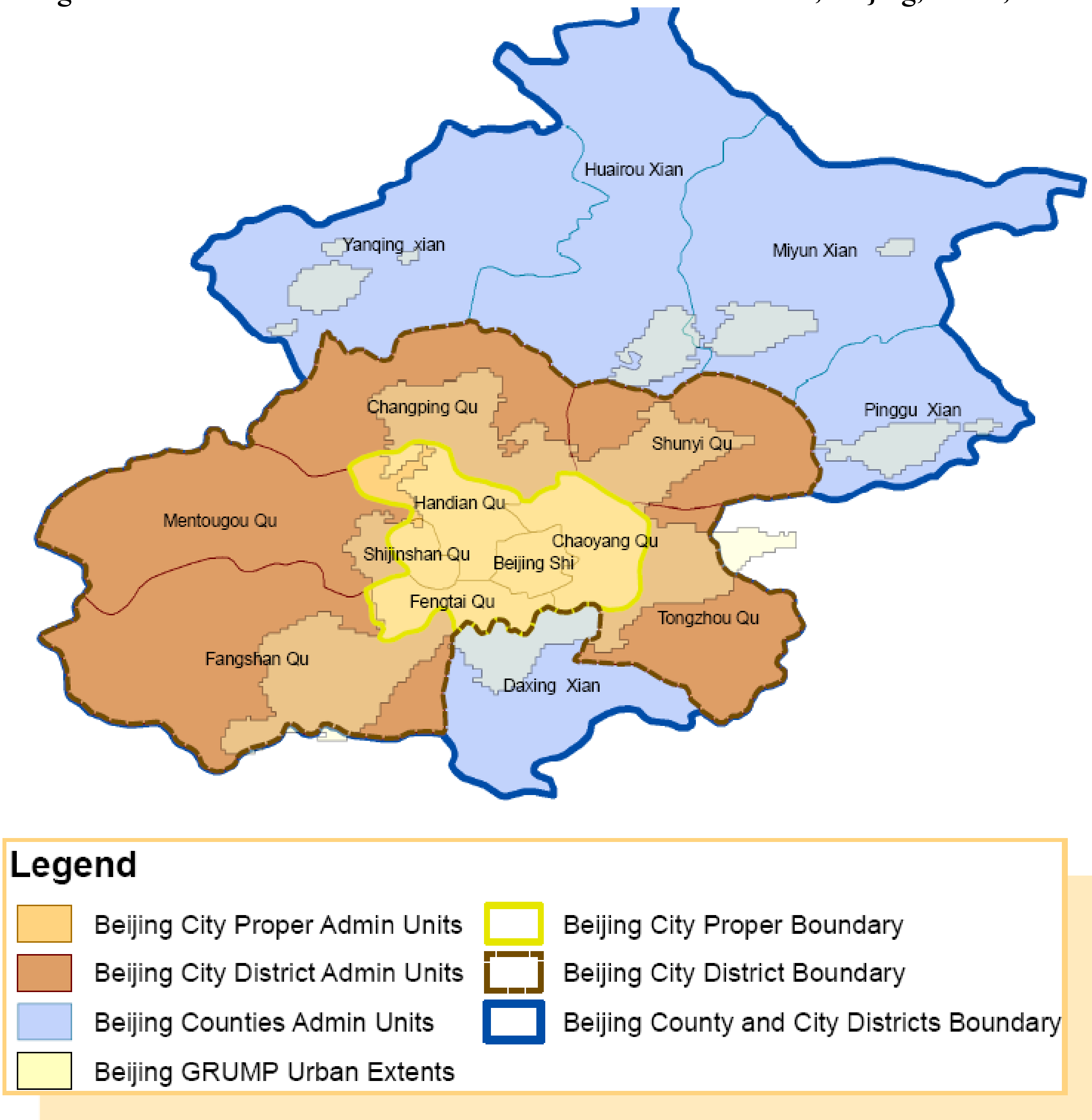


Table 2. Various Population Estimates for Beijing, China, 2000.

Administrative division type and name		Population			
		Total	Urban	Rural	% urban
City proper	Beijing Shi	2,114,586	2,114,586	-	100
City proper	Chaoyang Qu	2,289,756	2,289,756	-	100
City proper	Fengtai Qu	1,369,480	1,369,480	-	100
City proper	Shijinshan Qu	489,439	489,439	-	100
City proper	Handian Qu	2,240,124	2,240,124	-	100
subtotal A		*8,503,385*	*8,503,385*		
City District	Mentougou Qu	266,591	187,616	78,975	70%
City District	Fangshan Qu	814,367	379,882	434,485	47%
City District	Tongzhou Qu	673,952	346,645	327,307	51%
City District	Shunyi Qu	636,479	207,341	429,138	33%
City District	Changping Qu	614,821	251,792	363,029	41%
subtotal B		*11,509,595*			
subtotal C			*9,876,661*		
Beijing County	Daxing Xian	671,444	188,109	483,335	28%
Beijing County	Pinggu Xian	396,701	119,053	277,648	30%
Beijing County	Huairou Xian	296,002	116,900	179,102	39%
Beijing County	Miyun Xian	420,019	128,999	291,020	31%
Beijing County	Yanqing xian	275,433	92,742	182,691	34%
subtotal D		*13,569,194*			
subtotal E			*10,522,464*		

Table 3. Basic Comparison of the Number of Urban Areas in the UN City and GRUMP extents Databases, by city size (c. 2000).

Number of cities in GRUMP extents and UN City Databases

City Size	GRUMP*	UN**
Less Than 100,000 persons	16,299	1,079
100,000 - 500,000 persons	2,028	1,429
500,000 - 1 million persons	264	267
1 million persons +	292	255
Total	18,883	3,030

* City-size classification based on population estimated for 2000;
** City-size classification based on UN value for population for the year closest to 2000. 92% of UN settlements have a population value within 5 years of 2000.

Figure 4. UN Cities, Rendered within GRUMP urban Extents, and additional GRUMP urban Extents, India.

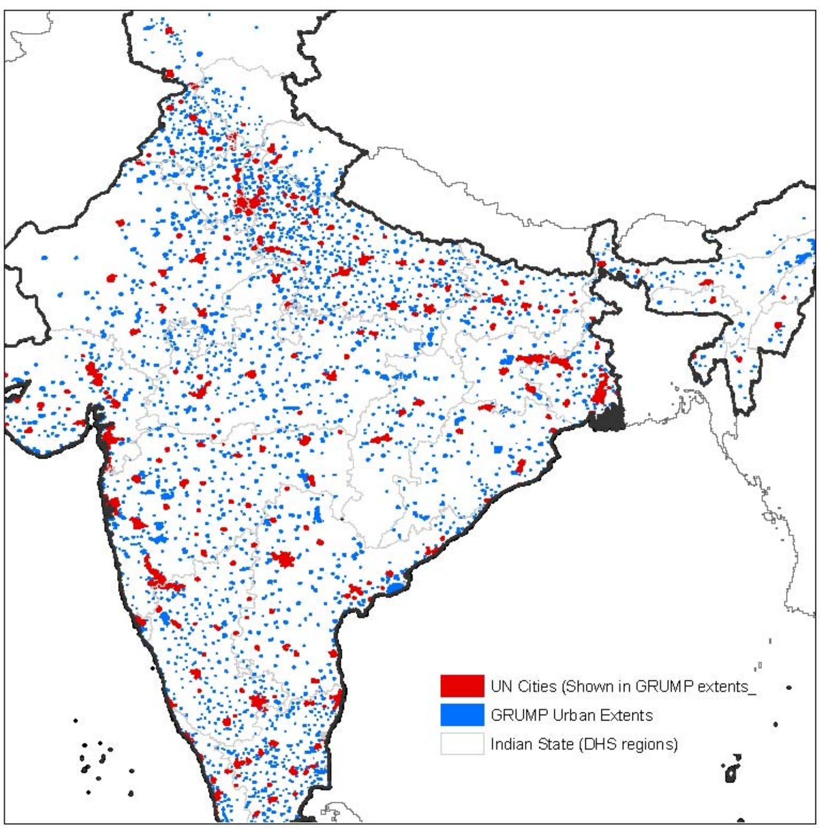


Figure 5. Difference between estimated population (2000) for UN Cities and GRUMP Estimates (GRUMP: Sum of Points (blue)and Model (orange))

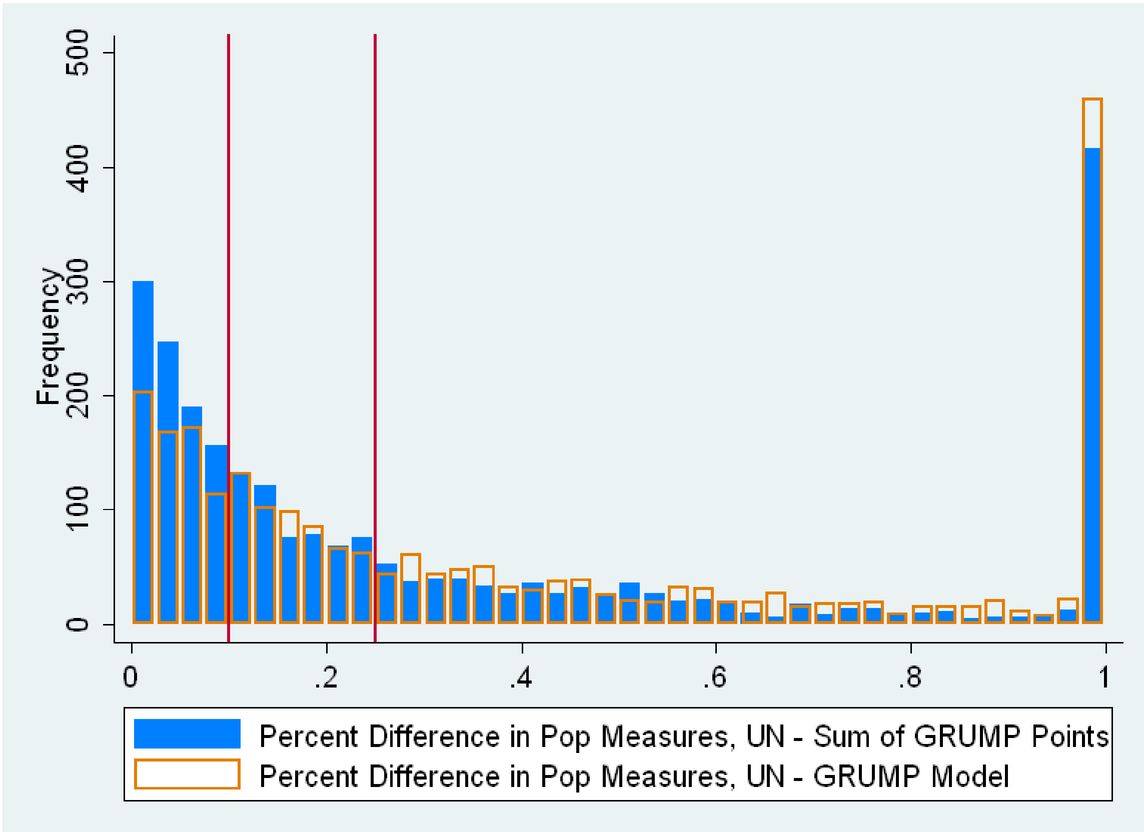


Figure 6. Difference between estimated population (2000) for UN Cities and GRUMP Estimates, by city size (GRUMP: Sum of Points (left) and Model (right))

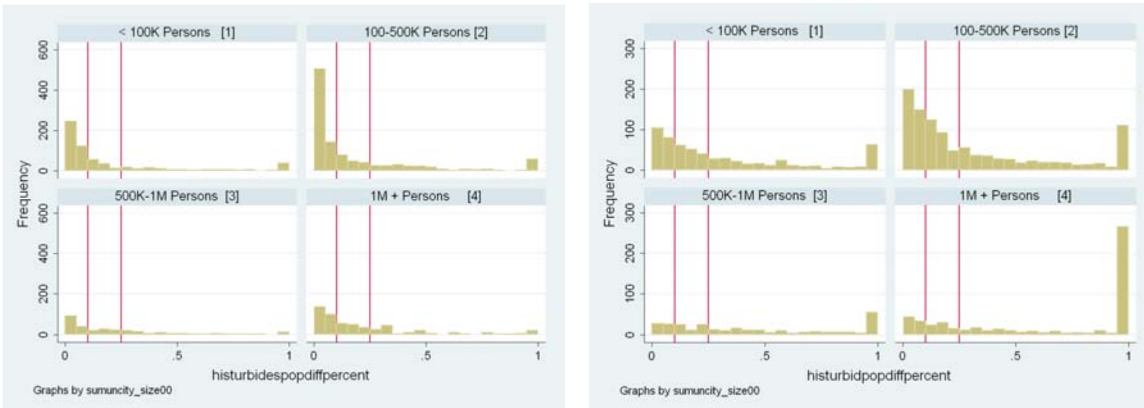
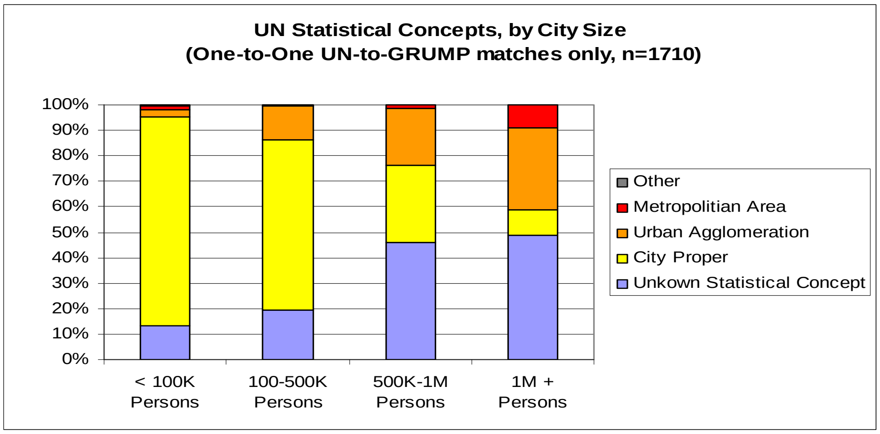


Table 4. Urban Population (000s) according to GRUMP extents and UN City Databases, 2000

City Size	GRUMP*	UN**
Less Than 100,000 persons	33,102	32,243
100,000 - 500,000 persons	233,696	235,976
500,000 - 1 million persons	186,415	194,452
1 million persons +	2,659,752	2,313,755
Total	3,112,964	2,776,426

Figure 7. UN Statistical Concepts by City Size.



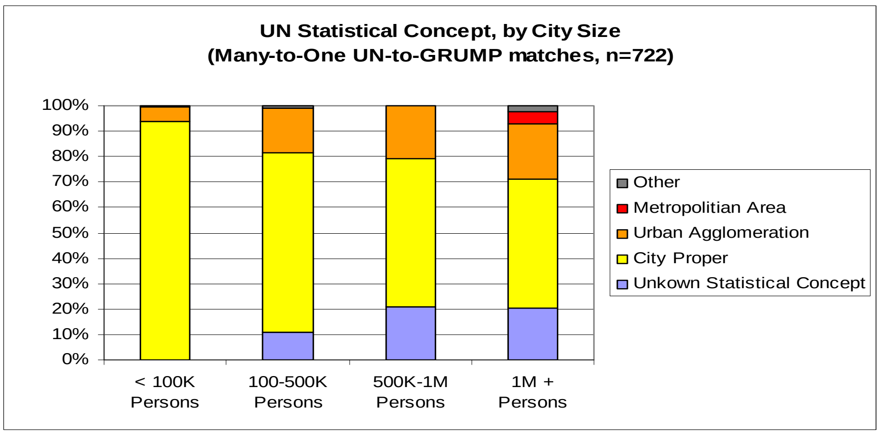
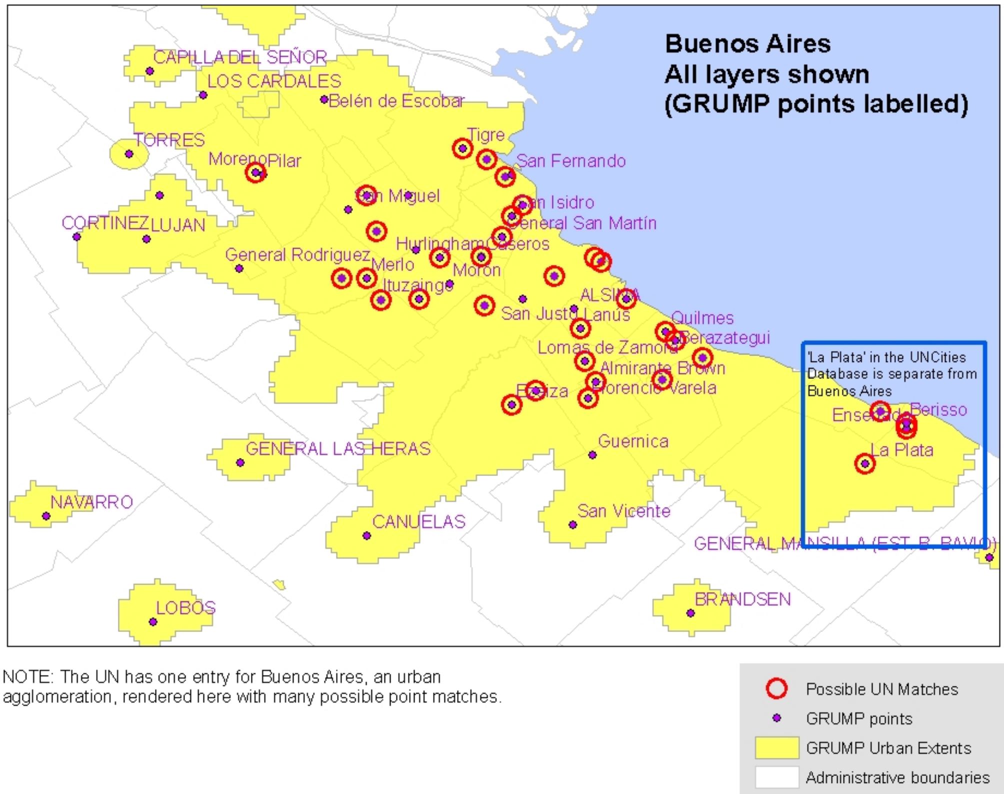


Table 5. Regression estimation of differences in population size (2000) between UN City Database and two GRUMP estimates

	GRUMP: Model			GRUMP: Sum of points	
	Coef.	SE		Coef.	SE
UN Statistical Concept					
Unkown Statistical Concept (reference)					
City Proper	-0.002	(0.048)		0.045	(0.026)
Urban Agglomeration	-0.025	(0.060)		0.068	(0.033)
Metropolitan Area	0.485	(0.144)	**	0.532	(0.079)
Other Statistical Concept	0.412	(0.490)		0.248	(0.269)
City size (based on UN population)					
< 100K Persons (reference)					
100-500K Persons	0.158	(0.043)	**	0.046	(0.024)
500K-1M Persons	0.418	(0.066)	**	0.211	(0.036)
1M + Persons	0.578	(0.074)	**	0.316	(0.041)
GRUMP data quality					
Number of points per extent	-0.100	(0.006)	**	-0.073	(0.004)
Number of admin units per extent	-0.003	(0.000)	**	-0.002	(0.000)
Continent					
Africa (reference)					
Asia	-0.185	(0.042)	**	-0.092	(0.023)
Central America	-0.154	(0.089)		-0.059	(0.049)
South America	0.005	(0.057)		0.033	(0.031)
Constant	0.083	(0.064)		0.074	(0.035)
N	1,703				1,703
F(12, 1690)	33.130				52.770
Prob > F	0.000				0.000
R-squared	0.190				0.273
Adj R-sq	0.185				0.267

NB: The table omits all many-to-one UN-to-GRUMP-extent matches.

Figure 8. Buenos Aires and La Plata, Argentina. UN-GRUMP matched data.



NOTE: The UN has one entry for Buenos Aires, an urban agglomeration, rendered here with many possible point matches.