LINCOLN INSTITUTE
OF LAND POLICY

Urban Expansion in a Global Sample of Cities, 1990 – 2014

Working Paper WP18AB2

Alejandro M. Blei
Marron Institute of Urban Management

Shlomo Angel
Marron Institute of Urban Management

Daniel L. Civco
Center for Land Use Education and Research

Nicolas Galarza
Marron Institute of Urban Management

Achilles Kallergis
Marron Institute of Urban Management

Patrick Lamson-Hall
Marron Institute of Urban Management

Yang Liu
Marron Institute of Urban Management

Jason Parent
Marron Institute of Urban Management

November 2018

Abstract

Using a 200-city sample that was carefully selected to represent the universe of 4,231 cities in in 2010, we generated measures of cities' areas and populations at three points in time over a 24-year period. The 600 observations were used to calculate the growth rates of cities' areas, or their urban extents, and their populations, over three analysis periods: 1990 – 2000, 2000 – 2014, and 1990 – 2014. During 2000 – 2014, the most recent period, the median urban extent growth rate for less developed countries cities was 5.7 percent per year compared to 1.1 percent per year in more developed countries cities and 3.1 percent per year in all cities. Average growth rates were higher. A quantity that grows at 5.7 percent per year doubles in size in 12 years and triples in size in 19 years. The median population growth rate over this period in less developed country cities was 3.6 percent per year, compared to 0.7 percent per year in more developed country cities and 2.2 percent per year in all cities. We observed statistically significant declines in the average urban extent growth rate and the average population growth rate from 1990 – 2000 to 2000 – 2014 across all three analysis categories. Despite these declines, the average urban extent growth rate was greater than the average population growth rate at all time periods in each of the categories and this difference was statistically significant. The factors that could explain observed variation in urban extent and expansion were explored in multiple regression models. Population and income were the overwhelmingly dominant factors. These two factors alone explained 85 percent and 65 percent of the variation in urban extent and urban expansion in the 200-city sample. This analysis is very similar to the proposed UN Sustainable Development Goal (SDG) indicator 11.3.1 and offers a first glimpse at globally representative findings. The global sample of cities provides a platform that could be used to monitor progress of other SDG indicators measures as well.

Keywords: urban expansion, global monitoring, growth rate, population, SDGs

About the Authors

Alejandro M. Blei is a research scholar at the Marron Institute of Urban Management at New York University. He was a research coordinator for the Monitoring Global Urban Expansion research program, a tri-partite collaboration between New York University, UN-Habitat, and the Lincoln Institute of Land Policy. He is a co-author of the 2016 *Atlas of Urban Expansion.*

Address: 60 5th Avenue, 2nd Floor
New York, NY 10011
Telephone: 212-992-6872
Email: ablei@stern.nyu.edu

Shlomo (Solly) Angel is a Professor of City Planning and the Director of the NYU Urban Expansion Program at the Marron Institute of Urban Management at New York University.
Email: sangel@stern.nyu.edu

Daniel L. Civco is a Professor Emeritus of Geomatics and former director of the Center for Land Use Education and Research (CLEAR) at the Department of Natural Resources and the Environment of the University of Connecticut.

Address: 1376 Storrs Road, U-4087
Storrs, CT 06269
Telephone: 860-486-0148
Email: daniel.civco@uconn.edu

Nicolas Galarza is a research scholar at the Marron Institute of Urban Management and coordinator of the Colombia Urban Expansion Initiative.
Email: ngalarza@stern.nyu.edu

Achilles Kallergis is a research scholar at the Marron Institute of Urban Management at New York University. He was a coordinator for the Land and Housing Survey in a Global Sample of Cities.
Email: akallerg@stern.nyu.edu

Patrick Lamson-Hall is a research scholar at the Marron Institute of Urban Management. He is the New York coordinator of the Ethiopia Urban Expansion Initiative and the India Urban Expansion Observatory. He is a co-author of the 2016 *Atlas of Urban Expansion*.
Email: plamsonh@stern.nyu.edu

Yang Liu is a research scholar and statistician at the Marron Institute of Urban Management at New York University.
Email: yl3371@nyu.edu

Jason Parent is an assistant research professor in the Department of Natural Resources and the Environment at the University of Connecticut's College of Agriculture, Health, and Natural Resources. He specializes in the application of remote sensing and geospatial technologies.

Address: 1080 Shennecossett Rd.
Groton, CT 06340
Telephone: 860-405-9214
Email: Jason.Parent@uconn.edu

Table of Contents

Urban Expansion in a Global Sample of Cities, 1990 - 2014

Introduction

In 2015, the human population of the planet was 7.3 billion. More than half of that total (3.96 billion or 54%) lived in cities and towns and that share is expected grow over the coming decades (United Nations 2014). By 2050, the world's urban population is projected to grow by 60%, adding 2.38 billion people to cities and towns. That increase will be highly skewed toward less developed countries, which will absorb 95 percent of all new urban dwellers. In other words, between 2015 and 2050, 19 persons will have been added to the urban population in less developed countries for every single person added to the urban population in more developed countries. And yet, surprisingly perhaps, the urbanization of our planet is slowing down. While the absolute number of people living in cities and towns continues to rise, the rate of that increase is decreasing and it will continue to decrease into the future. The movement of people from living closer to the land to living closer to each other, a process that began in earnest around the turn of the 19th century when only 5 percent of all people lived in cities and towns, is likely to end by 2100, when that number approaches 75-80 percent.

What are the implications of this added population to cities and towns, more than two billion people by 2050 and another billion and a half by the end of the century?[1] For one, it has helped bring about new views on the relation of cities and towns to development policy and to their roles as pathways to sustainable development in particular (Parnell 2016; Barnett and Parnell 2016). A stand-alone goal for cities and human settlements in the United Nation's post-2015 development agenda, the 2030 Agenda for Sustainable Development, was a historic first, and the goal's ten targets and fifteen indicator measures suggest global commitment to action and reporting at the city level.

Within Sustainable Development Goal (SDG) number 11, indicator measure 11.3.1, "Ratio of land consumption rate to population growth rate," makes information about the areas cities occupy, their populations, and their change over time, essential inputs to the reporting process. Shifting the discussion of urbanization toward a land-based perspective, effectively framing urbanization and its challenges in spatial terms, as target 11.3 begins to do, makes confronting the challenges of urbanization, we believe, a more manageable task. Namely, the growth of population cannot be effectively guided by policy but cities occupy land, and programs aimed at urban land, or at the conversion of land to urban use – for housing, public works, public open spaces, and other public amenities, is very much guided by public policy.

We know that cities and towns will absorb great numbers of new residents and that this entails the occupation of existing and new lands. But how much land? This is where our interest in studying the change in cities' areas and populations lies. When we have globally representative and historical data on this relationship and when we investigate the factors that influence it, we

[1] Assuming an urban population growth rate of 0.5 percent between 2050 and 2100.

introduce an evidence-based approach to understanding how much additional land cities are likely to occupy in the future. This knowledge can be applied in planning and preparing for the expansion of cities, so that decisions of how to respond proactively are based on realistic assumptions and targets.

In this paper, we report our findings on changes in the areas and populations of cities, in a global sample of 200 cities over the 1990 – 2014 period. While the data and analysis we generate is similar to that associated with indicator 11.3.1, our motivation for undertaking this work is different.

Our motivation is a pragmatic one. Making minimal preparations for land and infrastructure in advance of development is much easier, from financial and logistical points of view, when land on the periphery is unoccupied and relatively cheap. Once these areas are occupied, it is very difficult to relocate populations and it is expensive to reconfigure areas that are already built. In rapidly growing cities and towns, failure to properly service land, secure public open space, and obtain rights of way for roads and infrastructure in advance of development can have negative and lasting consequences for individuals, the economy, and the environment.

How much additional area rapidly growing cities will add need not be a mystery, however. We shed light on this question through our analysis. The basis of our investigation is a 200-city sample that was carefully selected from the universe of 4,231 cities with populations over 100,000 in the year 2010. The sample includes cities of all populations sizes, in every world region, in countries large and small. We studied the same 200 cities at three time points, which allowed us to measure and compare change over time. A city's urban extent, which defines the hard edge that was needed to bound the area and population calculations for a given city at a given date, was identified using a methodology developed by the research team.

A median urban extent growth rate of 3.1 percent per year was observed for all cities over the 2000 – 2014 period compared to an average growth rate of 5.0 percent per year. The summary values for all cities mask important differences between different groups of cities. In less developed country cities, the median urban extent growth rate was 5.7 percent per year compared to an average rate of 6.2 percent per year, while in more developed country cities, the median and average rates were 1.1 percent per year and 1.8 percent per year respectively. A growth rate of 5.7 percent per year implies a doubling time of 12 years and a tripling time of 19 years. A growth rate of 1.1 percent per year implies a doubling time of 64 years and a tripling time of 100 years. The situation in these two groups of cities is clearly very different.

Average urban extent and population growth rates were found to have declined from the 1990 – 2000 to the 2000 – 2014 period. Further declines in these growth rates can be expected if the trend continues. The observed declines did not affect the relationship between the urban extent growth rate and the population growth rate in cities, however. We looked at the difference between the urban extent growth rate and the population growth rate at the city level and tested whether the average difference for cities was significantly different than zero. We concluded that urban extent grew faster than population, on average, in both less developed and more developed country cities, at both time periods, and that this difference was statistically significant. We might also expect this trend to continue in the near future, barring major changes.

An exploration into the drivers of urban extent and expansion pointed to two main explanatory factors: population and income. As cities grow in population and as their inhabitants become wealthier they occupy more area. These two factors alone could explain 85 percent of observed variation in models of urban extent and 65 percent of variation in urban expansion. The inclusion of additional factors, related to geography and climate, transport cost, building regulations, agricultural land, and global integration, improved the models' explanatory power very marginally.

The analysis presented here represents work associated with the first of a three-phase research effort entitled *Monitoring Global Urban Expansion*. In Phase I we mapped and measured the built-up areas and open spaces of the 200 cities, or their urban extents, and calculated their populations, to obtain estimates of recent expansion, density decline, compactness, and fragmentation. In Phase II we compared the spatial organization of streets and blocks in cities' 1990 – 2014 expansion areas to their pre-1990 areas to understand how block sizes, street widths, the share of the share of the land in streets, road density, and street grids have changed over time. In Phase III we engaged local experts to complete surveys on housing affordability and the rules and regulations governing land and housing in cities and their expansion areas, to better understand how the two interact. Findings associated with different aspects of these three phases will be presented in a series of working papers.

The paper is structured as follows. Section 2 describes the methods and data that were used to generate urban extent and population data for the 200 cities. Section 3 presents findings for urban extent and population growth rates and their change over time in all cities, less developed country cities, and more developed country cities. Section 4 discusses the rationale, data, and results for models of urban extent and expansion. Section 5 concludes.

Empirical Framework

Selecting the 200-City Sample

<u>The Universe of Cities</u>

The study of urbanization trends has typically occurred at the country level, based on a distinction between urban and rural population that is made by national statistical offices. Since countries apply different criteria to make this urban-rural distinction, such as different population thresholds, and since the results are lumped together into a single national value, it is difficult to advance our knowledge about urbanization attributes and trends in cities when we use this country-based data (Cohen 2003).

In this work, we have focused on cities as the unit of analysis. How many cities in the world are there? There is no simple answer to this question. Just as there are many ways to define the urban-rural distinction there are many ways to define a city, including any combination of population thresholds, administrative boundaries, density thresholds, and commuting and activity patterns (Parr 2007; OECD 2012; Uchida and Nelson 2008; Deuskar and Stewart 2016; and Taubenbock et al. 2014). Identifying all cities in the world required a definition that we could apply universally with existing data sources.

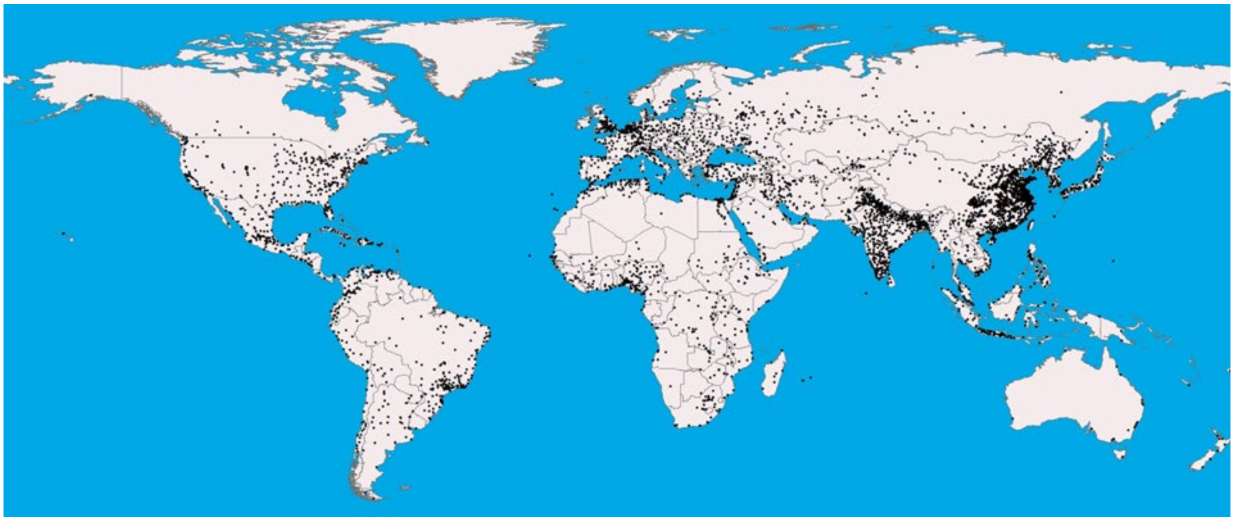
We chose to define cities by their geographical extents and by a population threshold of 100,000 in the year 2010. By geographical extent, we refer to the relatively contiguous built-up area extending out of a historical city center. The extent may stretch across many municipalities and is not constrained by administrative boundaries. The extent is visible to the naked eye using high resolution satellite imagery, such as that which can be seen on Google Earth. What the appropriate population threshold should be for defining cities is fairly subjective but few would disagree that a settlement with 100,000 inhabitants constitutes a city.

It was necessary to first identify candidate cities from lists of cities, municipalities, metropolitan areas and urban agglomerations containing a population value at 2010, or for which a population value at 2010 could be estimated. The two main data sources for this exercise were the UN Population Division, which provided data for settlements with populations of at least 300,000, and the website www.citypopulation.de, which reproduces census data and maps for all countries. Information from these lists was supplemented with internet research. According to official lists, China had only 662 cities in 2010. We identified 1,029 settlements in China with contiguous geographic extents that we believed had populations over 1000,000. The Chinese Academy of Sciences helped us estimate their populations.

Each candidate city was viewed on Google Earth to confirm its existence and to determine whether it was part of a larger agglomeration. Only cities that were not part of a larger agglomeration were included in our final list. This checking procedure led to the merging of many observations that were considered to be part of the same geographical extent as well as the exclusion of candidate cities that did not meet our criteria. For the largest agglomerated areas, such as the northeast corridor in the United States, and in many other locations as well, we used metropolitan area boundaries to limit the extent of a city. In doing so, we imply that the geographical extent of cities cannot extend forever and that they should roughly correspond to commuting areas or labor markets.

When the procedure was completed, our 2010 universe of cities contained 4,231 free standing cities with populations of at least 100,000. The locations of these cities are shown in Figure 1.

Figure 1: All 4,231 cities comprising the 2010 universe of cities.



The Global Sample of Cities

It is not possible to study each observation in the universe of cities and perhaps it should not be necessary, so long as there is a carefully constructed sample whose results can be generalized to the universe of cities as a whole. If we were to take a simple random sample of 200 cities from the universe of cities, we would end up with many small cities due to the fact that there are many more smaller cities than larger cities, and many cities in China, as approximately one quarter of the world's cities are located there. If we would like to know something about cities of different population sizes, in different world regions, and in countries large and small, it is necessary to organize the universe of cities into the relevant categories and to sample from these categories.

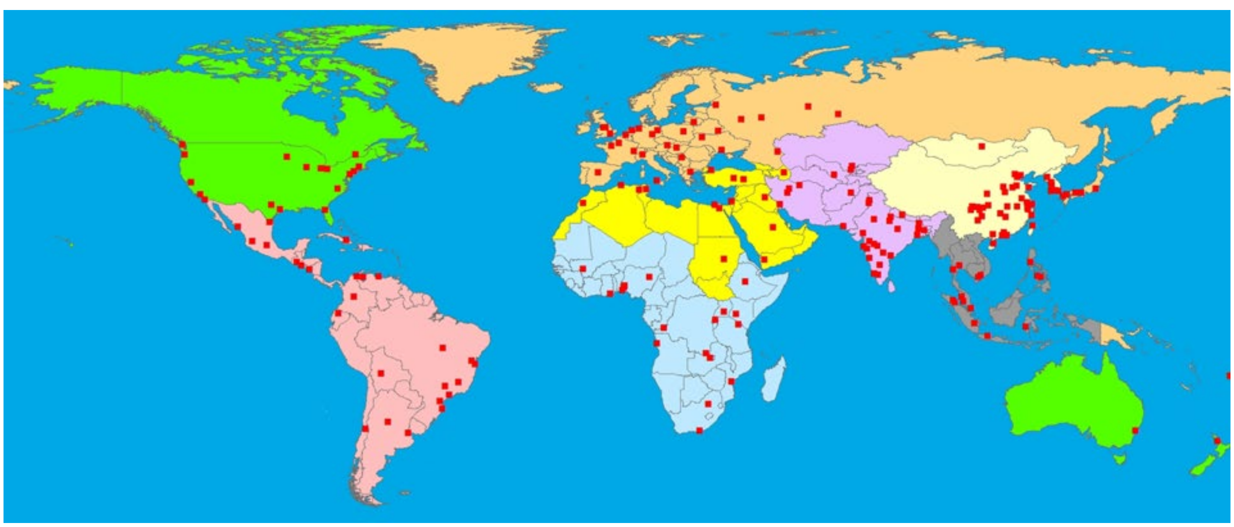
We organized the universe of cities along three strata with selecting a sample in mind. The first stratum was for the world region, of which there were eight: (1) East Asia and the Pacific, (2) Southeast Asia, (3) South and Central Asia, (4) Western Asia and North Africa, (5) Sub-Saharan Africa, (6) Latin America and the Caribbean, (7) Europe and Japan, and (8) Land-Rich Developed Countries. Land-rich developed countries include the United States, Canada, Australia, and New Zealand. The regional categories roughly follow the divisions in the United Nation's *World Urbanization Prospects.* That report also organizes the world into two mega-regions, more developed and less developed. The more developed regions category includes countries in North America, Australia, New Zealand, Europe, and Japan. The less developed region includes all other countries, even though some of them have high per capita incomes. We make use of the more developed less developed distinction later in our analysis. Cities were sampled from the eight regions in proportion to the population of the universe of cities in these regions.

The second stratum was for city population size. We created four categories of populations size roughly corresponding to small, medium, large, and very large: (1) 100,000 – 427,000, (2) 427,001 – 1,570,000, (3) 1,570,001 – 5,715,000, and (4) 5,715,001 and above. The total population of the universe of cities in each of these categories was approximately the same, about 622 million. An approximately equal number of cities was sampled for each of the four population size categories.

The third stratum was included so that the sample would contain cities from countries with few cities as well as cities from countries with many cities. The number of cities in the country stratum contained three categories: (1) 1 – 9 cities, (2) 10 – 19 cities, and (3) 20 or more cities. Cities were sampled from these categories in proportion to the population of the universe of cities in these categories.

We can combine the eight world regions, the four population size categories, and the three categories for the number of cities in the country to create 96 subcategories, or boxes, that represent every possible combination of world region – population size – number of cities in the country. Every observation in the universe of cities must fall into one of these boxes according to its region, its population, and the number of cities in the country to which it belongs. After we distributed all 4,231 cities among the 96 boxes there were 76 non-empty boxes. We selected cities at random from each box in rough proportion to the total population of the box. Ultimately, we selected cities from 61 boxes. The 15 boxes that were not directly represented by the sample contained 3 percent of the cities and 3 percent of the population in the universe of cities. We later added the information from these boxes to nearby boxes with the same region and with similar population and number of cities in the country attributes, so that these cities and their populations would be represented by the sample in later calculations. In this way, the entire universe of cities was represented by the sample of cities. The locations of the 200 sample cities and the eight world regions, in different colors, are shown in Figure 2.

Figure 2: The locations of the 200 sample cities and the eight world regions.



Weighting the Sample of Cities

Each of the 96 boxes from which the 200-city sample was selected is associated with a total number of cities and a total population. Consider a box that was used in our study, it contained 210 cities and the combined population of these cities was approximately 45 million. We sampled four cities from this box. Thus each of the four cities represents approximately 50 cities in its box. The combined population of the four sampled cities was 1.1 million. Thus each person in the sample in this box represents approximately 40 people in the universe of cities in this box.

We can expand the findings for the sample cities in this box to represent either the total number of cities in this box or the total population in this box. The values of 50 and 40 are in effect city-based and population-based weights, respectively. When we calculate a given metric for all cities in the sample, like the urban extent growth rate, and when we apply the same type of weight to all cities, we obtain the global weighted value for that metric. In this paper we have applied city-based weights to all calculations.

Measuring Urban Extent

The process of creating urban extent files for sample cities was organized into seven steps: (1) study area assessment, (2) spatial population data collection, (3) study area definition, (4) Landsat data collection, (5) Landsat classification, (6) landscape analysis, and (7) urban extent rule. A brief description of each step is outlined below.

Study Area Assessment

Our first task was to gauge how large an area the contiguous built up area extending out of an established historical core occupied circa 2014. This initial step was necessary to determine the area over which spatially explicit population data and Landsat satellite imagery would be needed to complete the analysis for each city. This task was undertaken by examining global nighttime light data for 2013 produced by the Earth Observation Group at the National Geophysical Data Center of the National Oceanographic and Atmospheric Administration. Stable nighttime light data (approximately 30m resolution at the equator) is publicly available and was downloaded from htpp://ngcd.noaa.gov/eog/dmsp.html. Since nighttime light data is known to overestimate built-up area extent, we believed it could be used to provide a sufficiently large enough area on which to base our population and satellite imagery data collection (Potere et al. 2009). Validation of the initial study area, to confirm the presence of built-up area, was conducted by viewing contemporaneous high-resolution satellite imagery over the same areas on Google Earth. The initial study area was increased or decreased based on the Google Earth comparison.

Spatial Population Data Collection

The twin goals of the analysis were to create urban extent boundaries for each city and to assign a population value to the urban extent. To assign a population value, we sought spatially explicit population data in the form of administrative boundaries, census enumeration zones, municipal districts, and city wards. Each of these spatial units had to list a residential population value associated with a date to be relevant for the analysis. Population estimates were sought for three dates: circa 1990, circa 2000, and circa 2014. To the extent that it was possible, we tried to use the same spatial units over time. Spatially explicit population data was used to apportion population to the urban extent using a procedure described in following section. Spatially explicit population data is typically made available by national or municipal governments in formats that can be visualized and manipulated in GIS software. When possible, we obtained the data directly from the relevant governmental agencies. Several non-governmental organizations that house repositories of population data, including the Socioeconomic Data and Applications Center (SEDAC) at Columbia University, the Chinese Academy of Sciences, the European Commission, and the website http://www.citypopulation.de were also consulted. Many times, however, especially in places with poor data programs, we communicated with scholars, development professionals, and local experts who provided us with maps, reports, and other documentation that we used to construct spatially explicit estimates over the area of interest. This sometimes required georeferencing and digitizing the material and assigning zones the relevant population values.

Study Area Definition

The delineation of the urban extent does not depend on population data directly; indeed, a key feature of this analysis is that urban extent boundaries are neither constrained nor defined by any type of boundary or zone, population or otherwise. To identify a city's urban extent, however, we needed to collect, classify, and analyze Landsat satellite imagery. Furthermore, to apportion population to the urban extent, we needed the set of population zones that completely contained the urban extent. To streamline the process of satellite imagery collection and analysis, urban extent creation, and population apportionment, we defined the study area by the set of population zones we believed would completely contain the urban extent. This decision was informed by our initial study area assessment and was almost always sufficient to complete the analysis of a city. In a handful of cases, however, the final definition of the study area was determined through an iterative process. More precisely, upon creating the urban extent, we sometimes observed that it ran up against the study area boundary rather than terminating successfully within the study area. In these cases, we acquired additional population data to increase the size of the study area. There were a handful of exceptions to this rule, where the iterative process would have led to urban extents that contained more than one functional urban area or more than one metropolitan labor market. In these cases, we kept the study area boundaries fixed, using local definitions of metropolitan area boundaries or basing the decision on expert opinion.

Landsat Data Collection

Landsat scenes from Landsat 4, 5, 7, and 8 satellites, corresponding to dates circa 1990, 2000, and 2014 were downloaded from the United States Geological Survey's Earth Explorer website: https://earthexplorer.usgs.gov. The satellites have revisit times of approximately 16 days. The Earth Explorer interface allows for the user to specify an area of interest and to preview all relevant Landsat scenes over this area. Scenes that were cloud free, especially in and around the cities' study areas, were selected for analysis. This sometimes resulted in the selection of images that were not exactly at the target dates of 1990, 2000, and 2014 but a few years before or after. Actual imagery dates associated with a particular city are the T1, T2, and T3 associated with that city in our analysis. Scenes measure 185 x 185 kilometers. The basic building block of a scene is a Landsat pixel, which has a typical size of 30-by-30 meters.

Landsat Classification

Study area boundaries were superimposed on Landsat scenes corresponding to the three time periods. The intersected area, with an additional 1 km buffer, was selected for classification. Our objective was to extract three types of land cover categories from the Landsat images: water, built-up, and other/open space. Water refers to any Landsat pixel that is comprised of surface water, including oceans, lakes, ponds, reservoirs, rivers, streams, canals, pools, and flooded wetlands. Built-up refers to any Landsat pixel that is comprised of structures or surfaces constructed by humans, including buildings, roads, parking lots, racetracks, railroads, and docks. Other/open space refers to any Landsat pixel comprised of vegetated surfaces and barren lands, including forests, agricultural crops, fallow agricultural fields, wetlands, grass lands, desert lands, beaches, mountaintops, and other land cover types that are neither water nor built-up. For a review of remote sensing concepts and methods and an explanation of the unsupervised classification technique employed in this analysis, including post classification processing and editing, see Angel et al. (2005), Chapter 3. The three-way classification of Baku, Azerbaijan into water, open space, and built-up area is shown in Figures 3 and 4.

Figure 3: The three-way classification of Baku, Azerbaijan into water (blue), open space (brown), and built-up (red).

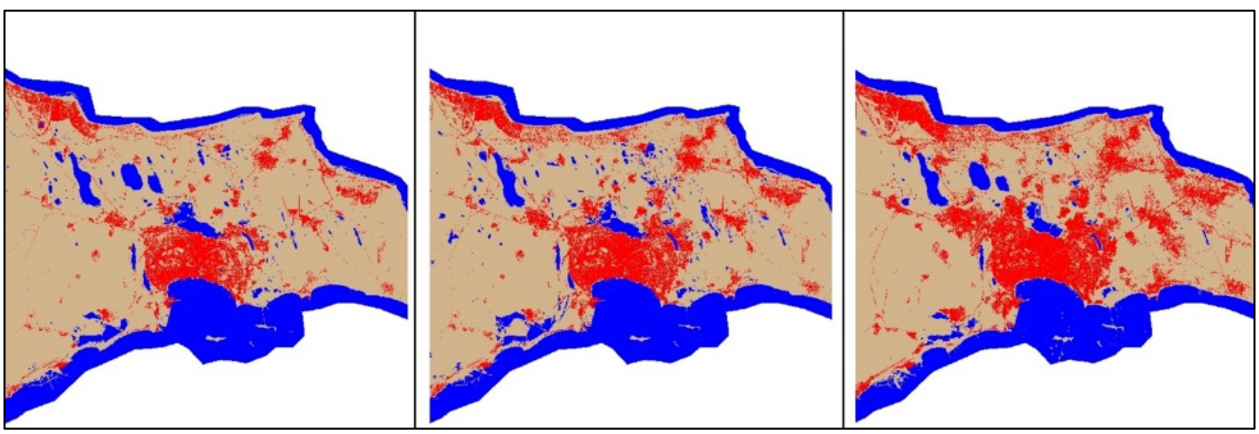
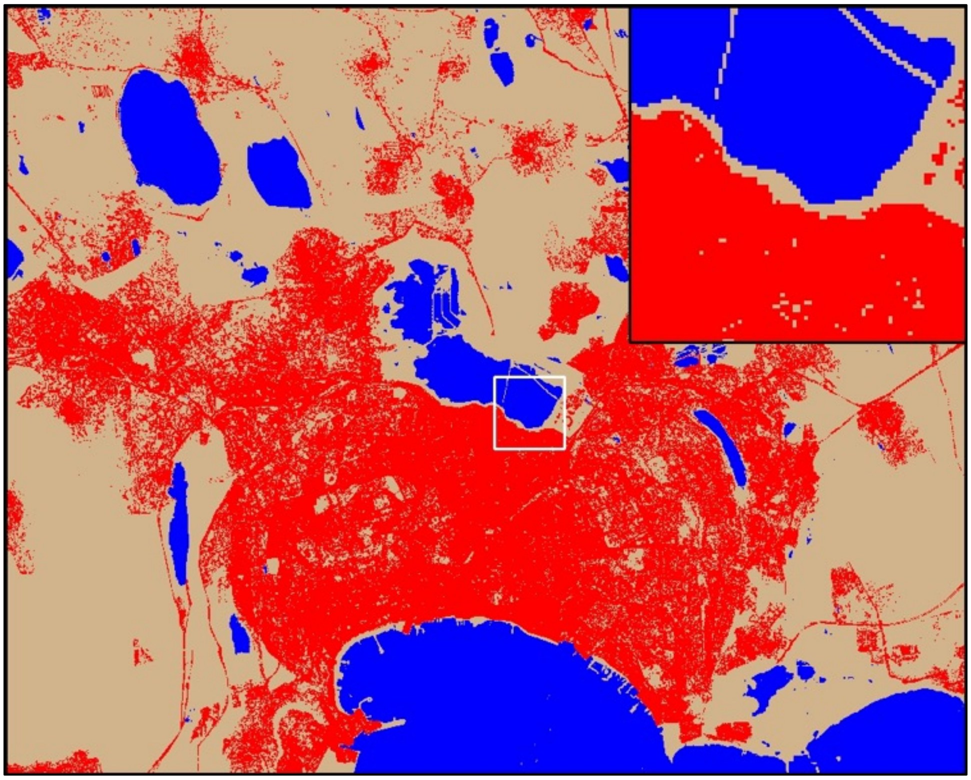


Figure 4: A close up of Baku's three-way classification illustrating the pixelated nature of the image.



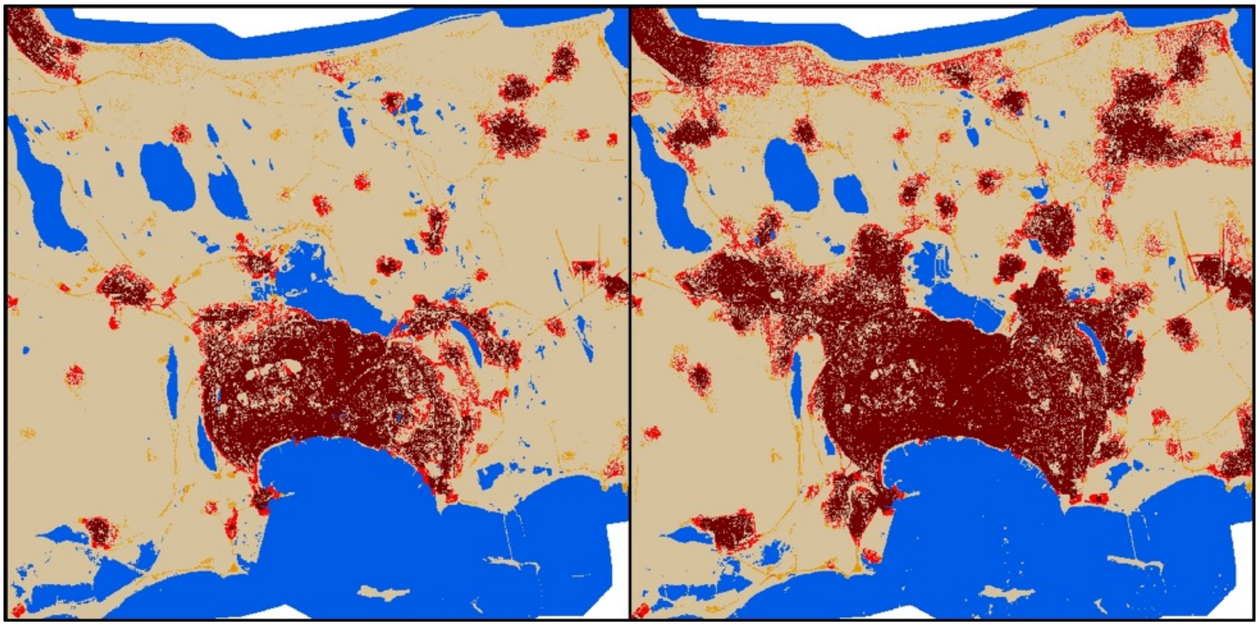
Landscape Analysis

The three-way classification of water, built-up, and open space was the input into a secondary analysis. This secondary analysis, or landscape analysis, sub-classified built-up and open space pixels into three categories each, allowing us to differentiate among different types of built-up and open space pixels. The sub-classification of the built-up class was based on the spatial density of built-up pixels within the Walking Distance Circle, defined as the 1 km^2 circle about a pixel. The three categories of built-up produced by the landscape analysis include:

1. *Urban* pixels, where the majority (> 50%) of pixels within the Walking Distance Circle are built up;
2. *Suburban* pixels, where 25-50% of pixels within the Walking Distance Circle are built-up; and
3. *Rural* pixels, where < 25% of pixels within the Walking Distance Circle are built-up.

The use of the terms urban, suburban, and rural to describe built-up pixels across the study area does not imply literal interpretations of how these terms manifest spatially. They were used in the sense that the areas they refer to generally correspond to our perceptions of what constitutes urban, suburban, and rural area in many cities throughout the world. The thresholds for the different categories are arbitrary and a different set of cutoffs would change the proportion of built up pixels in each category. We settled on these particular cutoff after experimenting with different combinations of values in many cities, examining the output, and deciding which combination of values was associated with the most consistent and intuitive results.

Figure 5: The classification of built up area into urban pixels (dark red), suburban pixels (red), and rural pixels (ochre) in Baku in July 1989 (left) and August 2014 (right).

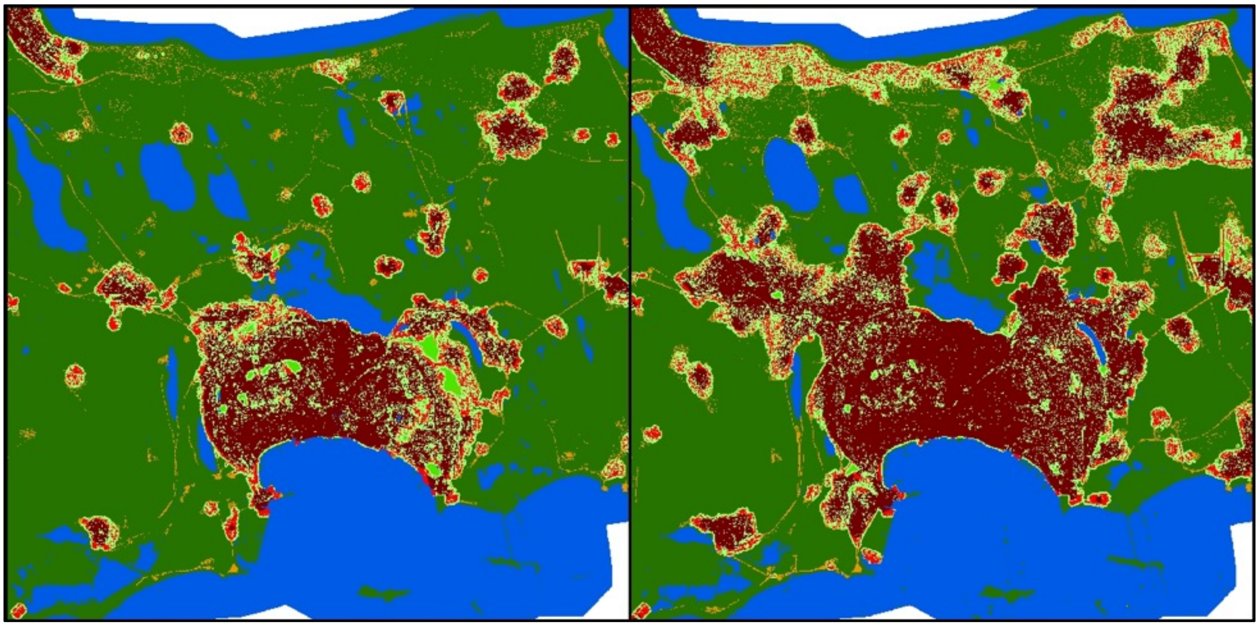


The three categories of open-space produced by the landscape analysis include:

1. *Fringe* open space pixels, all open space pixels within 100 meters of urban and suburban built-up pixels;
2. *Captured* open space pixels, clusters of open space pixels completely surrounded by fringe open space pixels less than 200 hectares in area; and
3. *Rural* open space pixels, all open space pixels that were neither fringe nor captured.

The rationale for the fringe open space category comes from the field of landscape ecology, where different studies have shown that settlements and built-up areas affect vegetation and wildlife along their edges, often in a belt up to 100m wide (Chen Franklin and Spies 1992; Winter Johnson and Faaborg 2000). Although captured open space is beyond the 100m belt, we identify it as a separate category based on the idea it is open space that may be degraded by isolation from other open spaces. Taken together, the fringe and captured open space within a study area constitute urbanized open space. Urbanized open space and rural open space together make up all of the open space within the study area.

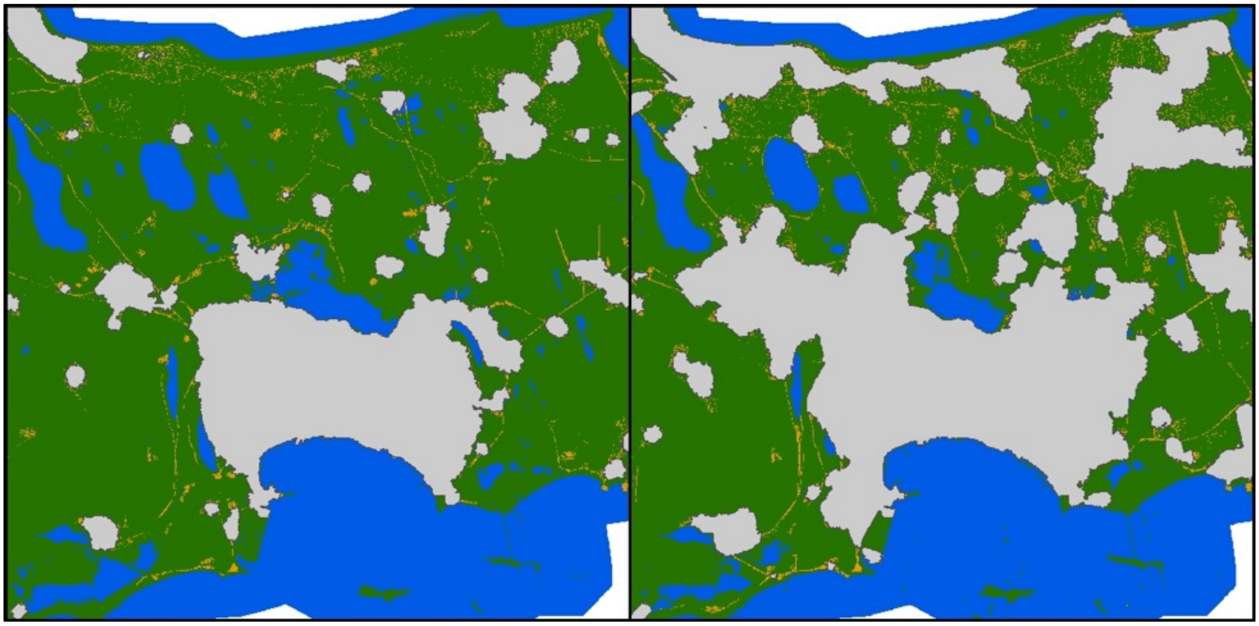
Figure 6: The classification of open space into fringe open space (light green), captured open space (bright green), and rural open space (dark green) in Baku in July 1989 (left) and August 2014 (right).



Urban Extent Rule

The landscape analysis differentiates the study area in a way that facilitates the creation of rules that can be used to identify urban clusters. We define urban clusters as discrete patches of urbanized open space that by definition contain urban and suburban built-up pixels. There is no limit to the number of urban clusters within a study area; sometimes there is only one cluster and sometimes there are thousands. In Baku, Figure 6 suggests that were dozens of urban clusters in 1989 and 2014. We can see the clusters more clearly in Figure 7 below. As a rule, the cluster containing the city hall location, which is usually indicative of a traditional city center and central business district, is included in the urban extent. Some of the other urban clusters within the study area may also become part of the city's urban extent. The challenge was to determine which other clusters to include.

Figure 7: Urban clusters across the Baku study area in July 1989 (left) and August 2014 (right).

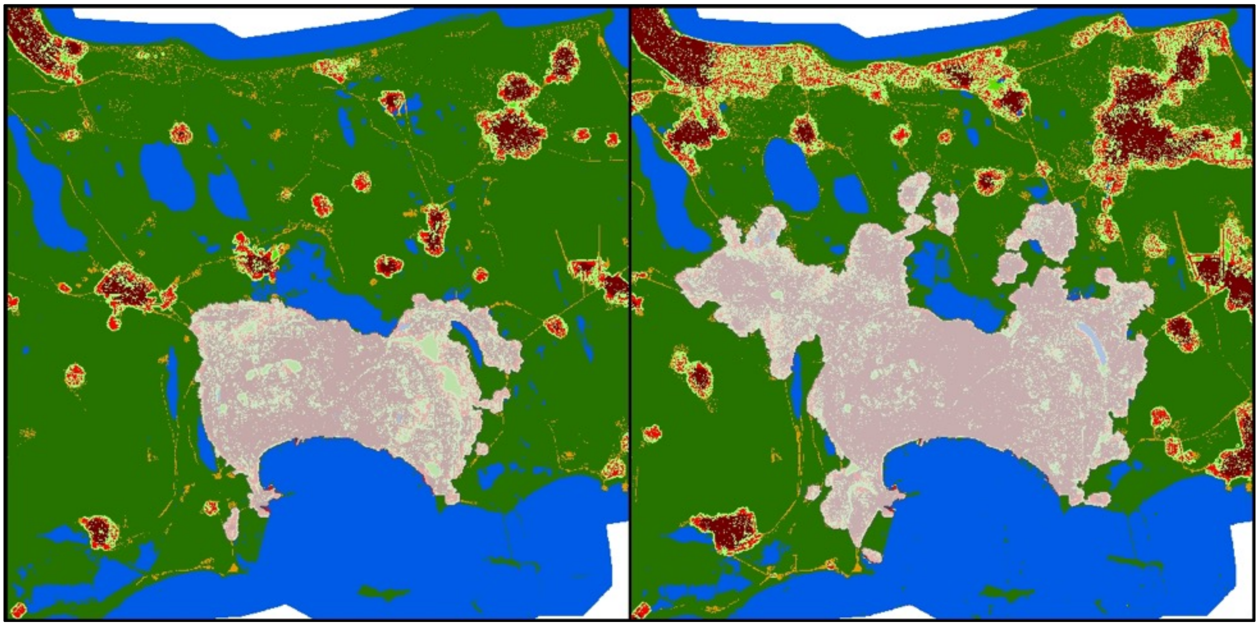


We employed rules based on the size and geographic proximity of clusters to each other to determine whether they should be grouped together into the same urban extent. We used these these rules in the absence of globally available data that could be used to measure the strength of commuting ties between clusters, for example, or local knowledge about whether separate clusters should be considered to be one or two distinct cities.

The decision of whether to group individual clusters together depended on an inclusion rule. We first generated a buffer around each cluster where the edge of the buffer area is always equidistant from edge of the cluster. The buffer distance for a given cluster is a function of the combined area of urban and suburban pixels within the cluster, more specifically: $buffer\ distance_{meters} = 6.659\sqrt{Area\ HA_{urban+suburban}}$. When applied, the buffered area contains one-quarter the area of the cluster. The inclusion rule unites all clusters whose buffers intersect one another. The new groupings of clusters become urban extents. The urban extent for the city in question is the grouping of clusters that contains the city hall location. Figure 8 shows final urban extent selections for Baku in 1989 and 2014.

The exact formulation of the inclusion rule was the result of attempts by the research team to group urban clusters in a way that corresponded with accepted notions of what constitutes the spatial extent or spatial footprint of a city. In a sense, the task was a form of pattern recognition. The pattern is sometimes easy to discern, such as a single large cluster completely surrounded by open countryside, or it may be more complex, such as clusters of varying sizes in different proximities to each other, similar to, but typically more complex than the Baku example. We apply a single rule to all situations and while it performs quite well, it is not perfect. In a small handful of cases we made manual corrections to add areas that should have been included in the urban extent, such as clusters on opposite sides of water bodies, as was the case in Hong Kong.

Figure 8: The Baku urban extent in July 1989 (left) and August 2014 (right).



Apportioning Population

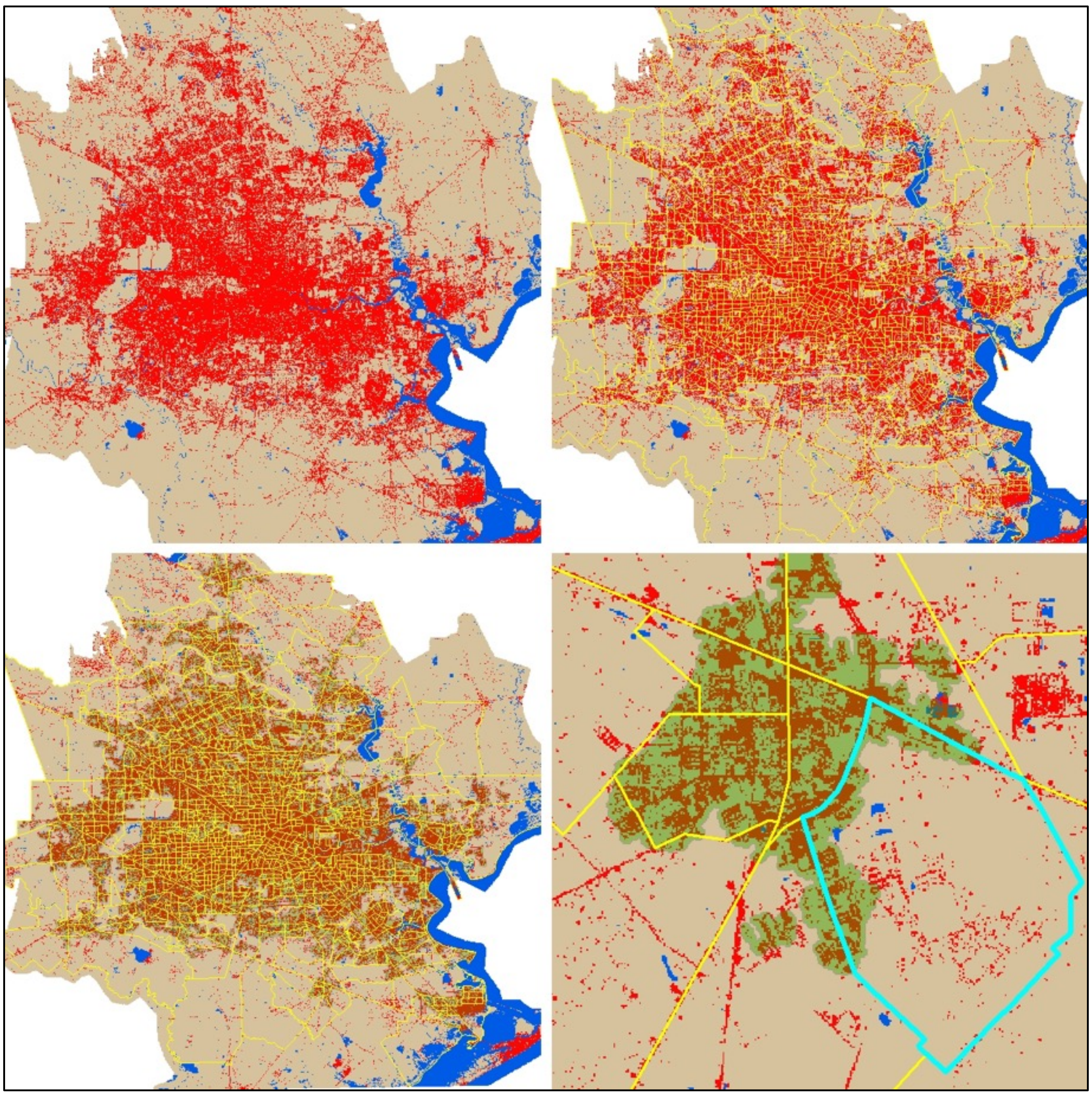
The process of assigning, or more precisely, of apportioning population to a city's urban extent entailed three pieces of data: (1) the urban extent boundary file, (2) the set of population zones that completely contained the urban extent, and (3) the three-way classification of the study area, containing the information for all built-up pixels.

The first task was to interpolate or extrapolate the population data to match the Landsat imagery dates. Linear interpolation/extrapolation was used to forward project or backward project population data using the closest available data points. In other words, if the imagery date was July 1992 but the population data corresponded to January 1990 and January 2000, we used linear interpolation to estimate the population of zones at July 1992 based on the populations of the zones at January 1990 and January 2000.

Next, we laid the spatially referenced population data over of the urban extent file and over the three-way classification. We made the assumption that the population of a zone could be equally distributed to all the built-up pixels within that zone as population resides in built areas as opposed to open spaces. In reality not all built-up area within a zone may be associated with population and the distribution of population within a zone may be unequal. Lacking information that could help us make these distinctions, we stuck with our assumption.

Within a given population zone, some built up pixels fall inside the urban extent and some pixels fall outside the urban extent. This this is clearly illustrated in Figure 9. Which contains an example for Houston in 2014. Starting in the upper left-hand corner, we see the three-way classification for the entire Houston study area. In the upper right hand corner we see Houston's population zones, in yellow, which correspond to census tract boundaries, overlaid on the three-way classification. In the lower left-hand image, we have introduced the T3 urban extent boundary, which lies in between the three-way classification and the population zones. It has been made transparent so that the red built-up pixels underneath are visible. In the lower right-hand corner we see the close up of an individual zone on Houston's periphery. Some of the red pixels in this image belong to the urban extent, namely those that fall within the green transparent cluster, some of the red pixels fall outside the urban extent.

Figure 9: The data required for population apportionment in Houston, Texas and an example on Houston's periphery.



We focus on the zone that is highlighted in fluorescent blue. Our goal was to estimate the population of the urban extent in this zone. If we assume that population is equally distributed within a zone, our target value is simply the share of the built-up pixels within the urban extent in the zone multiplied by the population of the zone. This is the population that we apportioned to the urban extent. We repeated this procedure for each population zone that intersected the urban extent. We summed the result to obtain the apportioned urban extent population.

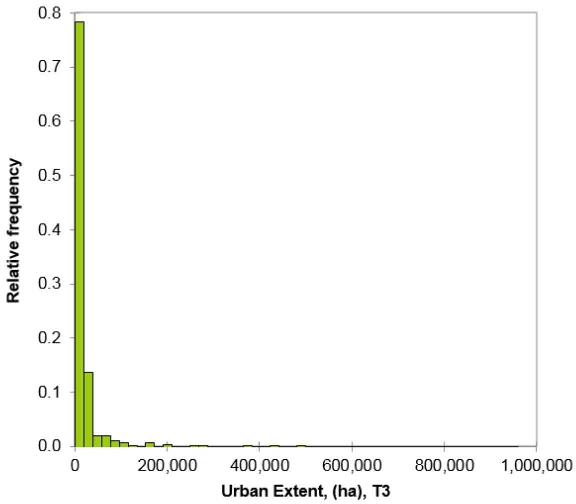
Findings

In this section we present our findings on urban extent and population, and their change over time, in the 200-city sample. We first examine the empirical distributions of urban extent and population to see whether known probability distributions can explain the observed frequencies obtained from our analysis. We then look at urban extent and population change in the sample as a whole, in less developed country cities, and in more developed country cities. For each of these three categories we conduct the same analyses based on data for three time periods: T1-T2, T2-T3, and T1-T3, roughly corresponding to 1990 – 2000, 2000 – 2014, and 1990 - 2014. First, we look at the distribution of urban extent growth rates and their average and median values. We then test whether the change in the urban extent growth rate from T1-T2 to T2-T3 is statistically significant, and if so, in what direction. Second, we look at the distribution of population growth rates and their average and median values. We then test whether the change in the population growth rate from T1-T2 to T2-T3 is statistically significant, and if so, in what direction. Third, we look at the city-level pairwise differences between the urban extent growth rate and the population growth rate to determine whether urban extent grew at a faster rate than population and whether this difference is statistically significant. Finally, we introduce the concept of the 1990 multiplier as an intuitive way to convey the magnitude of urban extent and population change over the study period. The section concludes with a summary of the findings.

Distributions of Urban Extent and Population

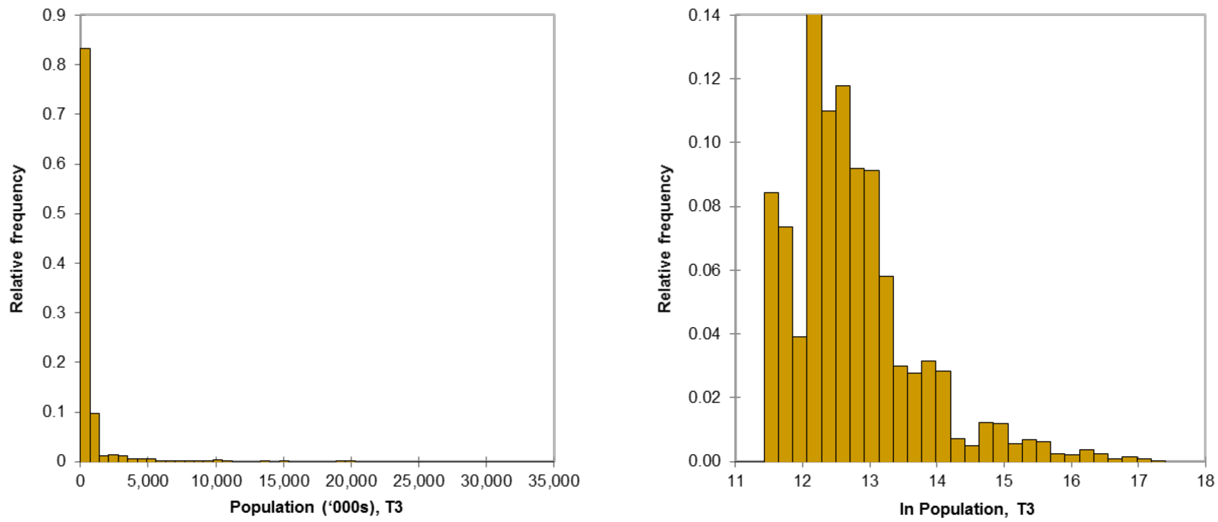
The 200-city sample was drawn from universe of cities which is known to contain many more smaller cities than larger cities. This leads to distributions of urban extent and population with long right tails. In Figure 10, below, on the left-hand side, we can see the weighted distribution of urban extent values for the T3 period. Most values are clustered on the extreme left and fall below 200,000 hectares. The tail extends out to 951,000 hectares on the right, corresponding to the New York City urban extent. The T1 and T2 distributions of urban extent follow similar patterns. We fitted probability distributions to the data using a statistical software package to determine whether a known distribution could explain the observed frequencies. At all three time periods, the log-normal distribution was associated with the best goodness-of-fit statistics. A log-normal distribution is one where the log values of the data follow a normal distribution. On the right side of Figure 10 we can see the weighted distribution of logged T3 urban extent. The shape of this distribution appears to be approximately normal. Using the estimated log-normal parameters associated with our data, we could explicitly test the hypothesis that the distribution of urban extent follows a log-normal distribution using the Kolmogorov-Smirnov test. For all three time periods, we failed to reject the null hypothesis that the data follows a log-normal distribution at a 0.01 significance level. This finding suggests the logged value is appropriate for use in regression models where urban extent is the dependent variable.

Figure 10: The untransformed and log-transformed distribution of urban extent in the 200-city sample.



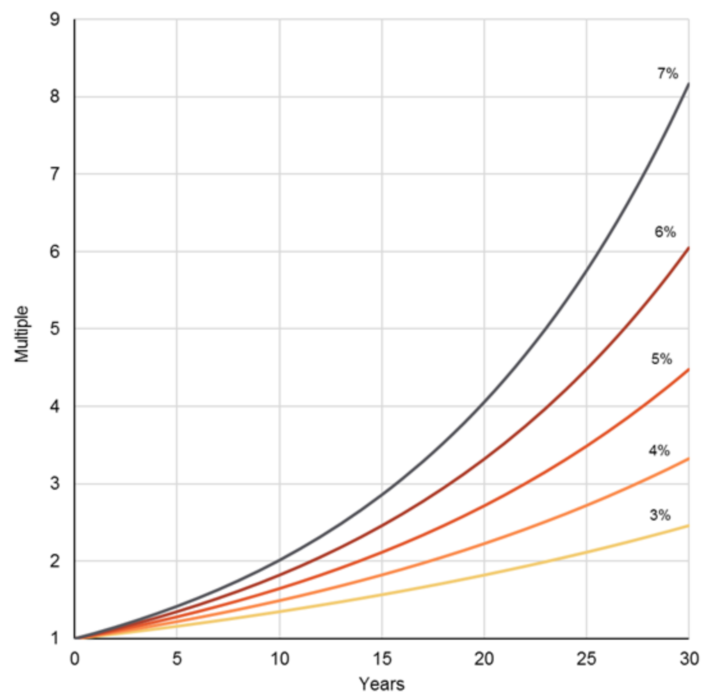
While urban extent values refer to the sample only, we can refer to either the sample or the universe of cities when we discuss the distribution of population. In this paper, we focus on the sample only. The weighted distribution of urban extent population based on sample cities is shown in Figure 11 on the left-hand side. As expected, the values are clustered at the extreme far left of the chart. Most cities fall below the 5 million mark and most are much smaller, reinforcing the notion that smaller cities, in absolute number, outnumber larger cities. The largest observation is 34.8 million, associated with the Tokyo urban extent. As there are only three cities in Tokyo's sampling box, it represents only three out of 4,231 cities, or 0.07% of the universe of cities, hence the negligible height of its bar. The T1 and T2 distributions of population show a similar pattern. As with urban extent, we tried to fit probability distributions to the data to determine whether a known distribution could explain the observed frequencies. At all three time periods, the log-normal distribution was associated with the best goodness-of-fit statistics. The right hand side of Figure 11 shows the weighted distribution of logged population values. The distribution appears less symmetric than that of urban extent and slightly more right skew. When we used the Kolmogorov-Smirnov to evaluate whether the estimated log-normal parameters follow a log-normal distribution, the test statistic fell within the critical region (alpha = 0.05), and we had to reject the null hypothesis that the data follow a log-normal distribution. A companion paper provides a detailed review of the statistical properties of the universe of cities and the 200-city sample and we refer the reader to that paper for further information.

Figure 11: The untransformed and log-transformed distribution of urban extent population in the 200-city sample.



Growth Rates

Urban extent and population growth rates refer to exponential growth models where the growth rate is based on years as the unit of time. Unlike linear growth rates, which are relatively easy to use to project by how much a quantity grows over time, projecting exponential growth rates is less intuitive. Following a growth curve on the graph in Figure 12 we can observe how long it takes, in years, for an initial quantity to double by identifying where that curve crosses the y-axis value of 2, and finding the corresponding x-axis value. At 5 percent growth per year, for example, the initial quantity doubles in 13.9 years and triples in 22 years. The table below the graph provides the doubling and tripling times associated with growth rates between 3 and 7 percent.



Growth Rate	3%	4%	5%	6%	7%
Time to double (yrs)	23.1	17.3	13.9	11.6	9.9
Time to triple (yrs)	36.6	27.5	22.0	18.3	15.7

Figure 12. Growth rates and their doubling and tripling times.

All Cities

Urban Extent

Urban extent was measured at three time points for all 200 cities. This allowed us to calcuate the growth rate of urban extent over the entire T1-T3 period, roughly a 25 year period, as well as over the T1-T2 and T2-T3 periods individually. As the urban extent grows over time, it may absorb built-up areas that existed at a previous time period. In this study, urban extent change is based on the total increase in the urban extent, which includes both newly converted open space to built-up area as well as previously existing built up area – area that was too far away or too small to be part of the city's urban extent at the previous time period. This differs from approaches that measure change in urban extent based on the conversion of open space to built-up area only (Schneider et al. 2015, Mertes et al. 2015). We have strutcured our analysis in such a way that when the urban extent grows from one time period to the next, we can identify which added pixels represent newly converted open space and which added pixels represent built-up area at the previous time period. We will report on the decomposition of added built-up area into four categories: infill, extension, leapfrog, and inclusion, this last category representing previously existing built up areas that were included into the urban extent as it grew, in a future paper. For all cities we find that on average about one-fifth to one-quarter of the change in the urban extent at each time period is inclusion.

Figure 13 shows the weighted distributions of growth rates associated with the three time periods in all sample cities. The average annual growth rate is binned along the x-axis and the relative frequency of that rate in the universe of cities is represented along the y-axis. It is immediately clear that growth rates are not the same everywhere. The distributions are right skew at all three time periods, with most rates clustered between 0 and 5 percent. At each time period there are also a substantial number of observations with rates above 5 percent. The greater range of x-axis vales in T1-T2 indicates higher growth rates observed over this period compared to T2-T3.

Figure 13: The distribution of urban extent growth rates for all cities, T1-T2, T2-T3, and T1-T3.

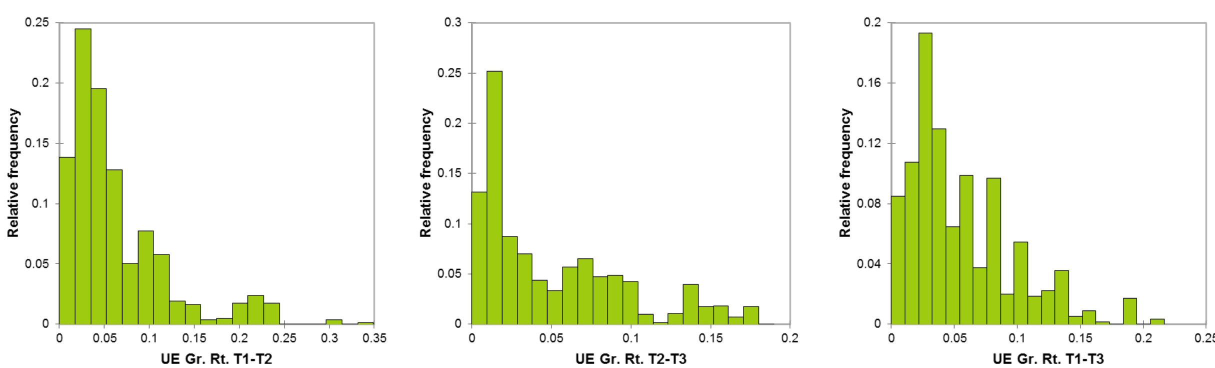


Figure 14 shows the weighted average urban extent growth rates associated with sample cities, and their 95 percent confidence intervals, over the three analysis periods. The weights expand the results to the universe of cities.

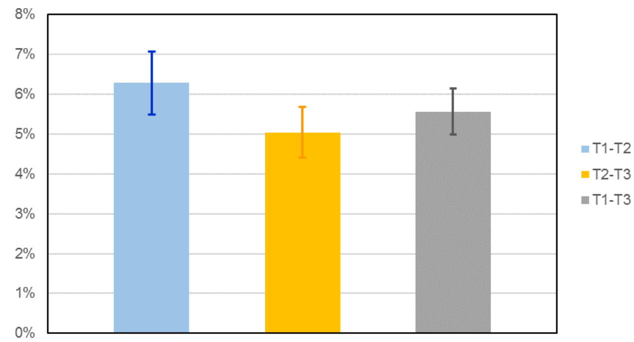


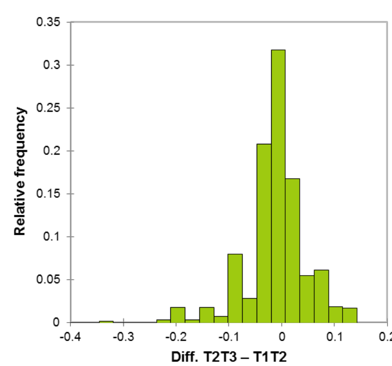
Figure 14. The average urban extent growth rate, all cities, T1-T2, T2-T3 and T1-T3.

Over the T1-T3 period, urban extent grew on average at 5.6 percent per annum with a 95 percent confidence interval of [5.0%, 6.1%]. The true average urban extent growth rate for all cities is unknown, it is a population parameter that we must estimate. We are 95 confident that the true average growth rate lies between 5.0 percent and 6.1 percent. The weighted median growth rate associated with the T1-T3 period was slightly lower, at 4.2%. A lower median was expected based on the skew distribution of urban extent growth rates. A table with descriptive statistics for the growth rates at each of the time period can be found in the appendix.

The point estimate for the growth rate over the T1-T2 period is 6.3 percent [95 percent C.I.: 5.5%, 7.1%] compared to 5.0 percent for the T2-T3 period [95 percent C.I.: 4.4%, 5.7%]. Median values for the two periods were 4.2 percent and 3.1 percent respectively. These differences suggest that growth rates are not constant over the entire T1-T3 period and that perhaps there has been a slowing down of the rate over time. While the confidence intervals for the two time periods overlap, suggesting that average weighted growth rates are not statistically different between T1-T2 and T2-T3, it is incorrect to base a judgement about change over time, at the city level, by comparing aggregate averages at the two time periods.

Our study is a repeated measures design, meaning that the values across the three time periods do not come from independent samples. Since these are paired observations, it is more appropriate to focus on the differences in growth rates at the city level between the T1-T2 and T2-T3 time periods. We evaluate whether these differences are different than zero, and if so, in what direction.

Figure 15 shows the distribution of these differences. For each city, the T1-T2 rate was subtracted from the T2-T3 rate. If the difference is positive, the T2-T3 rate was greater than the T1-T2 rate, and if negative, the T2-T3 rate was less than the T1-T2 growth rate. The differences are fairly symmetric and centered a little to the left of zero, with a median of -0.01 and an average of -0.012.



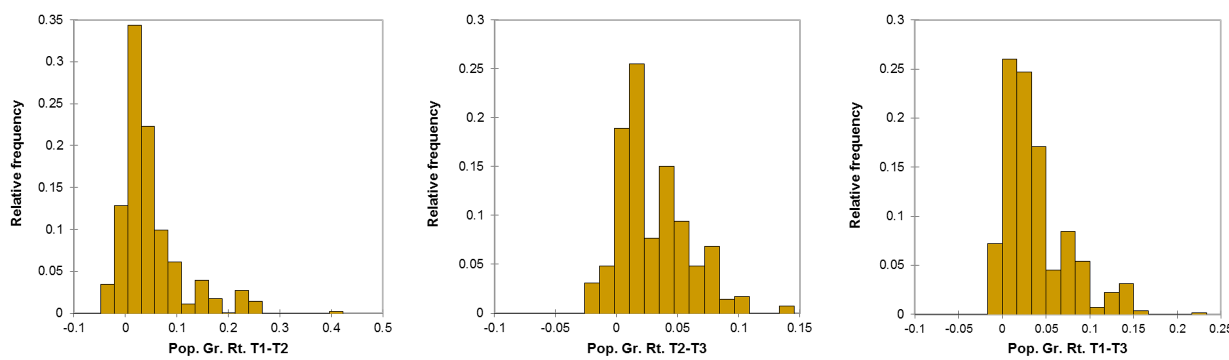
We use a weighted t-test on the differences to evaluate whether the average difference is statistically different than zero. A weighted t-test on the differences is equivalent to a weighted paired t-test for dependent samples. The result confirms that the difference in growth rates between the two time periods is different than zero, with a mean difference of -0.012 and a 95 percent confidence interval of [-2.1%, -0.4%]. We interpret this to mean that there has been a deceleration in the average growth rate of urban extent from the T1-T2 period to the T2-T3 period. This average difference was -1.2 percent in the sample, and we are 95% confident that that the difference lies between -2.1 percent and -0.4 percent in the universe of cities.

Figure 15. The distribution of differences in the urban extent growth rate, T3T2- T2T1, all cities.

Population

We repeat the analysis for urban extent on the populations of the 200 city sample. Figure 16 shows the weighted distributions of population growth rates. The overall shapes of the distributions appear less skew than the urban extent growth rate. If we exclude a very large value in T1-T2, corresponding to Hangzhou, China, the range of values along the x-axes is smaller for population growth rates than for urban extent growth rates at all three time periods.

Figure 16: The distribution of population growth rates for all cities, T1-T2, T2-T3, and T1-T3.



Average weighted population growth rates across the three time periods are shown in Figure 17. Over the T1-T3 period the average rate was 3.8 percent with a 95 percent confidence interval of [3.3%, 4.3%] and the median rate was 2.9 percent. This rate is not constant across the entire period, evidenced by different heights of the T1-T2 and T2-T3 bars. The average rate over the T1-T2 period was 4.7 percent with a 95 percent confidence interval of [3.9%, 5.6%] and the average rate over the T2-T3 period was 3.0 percent with a 95 percent

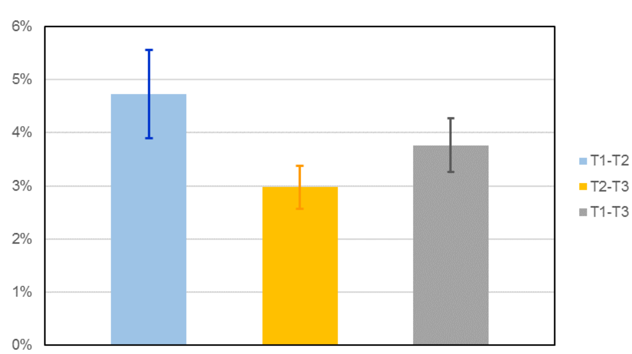


Figure 17. The average population growth rate, all cities, T1-T2, T2-T3 and T1-T3.

confidence interval of [2.6%, 3.4%]. The respective median growth rates were 3.0 percent and 2.2 percent. The trends suggest a slowing down of the population growth rate over time. Point estimates for average population growth rates across are smaller than urban extent growth rates across the three time periods.

We look at the differences in population growth rates over time, at the city level, to determine whether the average change over time is different than zero. Figure 18 shows the distribution of these differences, where the T1-T2 rate was subtracted from the T2-T3 rate. The distribution appears centered about zero with a left tail. The average difference is -0.017 and the median difference is -0.005. The weighted student's t-test on the differences confirms that the average is statistically different than zero and that that it is negative. The 95 percent confidence interval for the weighted average differences is [-2.5%, -1.0%]. We can interpret this to mean that the average population growth rate in the universe of cities has slowed down from T1-T2 to T2-T3. This average difference was 1.7 percent in the sample and we are 95 percent confident that the average difference is between -2.5 percent and -1.0 percent in the universe of cities.

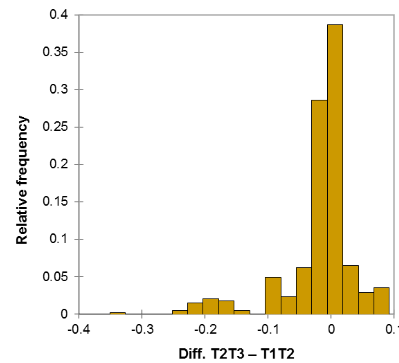


Figure 18. The distribution of differences in the population growth rate, T3T2- T2T1, all cities.

Urban Extent Growth Rate vs. Population Growth Rate

Are cities expanding outwards at faster rates than their populations are increasing? Figure 19 places the average weighted values for the urban extent growth rate and the population growth rate, with their 95 percent confidence intervals, side-by-side. The confidence intervals at T1-T2 overlap, but the overall pattern suggests higher rates of change for urban extent than population across the three analysis periods.

To answer this question, we must look at the pairwise differences between the urban extent growth rate and the population growth rate at the city level at a given analysis period. A weighted one-sample t-test on the paired differences tells us whether the differences are different than zero and whether they are significantly greater or less than zero.

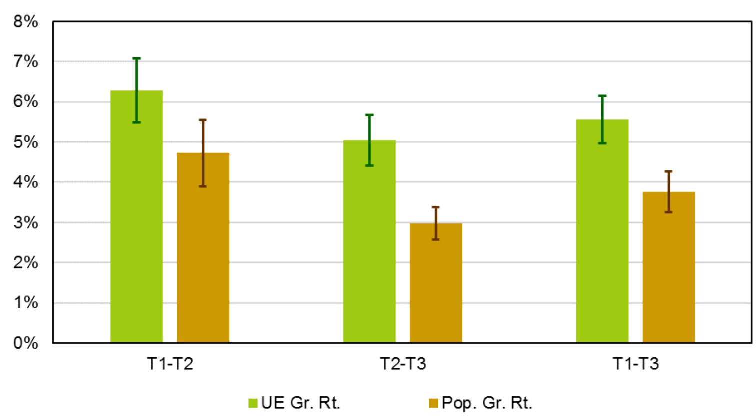


Figure 19. The average urban extent growth rate vs. the average population growth at all periods, all cities.

For each city, we consider its urban extent growth rate to represent an x value and its population growth rate to represent a y value. We subtract the y's from the x's for all cities to obtain the difference in the two growth rates over a given analysis period.

If the difference is positive, the urban extent growth rate is greater than the population growth rate. If the difference is negative, the population growth rate is greater than urban extent growth.

Figure 20: The distribution of differences, urban extent growth rate vs. population growth rate, all cities.

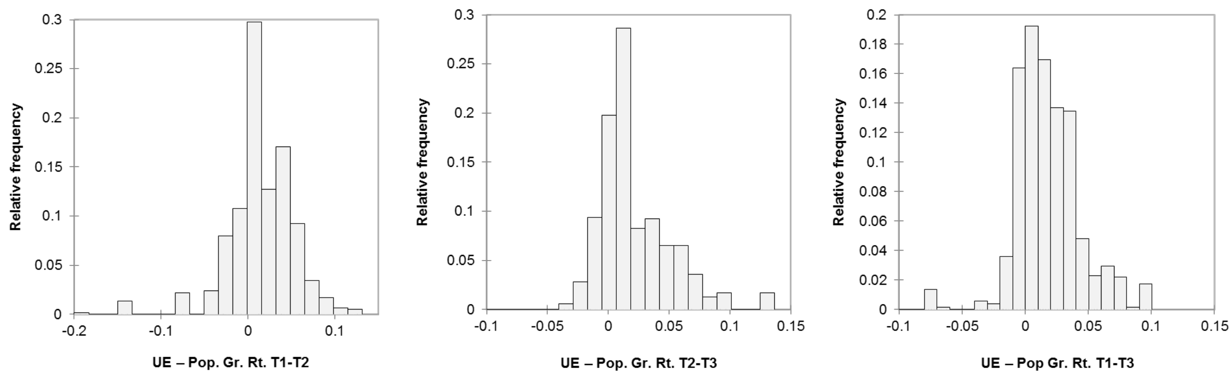


Figure 20 shows the weighted distributions of these differences. They are fairly symmetrical distributions with a modest right skew. Over the T1-T2 period, the mean difference is 0.016 compared to a median difference 0.014; at T2-T3 the mean difference is 0.021 compared to a median difference of 0.012 and over the T1-T3 period the mean difference is 0.018 compared to a median difference of 0.017. One-sample weighted t-tests on the differences at each time period, testing the null hypothesis that the differences are equal to zero, confirm that urban extent and population growth rates are different from zero at each time period, and more specifically, that the urban extent growth rate is greater than the population growth rate. Figure 21 shows the mean differences across the three time periods and their 95 percent confidence intervals. The data suggest that urban extent growth rates are greater than the population growth rates at all time periods and we know that both average urban extent and population growth rates have slowed down over time. Curiously, the overlapping confidence intervals for the difference between the urban extent rate and the population rate suggest that the degree to which the urban extent rate growth rate is different than the population growth rate is not statistically different over time.

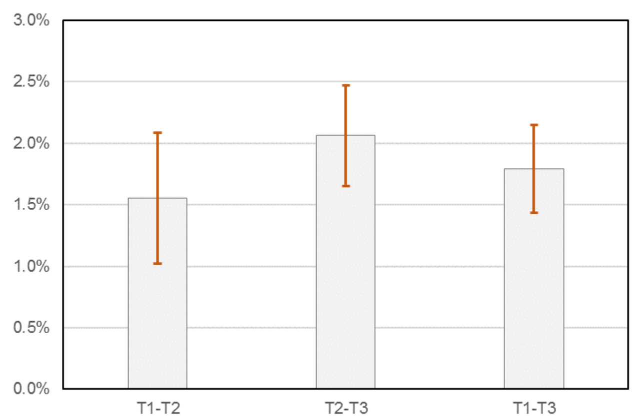


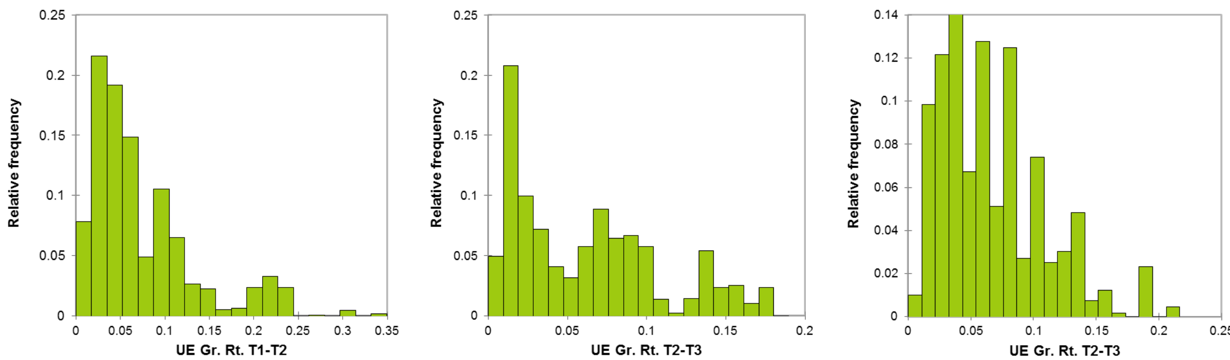
Figure 21. Average difference of the urban extent growth rate vs. the population growth rate in all cities

Less Developed Country Cities

Urban Extent

We focus now on cities in less developed countries. These cities are a subset of the 200-city sample. The weights expand the data so that the sub-sample is representative of all cities in less developed countries. In Figure 22, we see the weighted distributions of urban extent growth rates associated with the three time periods. The shapes of the distributions are very similar to the weighted distributions for all sample cities. Less developed country cities comprise 66 percent of the universe of cities. The greater range of x-axis values in T1-T2 indicates higher growth rates observed over this period compared to T2-T3.

Figure 22: The distribution of urban extent growth rates in less developed cities, T1-T2, T2-T3, and T1-T3.



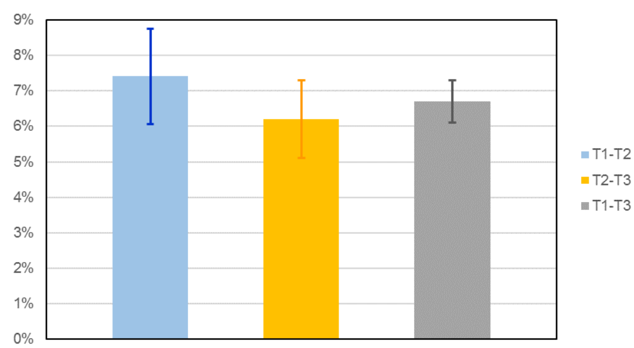


Figure 23. The average urban extent growth rate, less developed cities, T1-T2, T2-T3 and T1-T3.

Figure 23 shows the average weighted urban extent growth rates, and their 95 percent confidence intervals for cities in less developed countries. Over the T1-T3 period, urban extent grew on average at 6.7 percent per annum with a 95 percent confidence interval of [6.1%, 7.3%]. If we were to take repeated samples of similar size for cities in less developed countries, we would expect the average growth rate to fall between 6.1 percent and 7.3 percent 95 times out of 100. The weighted median growth rate associated with the T1-T3 period was 6.0 percent. The point estimate for the average T1-T2 growth rate is 7.4 percent [95 percent CI: 6.1%, 8.7%] compared to 6.2 percent [95 percent CI: 5.1%, 7.3%] for the T2-T3 period. The respective median values for these time periods are 5.5 percent and 5.7 percent. These estimates of average urban extent growth rates are higher, by about 1.1 percentage points, than estimates based on all sample cities. The difference in average growth rates between T1-T2 and T2-T3 suggests a slowing down of the growth rate over time. While the confidence intervals overlap, the repeated measures design calls for evaluating the differences in growth rates at the city level to determine whether the differences over time are significantly different than zero.

The distribution of the differences, where the T1-T2 growth rate for a given city was subtracted from the T2-T3 growth rate for the same city, can be seen in Figure 24. The values are centered about zero with a moderate left skew. The median difference is -0.010 and the mean difference is -0.012. Curiously, these are the same median and mean values associated with the analysis for all cities. A weighted t-test on the differences confirms that the difference in growth rates between the two time periods is different than zero at the 95 percent confidence level and, more specifically, that the average difference is lower than zero. The mean average difference has a value of -0.012 with a 95 percent confidence interval of [-0.0228, -0.0012]. We interpret this to mean that the urban extent growth rate has slowed down in less developed country cities from T1-T2 to T2-T3, and that we are 95 percent confident that this difference is between -2.3 percent and -0.01 percent in less developed country cities.

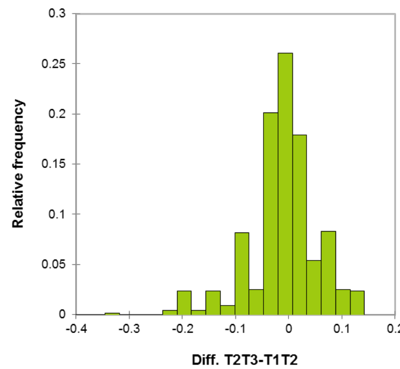
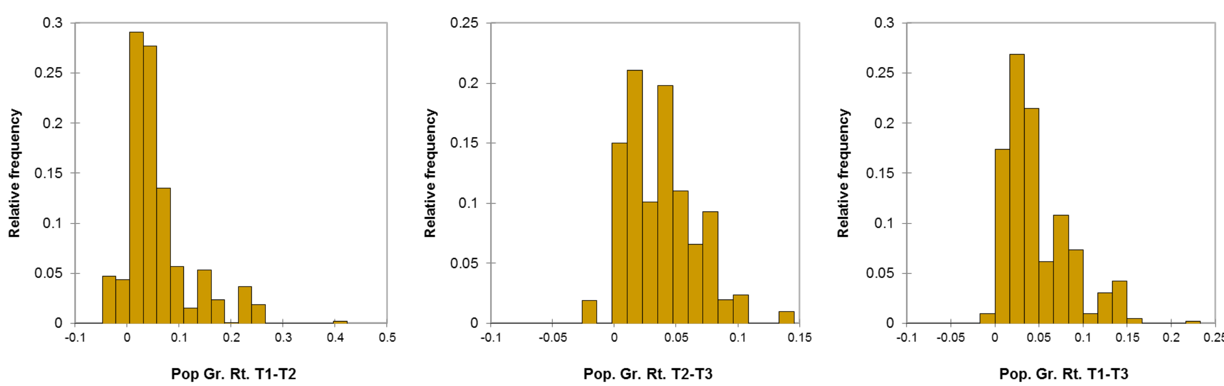


Figure 24. The distribution of differences in the urban extent growth rate, T3T2- T2T1, less developed cities.

Population

Figure 25 shows the weighted distribution of population growth rates in les developed country cities. As with urban extent, the distribution of population growth rates closely mirrors the distribution of the overall sample. Population growth rates appear more tightly clustered than urban extent growth rates.

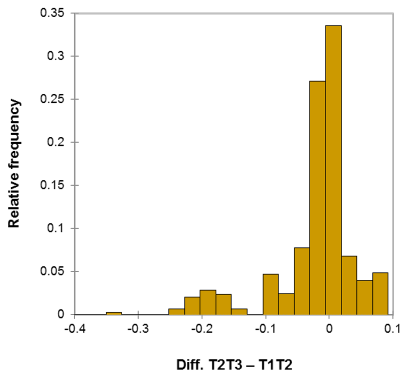
Figure 25: The distribution of population growth rates in less developed cities, T1-T2, T2-T3, and T1-T3.



Average weighted population growth rates across the three time periods are shown in Figure 26. Over the T1-T3 period the average growth rate was 4.7 percent per annum with a 95 percent confidence interval of [4.1%, 5.3%]. The median rate over this period was 3.9 percent. The average rate was not constant over the entire period. The average rate for T1-T2 period was 5.8

percent with a 95 percent confidence interval of [4.6%, 7.1%] and the average rate over the T2-T3 period was 3.8 percent with a 95 percent confidence interval of [3.1%, 4.4%]. The respective median growth rates for these two periods are 4.0 percent and 3.6 percent. Average population growth rates are smaller than urban extent growth rates at all the three periods. The trends suggest a slowing down of the population growth rate over time. An evaluation of the pairwise differences in population growth rates at the city level is needed to confirm whether the decrease in the population growth rate is statistically significant.

Figure 26: The average population growth rate, less developed cities, T1-T2, T2-T3, and T1-T3.



The distribution of differences in population growth rates in less developed country cities is shown in Figure 27. The T1-T2 rate was subtracted from the T2-T3 rate. Values less than zero indicate a lower growth rate in later time period. The distribution appears centered about zero with a left tail. The average difference has a value of -0.021 and the median difference is -0.008. The weighted t-test on the differences confirms that the average is statistically different than zero and that it is negative. The 95 percent confidence interval for the weighted average difference is [-3.1%, -1.0%]. We can interpret this to mean that that population growth rates have slowed down from T1-T2 to T2-T3, by 2.1 percent on average in the sample and between -3.1 percent and -1.0 percent in all less developed country cities.

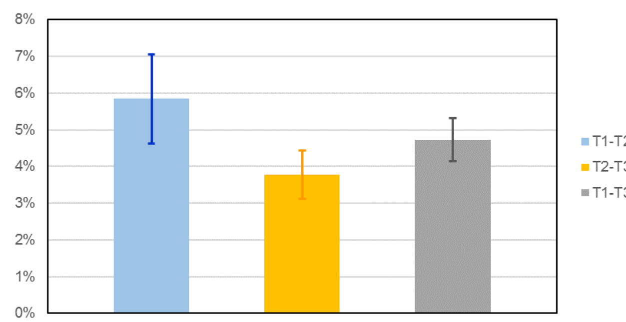


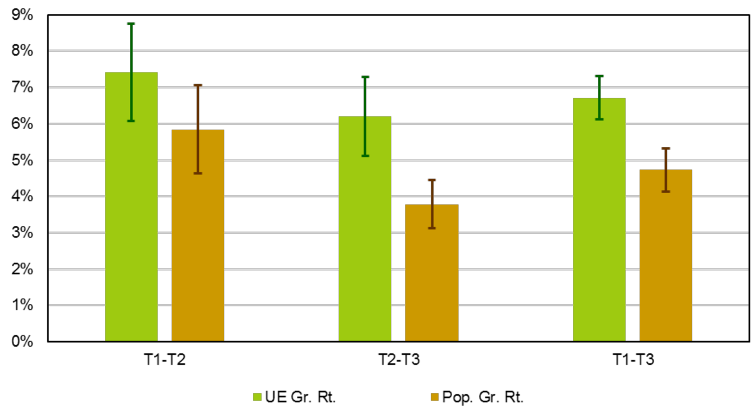
Figure 27. The distribution of differences in the urban population growth rate, T3T2-T2T1, less developed cities.

Urban Extent Growth Rate vs. Population Growth Rate

When we look at the entire sample, we know that cities' urban extents are increaing at faster rates than their populations. We might expect a similar result, and perhaps a more pronounced difference in less developed country cities considering the trends between these two groups.

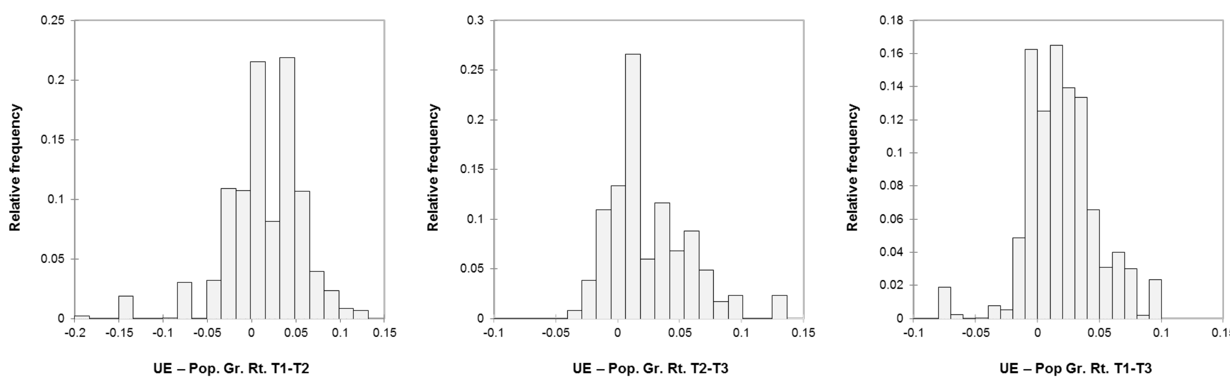
Figure 28 shows the average weighted urban extent and population growth rates, with their 95 percent confidence intervals, side-by-side. The relationship between urban extent and population at each time period and across time periods mirrors the relationship observed in the entire sample, with the notable exception that the values for less developed country cities are greater.

Figure 28: The average urban extent growth rate vs. the average population growth at all periods, less developed cities.



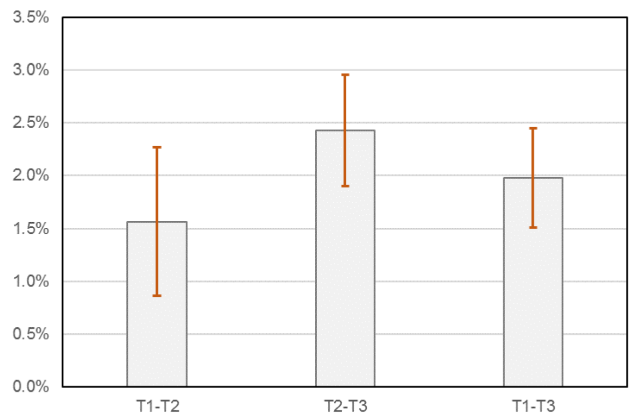
We examine the pairwise differences between the urban extent growth rate and the population growth rate at the city level at a given time period. We subtract the population growth rate from the urban extent growth rate to determine whether the average difference is different than zero. If the difference is positive, the urban extent growth rate is greater than the population growth rate. If the difference is negative, the population growth rate is greater than urban extent growth. The distributions of these differences are shown below in Figure 29.

Figure 29: The distribution of differences, urban extent growth rate vs. population growth rate, less developed cities.



All three distributions are fairly symmetrical centered to the right of zero. Over the T1-T2 period the mean difference is 0.016 compared to a median difference of 0.015; at T2-T3 the mean difference is 0.024 compared to a median difference of 0.015; and over the T1-T3 period the mean difference is 0.020 compared to a median difference of 0.019. One-sample weighted t-tests on the differences, testing the null hypothesis that the differences are equal to zero, confirm that the two rates are different from each other at each time period and that the urban extent growth rate is greater than the population growth rate at the 95 percent confidence level. Figure 30 shows the mean differences across the three time periods and their 95 percent confidence intervals. The overlapping confidence intervals for the difference between the urban extent rate and population rate suggest that the degree to which the urban extent rate growth rate is different than the population growth rate is not statistically different over time.

Figure 30: Average difference of the urban extent growth rate vs. the population growth rate in less developed cities.



<u>More Developed Country Cities</u>

Urban Extent

We now focus on the subset of sample cities in more developed countries. These cities comprise 26 percent of the sample and 33 percent of the universe of cities. Figure 31 shows the weighted distribution of the urban extent growth rates for these cities across the three analysis periods. The ranges of x-axis values correspond approximately the lower third of the x-axis ranges for cities in less developed countries; in other words, the observed rates of urban extent change are much smaller in more developed country cities. Over the T1-T2 period most rates are clustered below 5 percent while over the T2-T3 period most rates are clustered below 2 percent. The right skew of the distributions at all time periods is similar to the pattern observed in less developed countries cities.

Figure 31: The distribution of urban extent growth rates in more developed cities, T1-T2, T2-T3, and T1-T3.

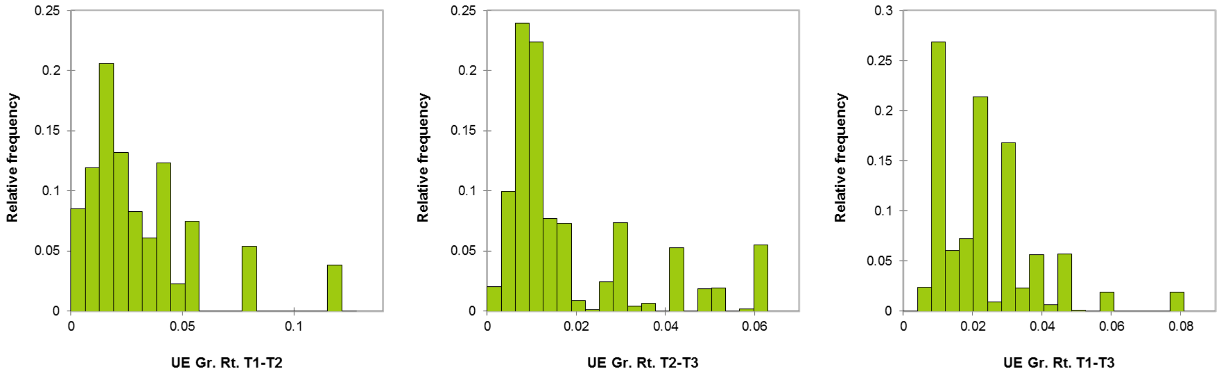


Figure 32 shows the average weighted urban extent growth rates and their 95 percent confidence intervals for cities in more developed countries. Over the T1-T3 period, urban extent grew on average at 2.4 percent per annum with a 95 percent confidence interval of [2.0%, 2.8%]. The weighted median growth rate was 2.2 percent.

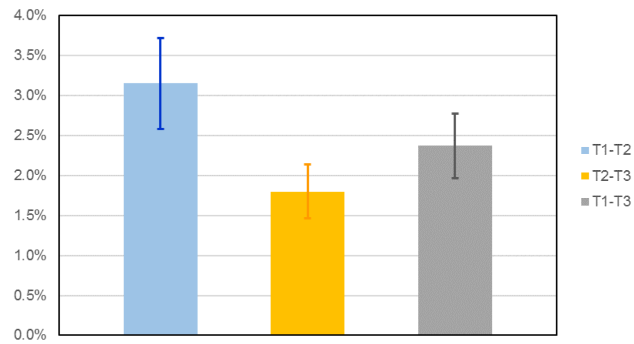


Figure 32. The average urban extent growth rate, more developed cities, T1-T2, T2-T3 and T1-T3.

The point estimate for T1-T2 growth rate is 3.1 percent [95 percent CI: 2.6%, 3.7%] compared to 1.8 percent [95 percent CI: 1.5%, 2.1%] for the T2-T3 period. The respective median values are 2.5 percent and 1.1 percent. The average and median urban extent growth rates are all lower than the rates observed in less developed country cities across all three periods.

As before, we evaluate the city-level differences in growth rates between the two analysis periods to determine whether the average difference is different than zero. We subtract the T1-T2 rate from the T2-T3 rate. If the difference is greater than zero, the T2-T3 rate is greater than the T1-T2 rate, if the difference is less than zero, the T1-T2 rate is greater than the T2-T3 rate. Figure 33 shows the distribution of these differences. The distribution is centered a little to the left of zero. The average difference is -0.013 and the median difference is -0.010. A weighted t-test on the differences confirms that the average difference is statistically different than zero and that it is negative. The 95 percent confidence interval for the average difference is [-2.16%, -0.05%]. We interpret this to mean that the average urban extent growth rate has slowed down over time and that we are 95 percent confident that the difference in the rate is between -2.16 percent and -0.05 percent for more developed country cities.

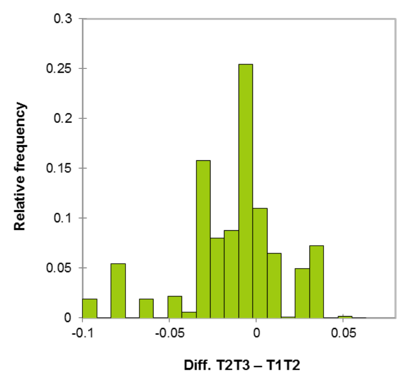
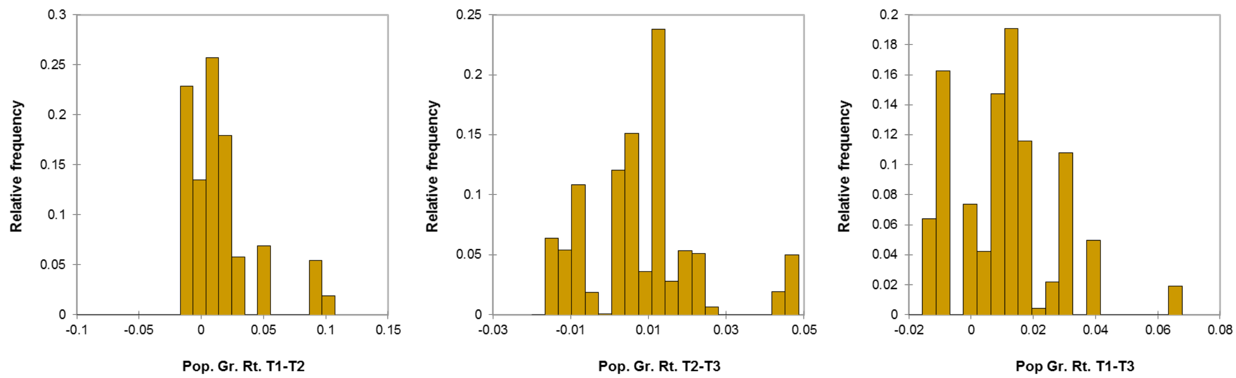


Figure 33. The distribution of differences in the urban extent growth rate, T3T2- T2T1, more developed cities.

Population

Figure 34 shows the weighted distributions of population growth rates in more developed country cities. While we can observe rates of 5 percent per year or higher at each time period, the majority of values are clustered around zero. It appears approximately 22 percent of cities lost population between T1-T3, which can be deduced by the heights of the bars associated with negative values along the x-axis.

Figure 34: The distribution of population growth rates in more developed cities, T1-T2, T2-T3, and T1-T3.



The average weighted population growth rates across the three time periods are shown in Figure 35. Over the T1-T3 period the average rate was 1.1 percent per annum with a 95 percent confidence interval of [0.6%, 1.5%]. The median rate was higher, at 1.3 percent. The average rate was not constant over the entire period. The average growth over the T1-T2 period was 1.6 percent with a 95 percent confidence interval of [1.2%, 2.0%], and the average rate over the T2-T3 period was 0.7 percent, with a 95 percent confidence interval of [0.5%, 1.0%]. The respective median rates were 1.3 percent and 0.7 percent. The trends suggest a slowing of the average population growth rate over time. We evaluate the city-level differences in growth rates between T1-T2 and T2-T3 to determine whether the average difference is different than zero and in what direction.

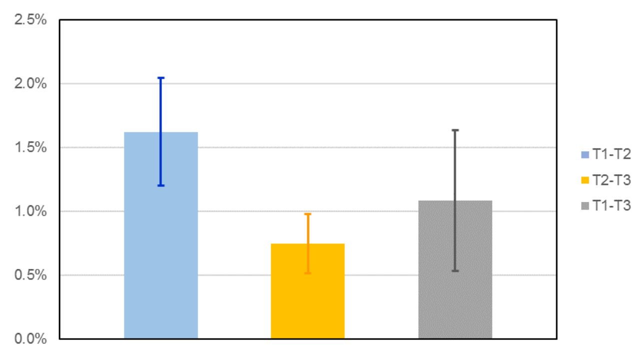
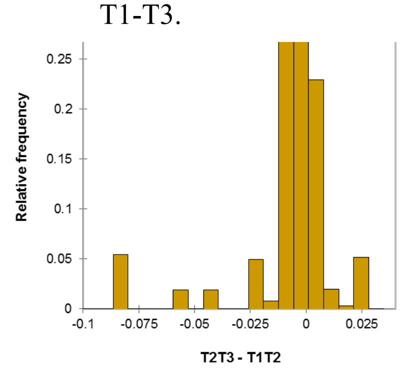


Figure 35. The average population growth rate, more developed cities, T1-T2, T2-T3 and T1-T3.



The distribution of these differences, where the T1-T2 growth rate was subtracted from the T2-T3 rate is shown in Figure 36. They appear centered around zero. The mean difference is -0.009 and the median difference is -0.002. The weighted t-test on the differences confirms that the average difference is statistically different than zero and that it is negative. The 95 percent confidence interval for the weighted average difference is [-1.5%, -0.25%]

Figure 36. The distribution of differences in the population growth rate, T3T2- T2T1, more developed cities.

Urban Extent Growth Rate vs. Population Growth Rate

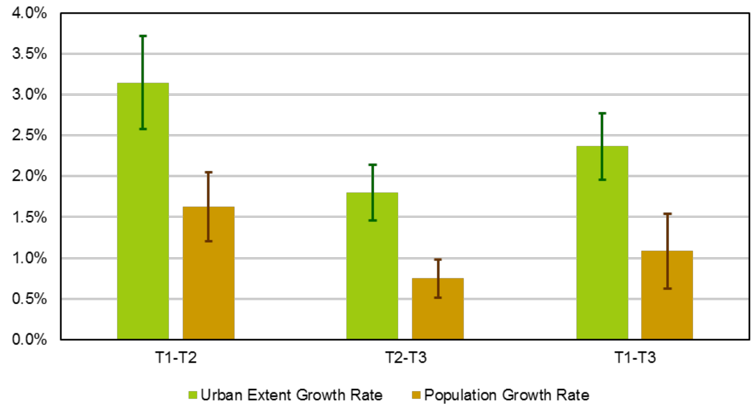
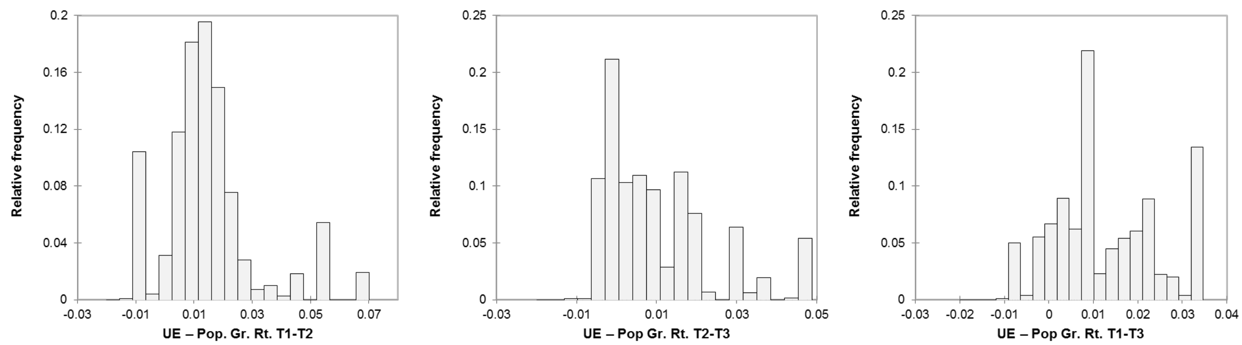


Figure 37. The average urban extent growth rate vs. the average population growth at all periods, more developed cities.

We have seen that the urban extent growth rate is greater than the population growth rate in the analyses for all cities and for cities in less developed countries. We now test whether the same is true for cities in more developed countries. Figure 37, shows the average weighted urban extent and population growth rates for more developed country cities side-by-side with their confidence intervals. The side-by-sde relationship between the average urban extent growth rate and population growth rates observed in all cities and less developed country cities holds true in more developed country as well, with the exception that rates in more developed countries are much smaller.

We examine the pairwise differences between the urban extent growth rate and the population growth rate at the city level at a given time period. We subtract the population growth rate from the urban extent growth rate to determine whether the average difference is different than zero. If the difference is positive, the urban extent growt rate is greater than the population growth rate. If the difference is negative, the population growth rate is greater than the ruban extent growth rate. The distributions of these differeces are shown in Figure 38.

Figure 38: The distribution of differences, urban extent growth vs. population growth rate, more developed cities.



At all three time periods, the differences appear centered to the right of zero. Over the T1-T2 period the mean difference is 0.015 compared to a median difference 0.013; at T2-T3 the mean difference is 0.011 compared to a median difference of 0.007; and over the T1-T3 period the mean difference is 0.013 compared to a median difference of 0.008. One sample weighted t-tests on the differences, testing the null hypothesis that the differences are equal to zero confirm that the two rates are different from each other at each time period and that the urban extent growth rate is greater than the population growth rate at the 95 percent confidence level. Figure 39 shows the mean differences across the three time periods and their 95 percent confidence intervals. We observe the same pattern that we saw in less developed country cities. Although average urban extent and population growth rates have slowed over time, the average difference between the two rates does not appear to be statistically different over time.

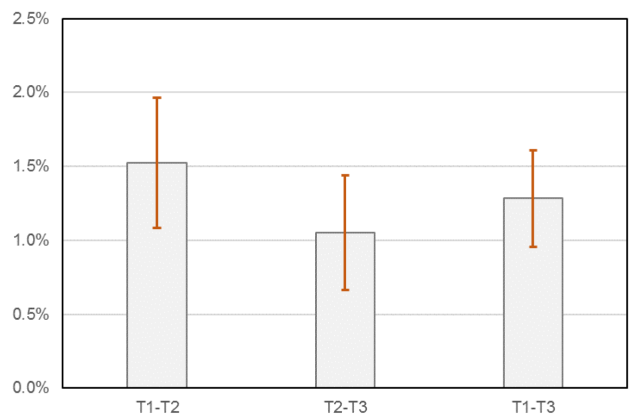


Figure 39. Average difference of the urban extent growth rate vs. the population growth rate in more developed cities

Multiples

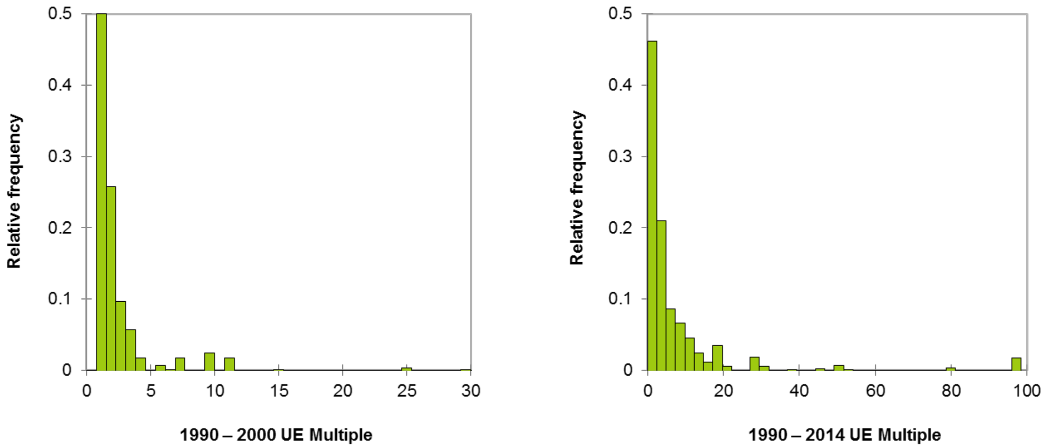
Unlike growth rates based on exponential growth models, which may be difficult to use to understand by how much a quantity grows over time, the concept of the multiple has a much more intuitive interpretation. When we use a multiple to describe the relationship between two quantities, it is the amount the reference quantity is multiplied by to obtain the target quantity. If a quantity grows from a value of 100 to value of 250, its multiple is 2.5. More concretely, the starting value doubled and a half. The multiple is similar to but not the same as total percent change. In this example, the total percent change is 150 percent. Converting the multiple to total percentage change is simply the multiple minus 1 times 100 percent. The same rules apply when the reference quantity loses value. If a quantity decreases from 300 to 180, its multiple is 0.6. More concretely, the target value is only 0.6 times the original quantity. To convert this multiple to total percent change, we take the multiple minus 1 and multiply by 100 percent to obtain a percent change of minus 40 percent.

When we use exponential growth rates, the rate is normalized with respect to the time between observations. If T1 and T2 in one city refer to observations in 1989 and 2000 and T1 and T2 in another city refer to observations in 1988 and 2001, it does not necessarily matter that the total time between observations is different in the two cities. That is because the rate in each city is annualized, and hence we refer to the average annual growth rate. The more similar the actual dates across observations are the better, but we can still reasonably compare annualized rates when the time periods are approximately equal. The situation is different when we use multiples. The multiple is not normalized with respect to time and different lengths of time between observations will result in multiples that are not directly comparable. We address this mismatch by projecting urban extent and population for all cities to three fixed time points: 1 July 1990, 1 July 2000, and 1 July 2014, corresponding to the midpoints of these years. These are the median years associated with all T1, T2, and T3 observations, respectively. The multiples we report are based off of these reference dates.

<u>All Cities</u>

Multiples were calculated for each city for urban extent and population for the 1990 – 2000 period and the 1990 – 2014 period. The distributions of the multiples, for both urban extent and population, are strongly right skew. The weighted distributions of urban extent multiples and population multiples are shown in Figures 40 and 41. The maximum urban extent multiple for the 1990 – 2000 period was 29.8, observed in Suva, Fiji, compared to an average multiple of 2.4 and a median multiple of 1.5. For the 1990 – 2014 period the maximum urban extent multiple was 97.3, observed in Xucheng, China, compared to an average multiple of 7.5 and a median multiple of 2.9. The extreme multiples at each time period pull the average to the right of the median and as a result the average multiple requires a careful interpretation. While the calculation of the average is correct, the median multiple will be a better descriptor of a typical city. For the 1990 – 2014 period, the five highest urban extent multiples, rounded to the nearest whole number were: 97 (Xucheng, China), 81 (Rajshahi, India); 52 (Kozhikode, India); 50 (Vinh Long, Vietnam) and 46 (Hangzhou, China). It is difficult to fathom urban extent multiples above fifty, yet maps and data tables from the *Atlas of Urban Expansion – 2016 Edition*, available for download from htpp://www.atlasofurbanexpansion.org, provide confirmation. *Atlas* map pages for Suva and Xucheng are included in the appendix.

Figure 40: The distributions of urban extent multiples in all cities, 1990-2000, and 2000-2014.



The maximum population multiple for the 1990 – 2000 period was 62.5, in Hangzhou, China, compared to an average multiple of 2.2 and a median multiple of 1.3. If we remove Hangzhou from the analysis, out of an abundance of caution for measurement error, the highest multiple is 13.1, in Palmas, Brazil. For the 1990 – 2014 period the maximum population multiple was 187.3, also observed in Hangzhou, China, compared to an average multiple of 4.0 and a median multiple of 1.9. If we remove Hangzhou out of caution, the highest multiple is 23.5, also in Palmas, Brazil. The x-axis in the 1990 – 2014 histogram in Figure 41 has been censored at 30 to exclude Hangzhou. For the 1990 – 2014 period, the five highest population multiples rounded to the nearest whole number, excluding Hangzhou, were: 23 (Palmas, Brazil); 23 (Rajshahi, India); 18 (Singrauli, India); 17 (Xucheng, China); and 17 (Rawang, Malaysia).

Figure 41: The distributions of population multiples in all cities 1990-2000, and 2000-2014.

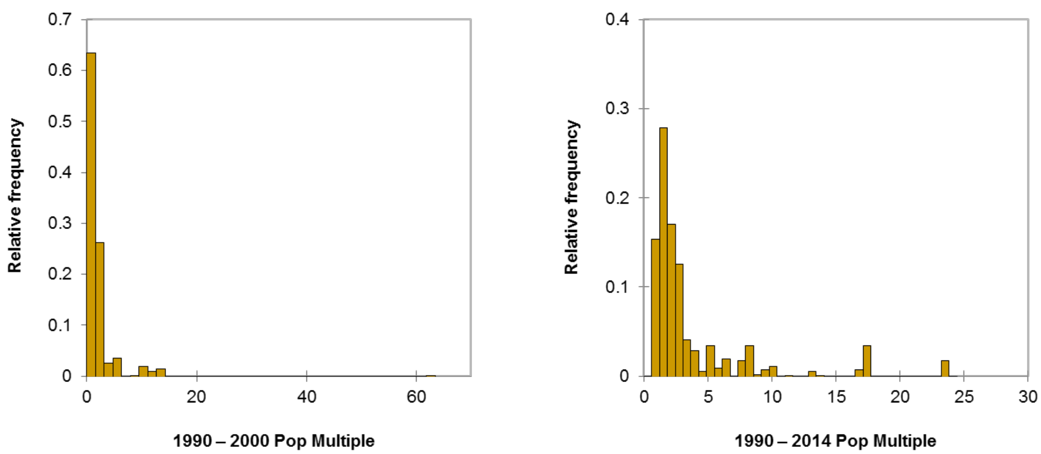
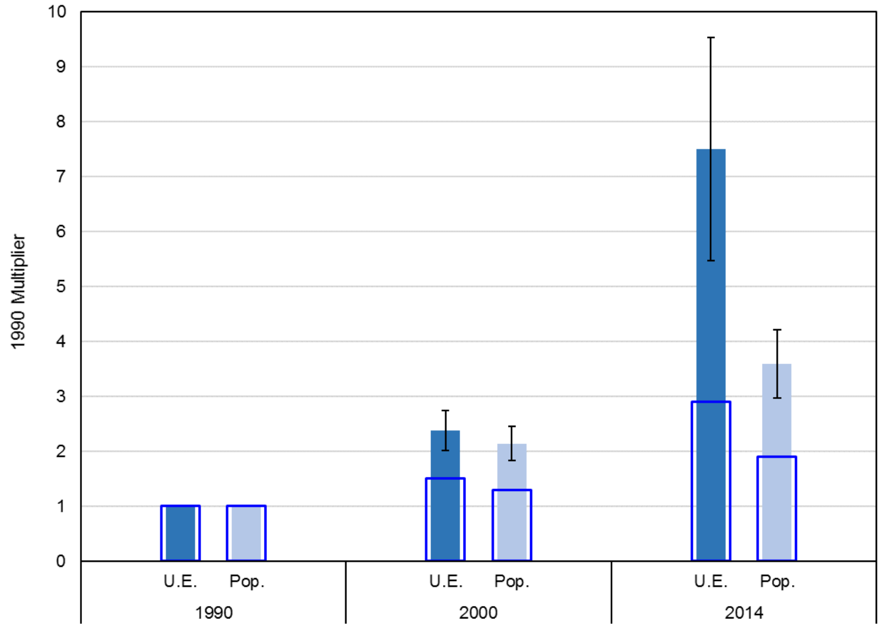


Figure 42 shows the weighted average median urban extent and population multiples for all sample cities, with the 95 percent confidence interval for the average. Average values are displayed in solid bars while median values are the wider hollow bars. Population data for Hangzhou has been removed out of an abundance of caution. The 1990 multiples have values of one as 1990 is the reference year on which the 2000 and 2014 multiple are based. A table of descriptive statistics for the multiples can be found in the appendix.

Figure 42: The average (solid) and median (hollow) values for the 1990 urban extent and population multiplier, all cities.



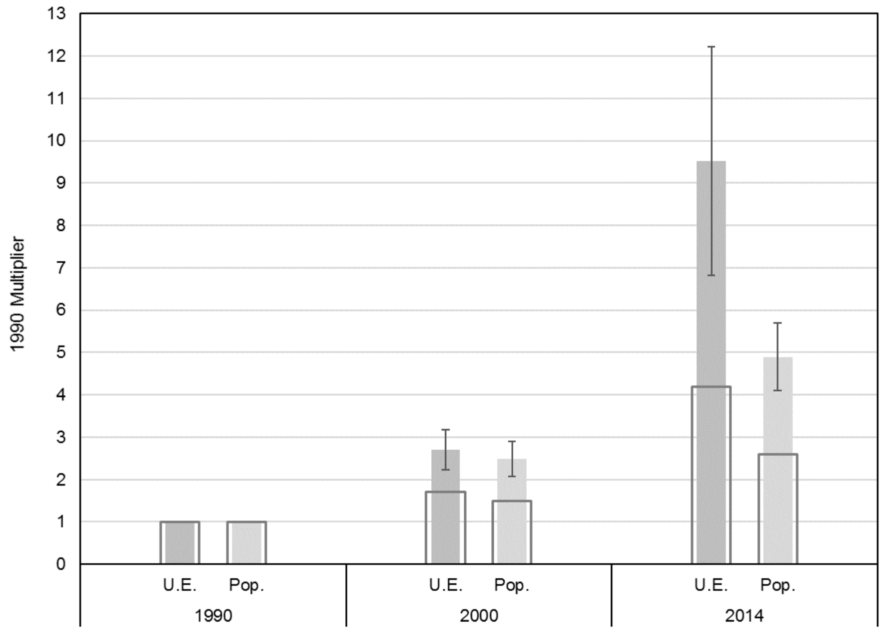
Less Developed Country Cities

The analysis of growth rates showed that trends in less developed country cities are more pronounced versions of the trends in all cities. This carries over to the analysis of multiples. As the distributions of urban extent and population multiples in less developed country cities mirror the those observed in the sample as a whole we do not show them.

Over the 1990 – 2000 period the maximum urban extent multiple was 29.8, in Suva, Fiji, compared to an average multiple of 2.7 and a median multiple of 1.7. Over the 1990 – 2014 period the maximum urban extent multiple was 97.3, in Xucheng, China, compared to an average multiple of 9.5 and a median multiple of 4.2. The top five urban extent multiples listed in the preceding section for all cities are located in less developed countries. We do not repeat them.

Over the 1990 – 2000 period the maximum population multiple was 62.5, in Hangzhou China, compared to an average multiple of 2.5 and a median multiple of 1.5. If we remove Hangzhou out of caution, the highest multiple is 13.1 in Palmas, Brazil. Remoivng Hangzhou again, the maximum multiple over the 1990 – 2014 period the maximum population multiple was 23.5, in Palmas, Brazil, compared to an average multiple of 4.9 and a median multiple of 2.6. Figure 43 shows the weighted average and median urban extent and population multiples for cities in less developed countries, with the 95 percent confidence intervals for the average. Population data for Hangzhou was not included in the creation of the graphic.

Figure 43: The average (solid) and median (hollow) values for the 1990 urban extent and population multiplier, less developed cities.

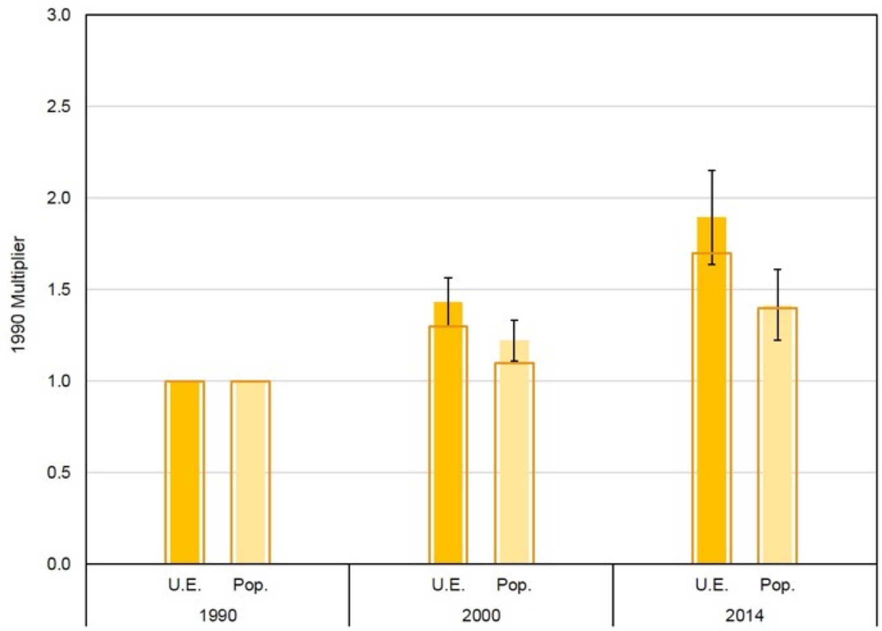


More Developed Country Cities

The distribtions of multiples in more developed countries are more tighly clustered and have relatively small values. The maximum urban extent multiple for the 1990 – 2000 period was 3.3, observed in Springfield, MA in the United States, compared to an average value of 1.4 and a median value of 1.3. For the 1990 – 2014 period, the maximum urban extent multiple was 6.6 observed in Raleigh-Durham, NC in the United States, compared to an average value of 1.9 and a median value of 1.7. The average and median multiples are close together, making the aveage a good descriptor of a typical city.

The maximum population mutiple for the 1990 – 2000 period was 2.7, observed in Raleigh-Durham, compared to an average value of 1.2 and a median value of 1.1. For the 1990 – 2014 period, the maximum multiple was 4.9, also observed in Raleigh-Durham, compared to an average value of 1.4 and a median value of 1.4. The strong similarity of the aveage and median multiples in more developed county cities stands in contrast to large distance between them in less developed country cities. The more developed average and median multpliers are shown in Figure 44.

Figure 44: The average (solid) and median (hollow) values for the 1990 urban extent and population multiplier, more developed cities.



Summary

This section presented evidence on the change in urban extent and population in the 200 city sample as a whole, in a subset of less developed country cities and in a subset of more developed country cities. Each city in the sample represents cities in its sampling sub-stratum, or its sampling box. The results we have discussed for all cities, less developed country cities, and more developed country cities are weighted results that adjust the findings of an individual sample city to account for the number of cities it represents.
From the T1-T2 to the T2-T3 period, we found that both the average urban extent growth rate and the average population growth rate declined in all three categories. These declines were statistically significant. There is considerable variation in the magnitude of growth rates across the categories. For the T2-T3 period, for example, the average and median urban extent growth rates in less developed country cities were 6.2 and 5.7 percent per year compared to rates of 1.8 and 1.1 percent per year in more developed country cities. The average and median population growth rates over this period were 3.8 and 3.6 percent per year in less developed country cities compared to 0.7 and 0.7 percent per year in more developed country cities. Are cities' urban extents growing at faster rates than their populations? Our statistical tests suggest that on average the answer is yes and that this was true for every analysis period in our study, in both less and more developed country cities, even though urban extent and population growth rates have declined.

The multiple was introduced as a more intuitive way to understand and communicate change over time compared to the average annual growth rate. Average and median multiples for the 1990 – 2000 and 1990 – 2014 periods may not be consistent with average and median growth rates over the T1-T2 and T1-T3 periods owing to the manner in which the multiple was constructed. We believe the median multiple is a more reliable measure and more directly comparable to the median growth rate. For example, the median urban extent multiple for all cities for 1990 - 2014 was 2.9; namely the typical city increased its area almost three-fold over the 24-year period. The median urban extent growth rate for all cities over the T1-T3 period was 4.2 percent per year. A quantity that grows 4.2 percent per year increases by a factor of 2.8 over 24 years. Similarly, the median 1990 – 2014 urban extent multiple for less developed country cities was 4.2; namely, the typical city in this category slightly more than quadrupled in size over the 24-year period. The median urban extent growth rate for these cities was 6.0 percent per year. A quantity that grows at that rate increases by a factor of 4.3 over 24 years.

Modeling Urban Extent and Expansion

The previous sections have addressed how spatial and population data for urban extent was collected, how we measured and defined it, how it is distributed in the 200-city sample, and how it has changed over time, based on 600 observations spanning a 24-year period. A quantitative understanding of the amount of land cities occupy, and its change over time, is essential for the promotion of evidence-based policies for planning and managing urban growth. When cities grow and expand, new urban lands must be properly serviced—with trunk urban roads that carry urban transport, with sewer and water systems, with public open spaces, for example—to be of optimum use to their inhabitants. Cities that can secure adequate lands and properly service them in advance of development are likely to become more productive, efficient, and sustainable than those that do not. It will be challenging to make adequate, forward-looking plans if there is no information for how fast cities are expanding, how fast they have expanded, or if there is no understanding of what the dominant trends are. As data collection methods and data processing capabilities improve, larger and more carefully constructed samples will refine our understanding of urban expansion in cities of different sizes, in different income categories, in different regions.

Measuring and reporting on urban extent and expansion should be repeated over time to update and improve the existing knowledge base. Planners and policy makers would be able to better plan and prepare for urban expansion, however, if there was accurate and contemporary information on the different forces that drive it. In this section we transition from an analysis that is primarily descriptive to one that is inferential. Whereas we have focused on describing the size of cities' urban extents and the rate at which they change over time, we now shift our attention to the factors that explain the changes we observe in the 200-city sample and the relative contributions of these factors to urban extent and expansion.

There are a multitude of analyses that have looked at urban land cover and urban expansion at different geographic scales. A 2010 meta-analysis identified no fewer than 326 case studies that used remotely sensed data to map urban land conversion at the metropolitan or regional scale, and that number has surely risen since then (Seto et al. 2010). Despite the large number of studies, the use of inferential statistics to understand how different independent variables can explain observed urban land cover and urban expansion, and how these relationships might be generalized to larger groups of cities, is seldom found in the academic literature (Burchfield et al. 2006; Angel et al. 2005; Seto et al. 2010; Angel et al. 2011). There are relatively few studies that employ regression analysis, for example, to quantify how the dependent variables of urban land cover and urban expansion, measured at the city level, relate to explanatory variables pertaining to the demand for land by households and firms, geographic constraints and environmental conditions, or policies and programs that affect travel behavior and the regulation of land and housing.

The reasons why this type of analysis is lacking may be due to diverse data issues, both on the on the side of the dependent variable and on the side of the independent variables, as well as a general tendency of researchers involved in the remote sensing of land to come from fields outside of urban planning, or the social sciences more broadly, though this is changing (Donaldson and Storeygard 2016).

On the dependent variable side, there are problems of sample size. While there may be many individual studies that map and measure change in urban land, these studies typically focus on one city only, one metropolitan region, or a small group of cities over a well-defined region. Combining information from various studies to obtain a larger sample size can lead to mismatches in the way cities are defined, in the spatial resolution of the underlying imagery, in the techniques used for land cover classification, and in the analysis dates. Simply put, data from different studies may not be directly comparable. Even when comparable data is assembled for a group of cities, there are questions about the generalizability of the dataset. Cities are not homogenous units; some are large and some are small, some are rich and some are poor, some are surrounded by water and mountains while others are surrounded by flat buildable land. A consequence of this heterogeneity is this is not all random city samples are equal. If the rationale of a sample-based study is to draw generalizations about the population from which it was drawn, samples must be reasonably representative of the populations they represent and they must be reasonably large. Sample based studies do exist in the literature but they are rare and sample sizes are relatively small (Schneider and Woodcock 2008; Angel et al. 2012; Taubenbock et al. 2012). Sample based studies should become more frequent in the future as new global land cover products become available, the time and effort required for data processing decreases, and the use geographic information system software in the urban sciences increases.

On the independent variable side, data availability and the spatial resolution at which data is collected are impediments to the development of models for large heterogenous samples based on cities in different countries. While there is plentiful social and economic data available at the global level from resources like the World Bank's World Development Indicators (WDI), these databases contain information reported at the national level. National level data is sometimes meaningful for inclusion in models that are fundamentally about cities, but more often than not it leads to analyses that become less about cities and their unique characteristics and more about the characteristics of countries to which they belong. There is no analogue to WDI for the world's cities. Our sample contains 34 observations in China, 17 in India and 14 in the United States. In each of these countries the cities in question have different populations, different economic profiles, different geographies, and different rules and regulations that may be determined at the local level. Ideally, independent variables in a study about cities would be collected at the same spatial level as the dependent variable. This is extremely challenging, especially when the sample is global in its composition. It may be possible to devote the time and resources to collect detailed city level data for a small group of cities, but this leads to a tradeoff in sample size and the generalizability of the findings. In our case, this would be further complicated by the fact that urban extent boundaries do not conform to statistical reporting units and some degree of spatial mismatch between the urban extent and the reported value would be practically unavoidable.

The situation is less challenging in countries or regions with robust data programs, or more generally, in places with a strong tradition of data collection and reporting. Detailed socioeconomic data may be publicly available at fine geographic resolutions in the United States and in many OECD countries but this is the exception rather than the rule. Perhaps data at fine geographic resolutions exists in other countries but it is not publicly shared, or it not shared with individuals outside the country. We experienced this during the collection of spatially referenced population data in certain countries. The situation is also less challenging when the variable in question involves information that can be remotely sensed or monitored as opposed to information that requires local informants or *in situ* data collection. Currently, remote sensing methods are typically used to collect information about geographic conditions and land uses, including elevation data and information about land cover categories, such as areas in cultivation. Perhaps in the future new data collection methods based on crowd sourcing, or big data approaches applied to mobile phone and internet data, will facilitate what information we can collect at the city level. Until that time, obtaining globally representative data at the city level, on regulations, programs, and local conditions related to the economy, land use, housing, and transport – variables which there is reason to believe might be important for explaining urban extent and expansion – remains a serious challenge.

Theories and Hypotheses

There are many potential factors that could explain the amount of area cities occupy and its change over time. These include economic factors such as household income, the costs of transport and housing, and the marginal productivity of land in different uses; geographic factors, such as steep slopes and water that limit outward growth or climactic conditions that make construction easier; social and cultural factors, that may lead people to live further away from each other or that explain different preferences and attitudes for housing and lifestyle; or political and institutional factors, that result in policies, regulations, and enforcement of those policies and regulations, that affect land, housing, and transport. Surely other potential factors may exist. Nothing has been said of population as a factor, though it is probably the most intuitive of all; as cities grow in population they upwards and outwards.

Developing theories about the factors that explain urban extent and its change over time is important, but to move these theories from intellectual exercises to testable hypotheses is also important, especially in the context of evidence-based policy for managing urban growth. This is a point at which international research on cities must confront data limitations that effectively constraint what the breadth and depth of testable hypotheses can be.

As they relate to planning for and managing urban growth, the hypotheses can be divided into two broad categories. First, how is amount of area occupied by cities influenced by factors generally assumed to be outside the realm of policy intervention? What is the relative influence of these factors in the sample of cities? This includes things like population growth and geography. Second, how do variables that may be influenced by public policy influence urban extent and expansion and what is the relative influence of these factors on observed otucomes? This includes things like the cost of transport and regulations on land and housing.

The point of departure for our hypotheses is the standard economic model of urban spatial structure, developed by Alonso (1964), Mills (1967), and Muth (1969) later refined by Wheaton (1976) and restated by Brueckner (1987). The model and its extensions yield a number of testable hypotheses on which our analysis is based. These include:

1. Other things being equal, the larger the population of a city, the larger its urban extent and the larger the increase the larger its expansion;
2. Other things being equal, the higher the average income in a city, the larger its urban extent, and the larger the increase the larger its expansion;
3. Other things being equal; the higher the share of buildable land available for housing and development, the larger its urban extent and expansion;
4. Other things being equal, the higher the price of agricultural land on the urban periphery, the smaller the urban extent
5. Other things being equal, the higher the cost of commuting in the smaller its urban extent,
6. Other things being equal, an increase in the world price of the export good will increase urban extent and expansion

We have tested these relationships in the 200-city sample using data that is described in the following section. To the above hypotheses we also added a seventh, related to regulatory red tape, or the overall stringency of the rules and regulations governing land and housing in cities, Other things being equal, the more stringent the rules and regulations governing land and housing, the smaller the urban extent. The city-based focus in the collection of explanatory variables for these models makes them the second generation of models of urban extent and expansion described in Angel et al. (2005).

Data for Model

<u>Average Income</u>

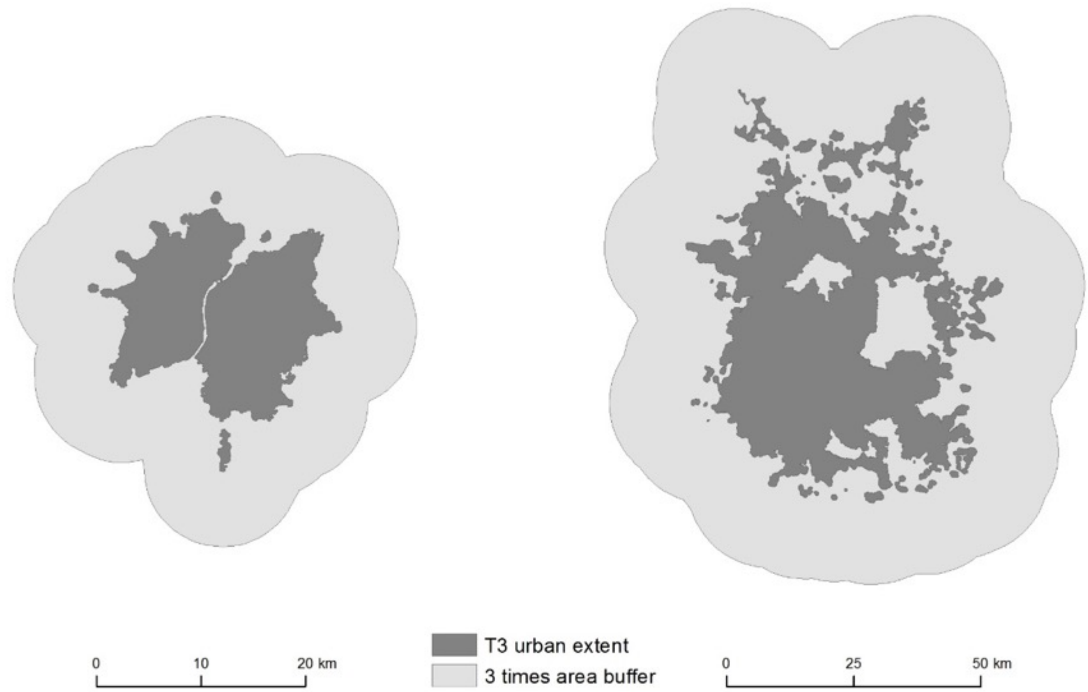
Time-series data for average income is readily available at the country level, reported as GDP per capita, but city-level measures of average income are much less common, especially in non-OECD countries. In the United States, the Bureau of Economic Analysis provides GDP per capita estimates for 382 metropolitan statistical areas. The OECD's statistical database provides GDP per capita measures for 281 metropolitan areas with populations over 500,000 in 29 OECD countries. Collecting GDP per capita data at the city level for all 200 cities presented numerous challenges as the sample contains 63 cities with populations under 500,000 and two-thirds of sample cities are in non-OECD countries. Moreover, the urban extent boundaries we developed do not conform to statistical reporting units. Even for sample cities with entries on the United States and OECD databases, we would not obtain a one-to-one spatial match between the boundary of the reported value and the boundary of the urban extent.

Since the average income - urban extent relationship that we want to test is spatially explicit, we sought city-level measure of GDP per capita for all 200 sample-cities as opposed to country level measures. We could obtain city-level estimates of GDP per capita (PPP) for all but two sample cities from the McKinsey Global Institute's CityScope database, a proprietary product. The database contains income information for 2,910 cities and metropolitan areas with populations of at least 150,000 in 2012. The two sample cities absent from the database were Gaoyou, Jiangsu Province, China, and Dzherzhinsk, Russia. Using the data for all cities within each of these countries, we developed models to estimate GDP per capita for these two missing cities. While the McKinsey average income estimates are city-based they are not explicitly spatial at least not in the database; we never knew exactly over which areas the estimates corresponded. We could gauge the agreement between their definition of a city and our own by comparing our urban extent populations to their reported city populations. Differences were small enough to warrant the assignment of GDP per capita estimates to the sample cities. For countries with multiple cities in the sample, such as China (34), India (17), the United States (14), Brazil (8), Russia (6), and many others, we have heterogenous estimates of average income for each city observation within the country; in China for example, the difference between the minimum and maximum value varies by a factor of 10.

Buildable Land

When the slope of land rises above a certain threshold it becomes unbuildable, or at the least, more difficult to build upon. The buildability threshold may vary by location according to the composition of the terrain, climactic conditions, building materials, accepted norms, or laws. To measure buildable land as it relates to models of urban extent and urban expansion, we focus on a city's expansion area. Built-up area within the urban extent or outside the urban extent, regardless of the slope of the underling terrain is buildable land – the built area already exists. We define the expansion area, or the analysis area, as the area surrounding the city that contains exactly three times the area of the urban extent. This analysis area is typically open space, but it may contain built-up areas or water bodies. Expansion areas for the buildable land analysis were automated using a Python script. Regardless of the shape of the urban extent of the number of polygons that comprise it, the script produces an analysis area that contains three times the urban extent within one percent error range. In Figure 45, below, we see the urban extents and expansion areas for the buildable land calculations in Ahmeadbad, India and Mexico City, Mexico.

Figure 45: The expansion areas of Ahmedabad, India (left) and Mexico City, Mexico (right).



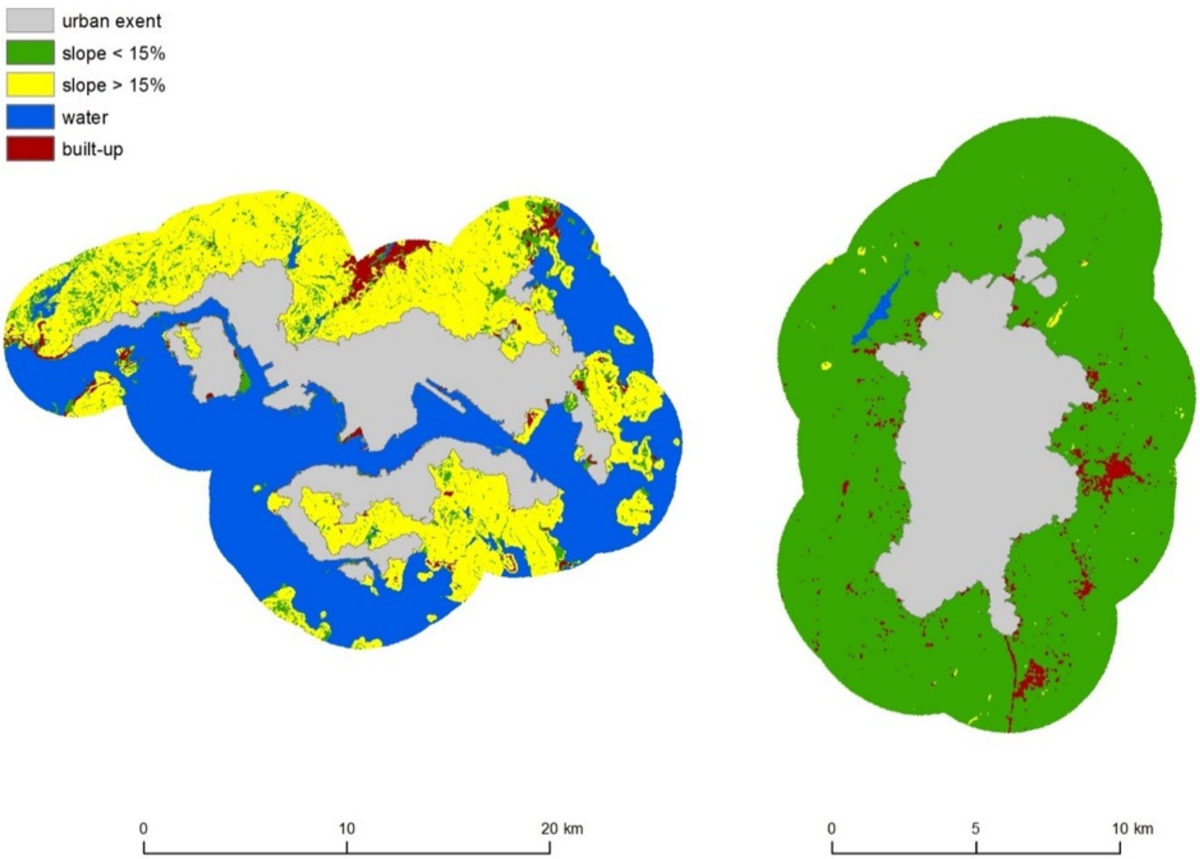
We apply a slope threshold of 15 percent rise, or 8.53 degrees, to calculate the share of buildable land within the analysis area. This threshold was chosen after conversations with builders and real estate professionals in the United States where slope values greater than a 15 percent rise are associated with increased development costs. It is possible to build on steeper slopes, of say 20 percent or even 50 fifty percent, but these raise development costs substantially and often require complex engineering solutions.

Buildable land was calculated by combining information from two datasets: NASA's Shuttle Radar Topography Mission (SRTM) dataset, which contains a digitial elevation model (DEM) and a water file, and the European Commission's Global Human Settlements (GHS) built-up grid. The GHS dataset contains built-up area information for the entire planet at 38m spatial resolution for the year 2014. SRTM data is 30m resolution and contains elevation data for the entire planet based on information collected in the year 2000. SRTM pixels were resampled to match the GHS resolution. The three pixel categories were mosaiced in such a manner that GHS pixels lay on top of water pixels, which lay on top of elevation pixels. Slope was calculated on this bottom layer of non-GHS, non-water pixels, using the ArcPy Python package. The buildable land ratio for a city is the area within its buffer with slope less than 15 percent divided by the total area of the buffer. In Figure 46, below, we can see two extreme examples of the buildable land ratio. Hong Kong, on the left, is surrounded by steep hillsides and water and has very little buildable land (7.1%) while Oyo, Nigeria on the right, is surrounded by flat open land, resulting in plentiful buildable land (96.3%). Green areas correspond to land with slope less than 15 percent and yellow areas to slope greater than 15 percent.

Figure 46: Buildable land in the expansion areas of Hong Kong (left) and Oyo (right).

Airport Score

Airport names and locations and information for the airline connections between airports were used to create an origin-destination matrix for the world's airports. The dataset used to create this matrix pertains to airport network linkages in June 2014 and was purchased from the website



www.openflights.org.The dataset lists 3,316 airports in 214 countries and contains over 65,000 unique origin-destination pairs. An origin-destination pair occurs at the airport level.

While the airport connectivity matrix provides information about the total number of connections associated with individual airports, not all airport connections carry equal weight in practice. A single connection to a large airport with many connections is ostensibly more important than a single connection to a small airport with few connections. For the purpose of developing a connectivity measure, therefore, a node (airport) within a network should carry greater importance if it is connected to other influential nodes. We incorporate this idea into our measure of airport connectivity, or airport centrality, by adopting the graph theoretical measure of eigenvector centrality. The eigenvector centrality analysis produces a measure of relative scores that can be used to rank the 3,316 airports in our dataset. The five top ranked airports align with what we might expect from such analysis: Atlanta's Hartsfield-Jackson International (ATL) takes the top ranking, followed by London Heathrow (LHR), Chicago's O'Hare International Airport (ORD), John F. Kennedy International Airport (JFK) in New York, and Los Angeles International Airport (LAX). The distribution of the relative scores is highly skewed to the right, with a small number of very important airports with high scores and a large number of airports with low scores

Relative scores are associated with the lat-long locations of airports. *City-based* airport scores were created for each of the 200 sample cities by adopting a gravity model formulation shown below:

$$C_i = \sum_{j=1}^{n} \frac{A_j}{\left(d_{ij}\right)^{\beta}}$$

where the city score C for city i is the sumproduct of the relative airport scores A for airports j weighted by the inverse distance between city i and airport j. Distance was calculated using the lat-long associated with a city's CBD location. Thus the city-based airport scores for the 200 sample cities incorporate the relative scores of all 3,316 airports. We experimented with different values for the β term and decided to calibrate it using a ranking of cities developed by Globalization and World Cities (GaWC) research network (2018). We obtained very high correspondence between our ranking and GaWC's: the highest ranking 40 cities by our method contained 84 percent (26 out of 31) of top ranked cities defined by GAWC. The high correspondence suggests that the city-based airport score may be a meaningful proxy for a global economic connectedness and the price that domestically produced export goods are likely to receive in global markets. There is evidence for this

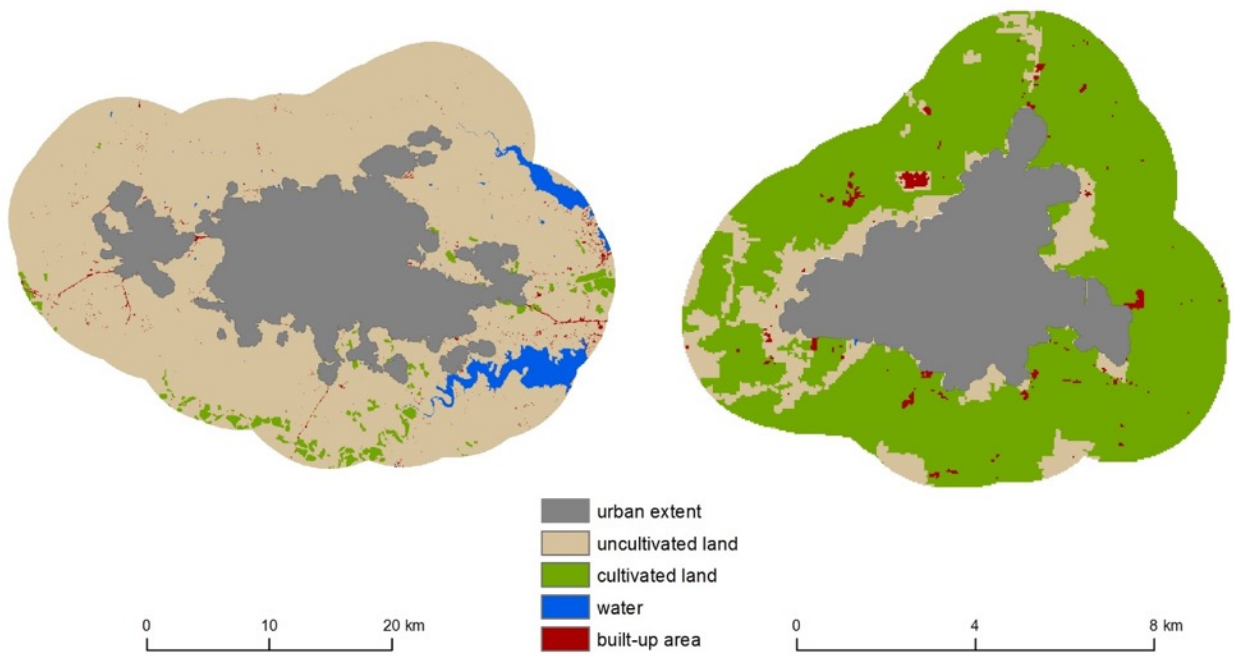
Agricultural Land

Cultivated Land in the Expansion Area

Spatially explicit data for land in agricultural use around is now available from global datasets produced by the National Geomatics Center of China (GlobeLand30) and the Global Food Security Analysis Support Data GFSASD), a project led by the United States Geological Service. When we undertook the analysis we were only aware of the Chinese dataset and learned of the GFSASD product when our analysis was nearly complete. The GlobeLand30 technical documentation states an accuracy of 83% for the cultivated land class. We are unaware of the accuracy of the GFSASD dataset or how using the two datasets might lead to different conclusions.

The GL30 product is 30m resolution and contains information for nine land cover classes. The cultivated land class is described as "Lands for use in agriculture, horticulture, and gardens, including paddy fields, irrigated and dry farmland, vegetation and fruit gardens, etc…" We downloaded data for the cultivated class for the entire planet and clipped the data by the expansion area boundaries used in the buildable land analysis. In Figure 47 below, we observe cultivated land in the expansion areas of Killeen, Texas on the left with only 3 percent of its expansion area in cultivated land and Leon, Nicaragua on the right, where 83 percent of the land in the expansion area is in cultivation.

Figure 47: Cultivated land in the expansion areas of Killeen, Texas (left) and Leon, Nicaragua (right).



Buildable Land That Is Cultivated

We determine the share of buildable land that is cultivated by combining information from the SRTM and GlobeLand30 datasets. From the buildable land analysis we identify all pixels within the expansion area that are neither built-up nor water. We assign these pixels a slope value. We determine which of these pixels is cultivated by resampling the cultivated land data to match the resolution and pixel grid associated with buildable land analysis. When we overlay the cultivated land data on the buildable land data, we obtain four class of non-built, non water pixels: (1) slope less than 15 percent cultivated, (2) slope less than 15 percent not cultivated, (3) slope greater than 15 percent cultivated, and (4) slope greater than 15 percent not cultivated.

Agricultural Land per Capita

Agricultural land refers to the land area that is under permanent crops, permanent pastures, or land that is arable. Approximately one third of agricultural land is arable land, defined by the Food and Agriculture Organization as land in temporary crops, temporary meadows, or land that is temporarily fallow. Total agricultural land in the country was obtained from the World Bank's World Development Indicators (WDI). We calculated 2014 per capita estimates using country population estimates from the WDI.

Value of Agricultural Land

The statistical database of the Food and Agriculture Organization (FAOSTAT) contains information for the production value of agriculture which is disaggregated by agricultural item by country. Production value is defined as gross production in physical terms multiplied by output prices at the farm gate. A measure of the average value of a hectare of agricultural land within a country was calculated by dividing the total production value by the land in agricultural use.

Building Regulations

A companion project on the 200 city sample by the research team, *The Land and Housing Survey in a Global Sample of Cities*, involved the completion of surveys by local experts on the rules and regulations governing the development of new residential land and housing and their enforcement. No single variable from the regulatory survey adequately summarizes the overall regulatory regime within a city. We tried creating new variables out of existing ones but these efforts proved unsuccessful for explaining urban extent and expansion. The variable we have chosen to use pertains to the subdivision of land on the urban periphery. More specifically, we use the typical time need to obtain the necessary permits for a 200-unit land subdivision on land already converted to urban use. Values range from a minimum of one month in Kampala, Uganda to 60 months in Taipei, Taiwan with an average value of 12 months for all 200 cities. This variable is used as a proxy for the regulatory red tape surrounding new development.

Climate

A revised Koppen-Geiger climate map was used to assign a climate classification to each of the sample cities (Peel Finlayson and McMahon 2007). While the Koppen-Geiger system contains as many as 29 sub-classifications of climate based on combinations of climate group, precipitation type, and heat level, we only assigned one of the five main climate groups: tropical, arid, temperate, continental, and polar to each sample city, based the intersection of the CBD point location with climate group boundaries.

Gasoline Price and Vehicle Ownership

In the absence of global data for local gasoline price prices at the city level, we relied on average national level per liter pump prices for 2014 from the World Bank's WDI dataset. As the WDI contains price information for both gasoline and diesel prices, we attempted to construct a weighted average fuel price using information for the shares of a country's passenger vehicle fleet that is gasoline vs. diesel based. We could not construct these shares for all countries owing to the limited public availability of necessary data. The International Energy Agency collects this information but does not make it freely available to researchers. A report by the Global Fuel Economy Initiative claims that gasoline vehicles represent the majority of cars sold in every country, except for a handful where diesel sales are higher, and Brazil, where flex fuel vehicles dominate (Fulton Jenn and Tal 2017). For this reason we chose to use the gasoline price.

Vehicles per thousand population, or the motorization rate, refers to all registered vehicles on the road in a country normalized by the country's population. Motorization rates for 2014 were obtained from the International Organization of Motor Vehicle Manufacturers (OICA), which aggregates data from national trade organizations, OICA members, national statistics offices, and ministries of transport. The motorization rate calculation uses United Nations population estimates for countries.

Model Result

We tested our hypotheses using models where (1) observed urban extent, measured in hectares and corresponding to the most recent analysis period, and (2) the annualized urban extent growth rate between the T2 and T3 periods, being a normalized measure of urban expansion, were the two dependent variables. We employed log-log, or loglinear models, where both the dependent variables and the independent variables have been log transformed and the independent variables are linear in their parameters. This type of model specification allows for a clear and convenient interpretation of the regression coefficients. It is also called a constant elasticity model because the elasticity of the dependent variable with respect to the independent variable does not change as value of the independent variable changes. More precisely, the coefficients tell us about the percent change in the dependent variable when the independent variable increases by one percent.

Output for models with urban extent and urban expansion as the dependent variable can be found in Tables 2 and 3 respectively. The estimated models are presented in a single column with blank spots indicating that a particular independent variable, listed in the left-hand column, has been excluded from that model's particular specification. There are nine models of urban extent and five models of urban expansion. Parameter estimates are listed first, followed by the standard error of the estimate, followed by the associated p-value, indicating whether that variable was found to be a statistically significant predictor. Statistical significance at the 5 percent level is indicated with a single asterisk, and at the 1 percent level with three asterisks.

<u>Logarithmic Models of Urban Extent</u>

Different iterations of the model of urban extent are tied to different categories of variables and different hypotheses. Model 1 looks at population as the single explanatory factor. The finding is powerful as it suggests that urban extent is first and foremost a function of a cities populations The adjusted R-squared value of 0.75 reveals that variations in population explain 75 percent of the variation in urban extent. The parameter estimate of 0.85 can be interpreted as follows: a one percent increase in population is associated with 0.85 percent increase in urban extent, or alternatively, a 10 percent increase in population is associated with a 8.5 percent increase in urban extent, or a 100 percent increase in population is associated with an 85 percent increase in urban extent.

Model 2 introduces information for average income at the city level. The explanatory power of the model increases and with only two variables we can explain 85 percent of the variation in urban extent. A 10 percent increase in average income at the city level is associated with 5.4 percent increase in urban extent. The relative contribution of population to explaining urban extent decreases slightly, a 10 percent increase in population is now associated with a 7.8 increase in urban extent.

Model 3 introduces geographic variables. There are 199 observations because we could calculate slope for St. Petersburg, Russia owing to SRTM data constraints. Adding information for the share of buildable land in a city's expansion area as well as information for climate barely improves the model's predictive power even though both are statistically significant. The direction of the sign on buildable land is as expected, namely when the share of buildable land increases by 10 percent, urban extent increases by 2.7 percent, all things being equal. We have used the 15 percent slope threshold to classify buildable land but it is likely that different thresholds would affect the parameter estimate. It is also possible that the area over which the buildable land ratio was calculated may be too large to capture the effect buildable land has on urban extent. We may experiment with different buildability thresholds and different sized analysis areas in the future to understand how topography influences urban extent and expansion. For the climate variable, we use the temperate climate category as the reference group. The interpretation of the coefficient is that all things being equal, cities in temperate climates have 34 percent ($e^{0.29} - 1$) larger areas than cities in non-temperate zones. This finding is a little surprising as we expected tropical climates would be more hospitable to building and construction and would be associated with larger extents, all things being equal.

The fourth model is composed of three submodels, 4.1, 4.2, and 4.3. that are all tied to agricultural hypotheses. Population and income are included but the geographic variables from model 3 have been excluded. If agricultural land exists on the periphery of the city, it creates competition for land between agricultural users and urban users. We first identify the share of buildable land in the expansion area that is cultivated. This variable combines information for buildable land as well as information about the area in agricultural use. All things being equal, we would expect that as the share of the area in cultivation increases, pressure against urban extent and expansion increases. Moreover, if the value of the area in cultivation is higher, namely, when the marginal productivity of agricultural land increases, we would expect even greater pressure against urban extent and expansion. We account for this by looking at the average value per hectare of agricultural land at the country level and dividing it by the city's GDP per capita. This is an estimate of how many average incomes it takes for a city dweller to purchase a hectare of agricultural land. When this ratio is low, agricultural land is relatively cheap and the influence of agricultural land should be less than when the ratio is high. If there is plentiful agricultural land in the country, however, measured as total agricultural land per capita, then perhaps neither the location of the agricultural land on the periphery of the city nor its average value matters much since agricultural uses can locate in many possible locations, thereby avoiding competition with urban uses.

Model 4.1 tests these three agricultural variables together, model 4.2 tests agricultural land per capita and the value of a hectare of agricultural land relative to average city income, and model 4.3 tests agricultural land per capita only. In all three models, only agricultural land per capita is significant and the parameter estimates for this variable across all three models are similar: a 10 percent increase in the agricultural land per capita at the country level is associated with a 3.2-3.4 percent increase in urban extent. This finding supports the notion that the influence of agricultural land on urban extent occurs at a larger spatial scale than that of the city and its environs. Replacing geographic variables with agricultural variables does not improve the model's explanatory power. In model 4.1 the lack of spatially referenced data for cultivated land resulted in seven fewer observations.

In model 5 we keep population and GDP per capita, remove geographic and agricultural variables, and include economic and policy variables. The airport score is a proxy for global economic connectedness. When the score increases we expect the economic profile of a city to increase, resulting in higher volumes of international trade and higher prices for domestically produced export goods. This should increase urban extent. The building regulations variable is a proxy for rules and regulations governing land and housing. Higher values indicate more onerous regulations and greater opposing pressure against urban extent. The gas price variable is a measure of transportation cost, and all things being equal we expect that places with higher transportation cost to have smaller urban extents. Only the regulatory variable was found to be significant, but the expected direction of the sign is reversed making none of the economic and policy variables meaningful in this particular model specification. A lack of building regulation data in 15 cities resulted in a fewer number of observations.

In Model 6, we test what happens when we remove the two variables with the strongest explanatory power. If population and GDP per capita alone explain 85 percent of the variation in urban extent, does that mean that other combinations of variables can explain at most 15 percent of observed variation? We include geographic, agricultural, economic and policy variables, and exclude population and GDP per capita. The model's explanatory power is 39 percent. The airport score is significant, and all other variables are not. A 100 percent increase in the airport score is associated with a 32 percent increase in urban extent. The airport score is moderately correlated with GDP per capita, and we suspect this explains its statistical significance in this model but not in model 5. The sign on the gas price variable is negative, as expected, but it is not a significant predictor.

The final model of urban extent, number 7 is a full test of all variables, including only agricultural land per capita as the sole agricultural variable. The model's explanatory power of 87.3 percent is only a 2.27 percent increase over a model that includes population and GDP per capita only. Population, GDP per capita, climate, agricultural land per capita, and building regulations are all statistically significant. The parameter estimates for population and GDP per capita are similar to previous specifications; a 10 percent increase in population is associated with a 7.8 percent increase in urban extent and a 10 percent increase in GDP per capita is associated with a 4.9 percent increase in urban extent. A 10 percent increase in agricultural land per capita is associated with a 3.1 percent increase in urban extent. As in model 5, the expected sign on building regulations is reversed. This may mean that the variable is not a good proxy for regulatory stringency or it may mean that stringent regulations on land and housing are not effective at limiting the outward expansion of urban areas. Indeed, that population and GDP alone explain 85 percent of the variation in urban extent strongly suggests that there is little to be explained by policy variables.

<u>Logarithmic Models of Urban Expansion</u>

As the dependent variable in models of urban expansion is the annualized urban extent growth rate between T2 and T3, we sought independent variables in terms of annualized growth rates over similar time periods to the extent this was possible. Apart from the population growth variable, which match the T2 and T3 dates exactly, growth rates for other independent variables, such as those at the country level, are based on observations at 2000 and 2014, the median dates of the T2 and T3 periods.

Model 1 looks at the population growth rate as the sole explanatory factor. It is statistically significant and a 10 percent increase in the population growth rate is associated with a 12 percent increase in urban extent growth rate. The coefficient on the population growth variable suggests that urban extents are increasing at faster rates than their populations are increasing and confirms the trend we observed in the findings section. Population growth alone explains 64 percent of observed variation in the urban extent growth rate.

Model 2 keeps the population growth rate and introduces economic variables, including the GDP per capita growth rate at the national level, the city GDP per capita at T3, and a dummy variable for if the city is in a developing country. We would have liked to have included the city GDP growth rate but we could not locate nor reasonably estimate year 2000 city GDP. Only the national GDP growth rate is significant and a 10 percent increase in the growth rate is associated with 1.1 percent increase in the urban extent growth rate. As GDP per capita within a country increases over time we would expect the average citizen to have higher income, which would lead to higher rates of expansion. The level of GDP per capita in a city is not a significant predictor of the expansion rate. Whether a city is in a developed country does not appear to be a significant predictor of the expansion rate once the population growth rate and the country GPD growth rate are held constant. Developed versus developing country status may not be a statistically significant predictor as it is correlated with the population growth rate. There is a practically negligible improvement in the explanatory power of this model over model 1.

Model 3 keeps the variables from model 1 and 2 and adds buildable land in the expansion area and agricultural land per capita at the country level. Neither buildable land nor agricultural land per capita refer to change over time. We would expect to see lower expansion rates in places with less buildable land. We saw that urban extents are larger, all things being equal, in countries where there is more agricultural land per capita. We might take this to mean that agricultural land use exerts less pressure against urban extent and expansion in places where there is plentiful agricultural land. In other words, we would expect higher urban extent growth rates in places with more agricultural land per capita. When we look at the parameter estimates, the population growth rate, the national GDP per capita growth rate remain significant. Agricultural land per capita is significant but its sign is in the opposite direction. The effect size is so small that we consider the finding insignificant for practical purposes, namely a 10 percent increase in the arable land per capita is associated with a 0.1% decrease in the expansion rate. Perhaps buildable land was found to be insignificant because the buildable land threshold and the buildable land analysis area are not capturing the true buildable land effect. There is a very small improvement in the explanatory power of the model, at 66.3 percent.

Model 4 is the full specification that includes all explanatory variables. To the variables in model 3 we add the growth rate in the average national per liter gas price, the growth rate in the motorization level, measured as vehicles per thousand, and the measure of regulatory stringency. The explanatory power of the model increases to 70 percent. The population growth rate remains significant but the GDP per capita growth rate is now insignificant. In its place, the motorization growth rate is significant, and we suspect this is because the two are moderately correlated. As a country becomes wealthier, we would expect the vehicle ownership rate to increase. The parameter estimate for the motorization is very similar to the GDP per capita growth rate in iterations 2 and 3; namely, a 10 percent increase in the motorization rate is associated with a 1.8 percent increase in expansion rate. City GDP per capita at T3 is now significant, and the negative sign indicates that places with higher average incomes are associated with smaller expansion rates. Being a city in a developing country is also a found to be a significant predictor of the expansion rate and the positive sign indicates they are associated with higher expansion rates. The signs for these last two variables are in the expected direction but their effect size is so small that the findings are insignificant for practical purposes.

Finally, in model 5, we test what happens if we remove information about change in population and change in GDP per capita. The model performs relatively poorly as expected, with an explanatory power of 32 percent. The motorization growth rate is the only significant variable and we suspect, as before, that it is acting as a proxy for change in income. Overall, the models of urban expansion show that changes in population and income are its key explanatory factors.

Table 1. Logarithmic Models of Urban Extent

Estimate (SE), p-value

Variables	Model 1	Model 2	Model 3	Model 4.1	Model 4.2	Model 4.3
Intercept	-1.79 (0.49), <0.001***	-5.89 (0.53) ,<.0001***	-5.96 (0.52) ,<.0001***	-5.79 (0.61), <.0001***	-5.67 (0.58), <.0001***	-5.89 (0.51), <.0001***
Log (Population)	0.85 (0.03) ,<.0001***	0.78 (0.03) ,<.0001***	0.80 (0.03) ,<.0001***	0.78 (0.03), <.0001***	0.78 (0.03), <.0001***	0.78 (0.03), <.0001***
Log (GDP per Capita)		0.54 (0.05) ,<.0001***	0.51 (0.05) ,<.0001***	0.49 (0.06), <.0001***	0.49 (0.06), <.0001***	0.51 (0.05), <.0001***
Log (Buildable Land)			0.27 (0.09) ,0.002*			
Log (Climate)			0.29 (0.08) ,<0.001***			
Log (Share Buildable Land Cultivated)				0.05 (0.23), 0.833		
Log (Agricultural Land per Cap)				0.32 (0.10), 0.001*	0.32 (0.09), 0.001*	0.34 (0.08), <.0001***
Log (Ag Val per ha ÷ GDP per cap)				-0.05 (0.07), 0.467	-0.06 (0.07), 0.438	
Log (Airport Score)						
Log (Building Regulations)						
Log (Gas Price)						
No.Observations	200	200	199	193	200	200
Adjusted R-square	0.7549	0.8503	0.8618	0.8606	0.861	0.8613

*<0.05 ***<0.001

Estimate (SE), p-value

Variables	Model 5	Model 6	Model 7
Intercept	-5.53 (0.80), <.0001***	11.53 (0.41), <.0001***	-6.16 (0.78), <.0001***
Log (Population)	0.75 (0.03), <.0001***		0.78 (0.03), <.0001***
Log (GDP per Capita)	0.52 (0.06), <.0001***		0.49 (0.06), <.0001***
Log (Buildable Land)		-0.08 (0.58), 0.887	0.66 (0.26), 0.014*
Log (Climate)		-0.12 (0.21), 0.559	0.17 (0.10), 0.089
Log (Share Buildable Land Cultivated)			
Log (Agricultural Land per Cap)		0.22 (0.18), 0.240	0.31 (0.08), <0.001***
Log (Ag Val per ha ÷ GDP per cap)			
Log (Airport Score)	0.02 (0.02), 0.370	0.32 (0.03), <.0001***	0.01 (0.02), 0.489
Log (Building Regulations)	0.15 (0.05), 0.003*	0.15 (0.10), 0.155	0.13 (0.05), 0.008*
Log (Gas Price)	0.02 (0.07), 0.744	-0.10 (0.16), 0.516	0.02 (0.07), 0.786
No.Observations	185	183	184
Adjusted R-square	0.8696	0.3884	0.8731

*<0.05 ***<0.001

Table 2. Logarithmic Models of Urban Expansion

Estimate (SE), p-value

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.008 (0.002), 0.002*	0.025 (0.022), 0.249	0.029 (0.024), 0.219	0.05 (0.024), 0.064	0.004 (0.010), 0.729
Log (Pop. Gr. Rt. T2-T3)	1.227 (0.066), <.0001***	1.166 (0.074), <.0001***	1.166 (0.073), <.0001***	1.094 (0.076), <.0001***	
Log (Country GDP/cap Gr. Rt. T2-T3)		0.112 (0.036), 0.002*	0.127 (0.036), 0.001*	-0.004 (0.052), 0.939	
Log (City GDP/cap T3)		-0.001 (0.001), 0.206	-0.001 (0.001), 0.217	-0.002 (0.001), 0.031**	
Developed		0.003 (0.004), 0.523	0.005 (0.004), 0.257	0.010 (0.005), 0.042*	
Log (Buildable Land)			-0.003 (0.010), 0.789	-0.002 (0.010), 0.817	0.020 (0.015), 0.174
Log (Agricultural Land per Cap)			-0.010 (0.003), 0.004*	-0.007 (0.004), 0.088	-0.007 (0.006), 0.220
Log (Gas Price Gr. Rt. T2-T3)				-0.006 (0.051), 0.911	0.059 (0.075), 0.433
Log (Motorization Gr.Rt. T2-T3)				0.179 (0.057), 0.002**	0.366 (0.054), <.0001***
Log (Building Regulations)				0.004 (0.002), 0.081	0.004 (0.003), 0.137
No.Observations	**200**	**196**	**195**	**175**	**175**
Adjusted R-square	**0.6364**	**0.6507**	**0.6627**	**0.7071**	**0.32**

*<0.05 ***<0.001

Conclusion

We now have up-to-date quantitative estimates for how rapidly cities have expanded outwards and how their populations changed over the 1990 – 2014 period. Our analysis yielded a number of important findings. We observed general trends that applied to all cities but we also noted that within these trends lay important differences that distinguished groups of cities from one another.

Both average urban extent and average population growth rates have declined over time, from the 1990 – 2000 to the 2000 – 2014 period in all cities, in less developed country cities, and in more developed country cities. Actual growth rates in less developed cities and more developed cities were vastly different, however.

The median and average urban extent growth rates in less developed cities were 5.7 percent per year and 6.2 percent per year for the 2000 – 2014 period compared to median and average rates of 1.1 percent per year and 1.8 percent per year in more developed country cities. If the median growth rate in less developed cities remains stable, we would expect half of the cities in less developed countries to at least double their areas in only 12 years. Making minimal preparations for land and infrastructure in advance of development in these cities would appear to be a most urgent task.

Cities' urban extents grew faster than their populations and this was true for both the 1990 – 2000 and 2000 – 2014 time periods. The median and average population growth rates in less developed cities were 3.6 percent per year and 3.8 percent per year for the 2000 – 2014 period compared to median and average rates of 0.7 percent per year and 0.7 percent per year in more developed cities. Multiples are a convenient way to communicate the total change implied by annualized growth rates. Over the 2000 -2014 period, the median less developed city increased its area by a multiple of 2.2 but only increased its population by a multiple of 1.6. Over the same period, the median more developed city increased its area by a multiple of 1.2 but only increased its population by a multiple of 1.1.

SDG indicator 11.3.1 promotes the ratio of the land consumption rate to the population growth rate as a measure of the relationship between the urban extent growth rate and the population growth rate. We used a different but related measure. We focused on the average difference between the urban extent growth rate and the population growth rate and we can now provide estimates that inform the monitoring of this indicator. For all cities over the 2000 – 2014 period, the average difference between the urban extent growth rate and the population growth rate was 2.1 percentage points. In less developed countries cities the average difference was higher, 2.4 percentage points, while in more developed cities it was lower, 1.1 percentage points. When we measure the sample again in the future, we can compare the average differences to determine whether they different from each other, and if so, in what direction.

While there are several potential factors that might explain how much area cities occupy and how the area they occupy changes over time, two factors dwarfed all others in their ability to explain the changes we observed in 200-city sample. Population and income alone explained 85 percent of the variation in urban extent and 65 percent of the variation in urban expansion. Simply put, population growth, and larger incomes that allow residents to consume more land, drive the differences we see. If income growth is an explicit development goal in less developed regions, then the combination of population growth and rising incomes less developed regions almost certainly implies a continuation of the outward expansion of cities.

When this exercise is repeated in the future to update and improve the existing knowledge base, perhaps in support of monitoring SDG indicator 11.3.1, there will be a possibility to incorporate changes that enhance our understanding of urban extent and population growth in the world, in different world regions, and in different countries.

The free, public availability of global time-series data for built-up area, such as the European Commission's Global Human Settlement Built-up Grid (Pesaresi et al. 2015) eliminates the time and effort that was previously needed to classify Landsat imagery. This makes larger sample sizes that would yield more precise estimates much more feasible. The global built-up data also makes it possible to reverse engineer the creation of a universe of cities. Instead of gathering data to build up a universe of cities, it should be possible to apply our urban extent script to disaggregate the entire dataset into its component urban extents. We are currently experimenting with this.

Completing the analysis also requires population data, and while there are new global population products that could be applied to such analyses (JRC CIESIN 2015; Worldpop 2018) our experiences have taught us that there is considerable variation in the reliability of small area estimates for the world at large. One of the benefits of working with a 200-city sample was that we could contact the relevant authorities, organizations, and experts in search of the most reliable local data, data that was not always included in national level datasets. This sort of detective works becomes more difficult when the sample size is larger, but improvements to population models and products, as well as the cooperation of national and local authorities for monitoring SDG indicator 11.3.1 will hopefully lead to reliable and readily available population data at geographical scales suited for city-based analyses.

References

Alonso, William. 1964. *Location and Land Use.* Cambridge, MA: Harvard University Press.

Angel, S., Sheppard, S.C., and Civco, D.L. 2005. *The dynamics of global urban expansion.* Washington D.C.: Transport and Urban Development Department, World Bank.

Angel S., Parent, J., Civco, D.L., Blei, A., and Potere, D. 2011. "The Dimensions of Global Urban Expansion: Estimates and Projections for All Countries: 2000 – 2050." *Progress in Planning, 75*(2): 53 – 107.

Angel, S., Parent, J., Civco, D. L., and Blei, A.M. 2012. *The Atlas of Urban Expansion.* Cambridge MA: Lincoln Institute of Land Policy.

Angel, S., Blei, A.M., Parent, J., Lamson-Hall, P., Galarza, N., Civco, D.L., Lei, Q., and Thom, K. 2016. *Atlas of Urban Expansion—2016 Edition*, Volume 1: Areas and Densities, New York: New York University, Nairobi: UN-Habitat, and Cambridge, MA: Lincoln Institute of Land Policy.

Barnett, C. and Parnell, S. 2016. "Ideas Implementation and Indicators: Epistemologies of the Post-2015 Urban Agenda." *Environment & Urbanization, 28(1): 87-98.*

Brueckner, J. 1987. "The Structure of Urban Equilibria: a Unified Treatment of the Muth-Mills Model." In E. Mills (Ed.), *Handbook of Regional and Urban Economics,* pp. 821 - 845. New York: Elsevier.

Burchfield, M., Overman, H.G., Puga, D. and Turner, M.A. 2006. "Causes of Sprawl: a Portrait from Space." *The Quarterly Journal of Economics*, *121*(2): 587 – 633.

Campante, F. and Yanagizawa-Drot, D. 2016. "Long-Range Growth: Economic Development in the Global Network of Air Links." NBER Working Paper No. 22653.

Chen, I., Franklin, J.F., and Spies, T.A. 1992. "Vegetation Responses to Edge Environments in Old Growth Douglas Fir Forests." *Ecological Applications, 2*(4): 387 – 396.

Cohen, B. 2004. "Urban Growth in Developing Countries: A Review of Current Trends and a Caution Regarding Existing Forecasts." *World Development, 32(1)*: 23 – 51.

Deuskar, C. and Stewart, B. 2016. "Measuring Global Urbanization Using a Standard Definition of Urban Areas: Analysis and Preliminary Results. Paper presented at World Bank Land and Poverty Conference 2016." https://www.conftool.com/landandpoverty2016/index.php/Deuskar-107-107_paper.pdf?page=downloadPaper&filename=Deuskar-107-107_paper.pdf&form_id=107&form_version=final

Donaldson D. and Storeygard, A. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives, 30*(4): 171 – 198.

European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network – CIESIN. 2015. GHS Population Grid, Derived From GPW4, Multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a

Fulton, L., Jenn, A., and Tal, G. 2017. "Can We Reach 100 Million Electric Carse Worldwide by 2030?" Working Paper 16. London: Global Fuel Economy Initiative.

Globalization and World Cities Research Network. "The World According to GaWC." www.lboro.ac.uk/gawc/gawcworlds.html

Mertes, C.M., Scheider, A., Sulla-Menashe, D., Tatem, A.J., and Tan, B. 2015. "Detecting Change in Urban Areas at Continental Scales with MODIS Data." *Remote Sensing of the Environment*, *158*: 331 – 347.

Mills, E.S. 1967. "An Aggregative Model Of Resource Allocation in a Metropolitan Area." *American Economic Review*, *57*: 197 – 210.

Muth, R. 1969. *Cities and Housing.* Chicago: University of Chicago Press.

OECD 2012. *Redefining "Urban": A New Way to Measure Metropolitan Areas,* Paris: OECD Publishing.

Parr, J.B. 2007. "Spatial Definitions Of The City: Four Perspectives." *Urban Studies, 44*(2): 381 – 392.

Parnell, S. 2016. "Defining a Global Urban Development Agenda." *World Development, 78: 529-540.*

Peel M.C., Finlayson B.L. and McMahon T.A. 2007. "Updated World Map Of The Köppen-Geiger Climate Classification." *Hydrology and Earth System Sciences*, *11*: 1633-1644.

Pesaresi, M., Ehrilch, D., Florczyk, A.J., Freire, S., Julea, A., Kemper, T., Soille, P., Syrris, V. 2015. GHS Built-Up Grid, Derived From Landsat, Multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b

Potere, D., Schneider, A., Angel, S., and Civco, D.L. 2009. "Mapping Urban Areas on a Global Scale: Which of the Eight Maps Now Available is More Accurate?" *International Journal of Remote Sensing, 30*(24): 6531 – 6558.

Qihao W., Esch, T., Gamba, P., Quattrochi, D., and Xian, G. 2014. "Global Urban Observation and Information: GEO's Effort to Address the Impacts of Human Settlements." In: Weng, Q. Ed. *Global Urban Monitoring and Assessment Through Earth Observation*, Boca Raton: CRC Press, 15-34.

Seto, K.C., Fragkias, M., Gunerlap, B., and Reilly, M. 2011. "A meta-analysis of global urban land expansion." *PlosOne*, *6*(8)*:* e23777.

Schneider, A., Mertes, C.M., Tatem, A.J., Tan, B., Sulla-Menashe, D., Graves, S.J., Patel, N.N., Horton, J.A., Rollo, J.T., Schelly, I.H., Stevens, F.R., and Dastur, A. 2015. "A New Urban Landscape in East-Southeast Asia." 2000-2010. *Environmental Research Letters, 10,* article 034002.

Schneider A., and Woodcock, C.E. 2008. "Compact, Dispersed, Fragmented, Extensive? A Comparison of Urban Expansion in Twenty-Five Global Cities Using Remotely Sensed Data, Pattern Metrics, And Census Information." *Urban Studies, 45:* 659-692.

Taubenbock, H., T. Esch, A. Felbier, M. Wiesner, A. Roth, and S. Dech. 2012. "Monitoring urbanization in mega cities from space." *Remote Sensing of the Environment*, *117:* 162-176.

Tauenbock, H., M. Wiesner, A. Felbier, M. Marconcini, T. ESch, and S. Dech. 2014. "New Dimensions of Urban Landscapes: The Spatio-Temporal Evolution From a Polynuclei Area to a Mega-Region Based on Remote Sensing Data." *Applied Geography, 47*: 137-153.

Uchida, H. and Nelson, A. 2016. "Agglomeration Index: Towards a New Measure of Urban Concentration." Background paper for *World Development Report 2009: Reshaping Economic Geography.* World Bank, Washington DC.

Wheaton, W.C. 1976. "On The Optimal Distribution of Income Among Cities." *Journal of Urban Economics, 3,* 31-44.

Winter, M., Johnson, D.H., and Faaborg, J. 2000. "Evidencef for Edge Effects on Multiple Levels in Tallgrass Prairie." *Condor*, *102*(2) 256-266.

Worldpop: www.worldpop.org

United Nations. 2014. *World Urbanization Prospects: The 2014 Revision.* Department of Economic and Social Affairs, Population Division. https://esa.un.org/unpd/wup/

Appendix A: Growth Rates

Descriptive statics for growth rates, all cities

	Urban Extent Growth Rates				Population Growth Rates				Difference, UE Gr. Rt.- Pop. Gr. Rt.		
Statistic	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3
Nbr. of observations	200	200	200	200	200	200	200	200	200	200	200
Minimum	0.005	0.001	0.006	-0.326	-0.023	-0.017	-0.013	-0.331	-0.188	-0.030	-0.073
Maximum	0.339	0.180	0.207	0.133	0.413	0.136	0.223	0.082	0.122	0.127	0.090
1st Quartile	0.025	0.014	0.023	-0.038	0.013	0.009	0.014	-0.017	-0.001	0.000	0.001
Median	0.042	0.031	0.042	-0.010	0.030	0.022	0.029	-0.005	0.014	0.012	0.017
3rd Quartile	0.081	0.079	0.082	0.018	0.059	0.045	0.050	0.005	0.042	0.035	0.033
Mean	0.063	0.050	0.056	-0.012	0.047	0.030	0.038	-0.017	0.016	0.021	0.018
Lower 95 CI	0.055	0.044	0.050	-0.021	0.039	0.026	0.033	-0.025	0.010	0.017	0.014
Upper 95 CI	0.061	0.071	0.057	-0.004	0.056	0.034	0.043	-0.010	0.021	0.025	0.022
Variance (n-1)	0.003	0.002	0.002	0.003	0.004	0.001	0.001	0.003	0.001	0.001	0.001
Standard deviation (n-1)	0.057	0.046	0.042	0.059	0.060	0.029	0.036	0.056	0.038	0.029	0.026

Descriptive statics for growth rates, less developed country cities

	Urban Extent Growth Rates				Population Growth Rates				Difference, UE Gr. Rt.- Pop. Gr. Rt.		
Statistic	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3
Nbr. of observations	148	148	148	148	148	148	148	148	148	148	148
Minimum	0.009	0.002	0.009	-0.326	-0.023	-0.017	-0.003	-0.331	-0.188	-0.030	-0.073
Maximum	0.339	0.180	0.207	0.133	0.413	0.136	0.223	0.082	0.122	0.127	0.090
1st Quartile	0.029	0.019	0.034	-0.041	0.020	0.016	0.021	-0.029	-0.008	0.004	0.000
Median	0.055	0.057	0.060	-0.010	0.040	0.036	0.039	-0.008	0.015	0.015	0.019
3rd Quartile	0.102	0.087	0.089	0.027	0.067	0.053	0.068	0.008	0.044	0.049	0.036
Mean	0.074	0.062	0.067	-0.012	0.058	0.038	0.047	-0.021	0.016	0.024	0.020
Lower 95 CI	0.061	0.051	0.061	-0.023	0.046	0.031	0.041	-0.031	0.009	0.019	0.015
Upper 95 CI	0.087	0.073	0.073	-0.001	0.071	0.044	0.053	-0.010	0.023	0.030	0.024
Variance (n-1)	0.004	0.002	0.002	0.004	0.004	0.001	0.001	0.004	0.002	0.001	0.001
Standard deviation (n-1)	0.061	0.047	0.043	0.066	0.064	0.029	0.037	0.064	0.043	0.032	0.029

Descriptive statics for growth rates, more developed country cities

	Urban Extent Growth Rates				Population Growth Rates				Difference, UE Gr. Rt.- Pop. Gr. Rt.		
Statistic	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3	Diff. T2T3 - T1T2	T1-T2	T2-T3	T1-T3
Nbr. of observations	52	52	52	52	52	52	52	52	52	52	52
Minimum	0.005	0.001	0.006	-0.099	-0.010	-0.016	-0.013	-0.086	-0.011	-0.011	-0.011
Maximum	0.118	0.062	0.080	0.053	0.097	0.048	0.067	0.025	0.069	0.048	0.034
1st Quartile	0.014	0.008	0.011	-0.027	0.000	0.001	0.001	-0.010	0.006	-0.001	0.004
Median	0.025	0.011	0.022	-0.010	0.013	0.007	0.013	-0.002	0.013	0.007	0.008
3rd Quartile	0.040	0.026	0.031	0.003	0.021	0.014	0.019	0.001	0.020	0.017	0.023
Mean	0.031	0.018	0.024	-0.013	0.016	0.007	0.011	-0.009	0.015	0.011	0.013
Lower 95 CI	0.026	0.015	0.020	-0.022	0.012	0.005	0.006	-0.015	0.011	0.007	0.010
Upper 95 CI	0.037	0.021	0.028	-0.005	0.020	0.010	0.015	-0.002	0.020	0.014	0.016
Variance (n-1)	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.000	0.000
Standard deviation (n-1)	0.026	0.016	0.015	0.029	0.026	0.015	0.017	0.022	0.016	0.014	0.012

Appendix B: Multiples

Descriptive statics for multiples, all cities

	Urban Extent Multiples			Population Multiples			UE Multiple - Pop. Mutiple		
Statistic	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14
Nbr. of observations	200	200	200	200	200	200	200	200	200
Minimum	0.93	0.94	1.15	0.80	0.79	0.74	-53.01	-1.04	-141.60
Maximum	29.75	12.38	97.35	62.48	6.68	187.34	20.14	10.27	80.20
1st Quartile	1.28	1.22	1.71	1.14	1.14	1.38	-0.04	0.00	0.13
Median	1.54	1.55	2.90	1.35	1.36	1.93	0.18	0.21	0.78
3rd Quartile	2.25	3.02	7.29	1.83	1.91	3.30	0.62	1.18	2.45
Mean	2.37	2.59	7.50	2.15	1.66	3.96	0.22	0.93	3.54
Lower 95 CI	2.02	2.27	5.46	1.69	1.54	2.74	-0.19	0.67	1.64
Upper 95 CI	2.73	2.91	9.53	2.62	1.78	5.18	0.64	1.19	5.44
Variance (n-1)	6.60	5.33	213.11	11.06	0.71	76.35	8.87	3.51	185.96
Standard deviation (n-1)	2.57	2.31	14.60	3.33	0.84	8.74	2.98	1.87	13.64

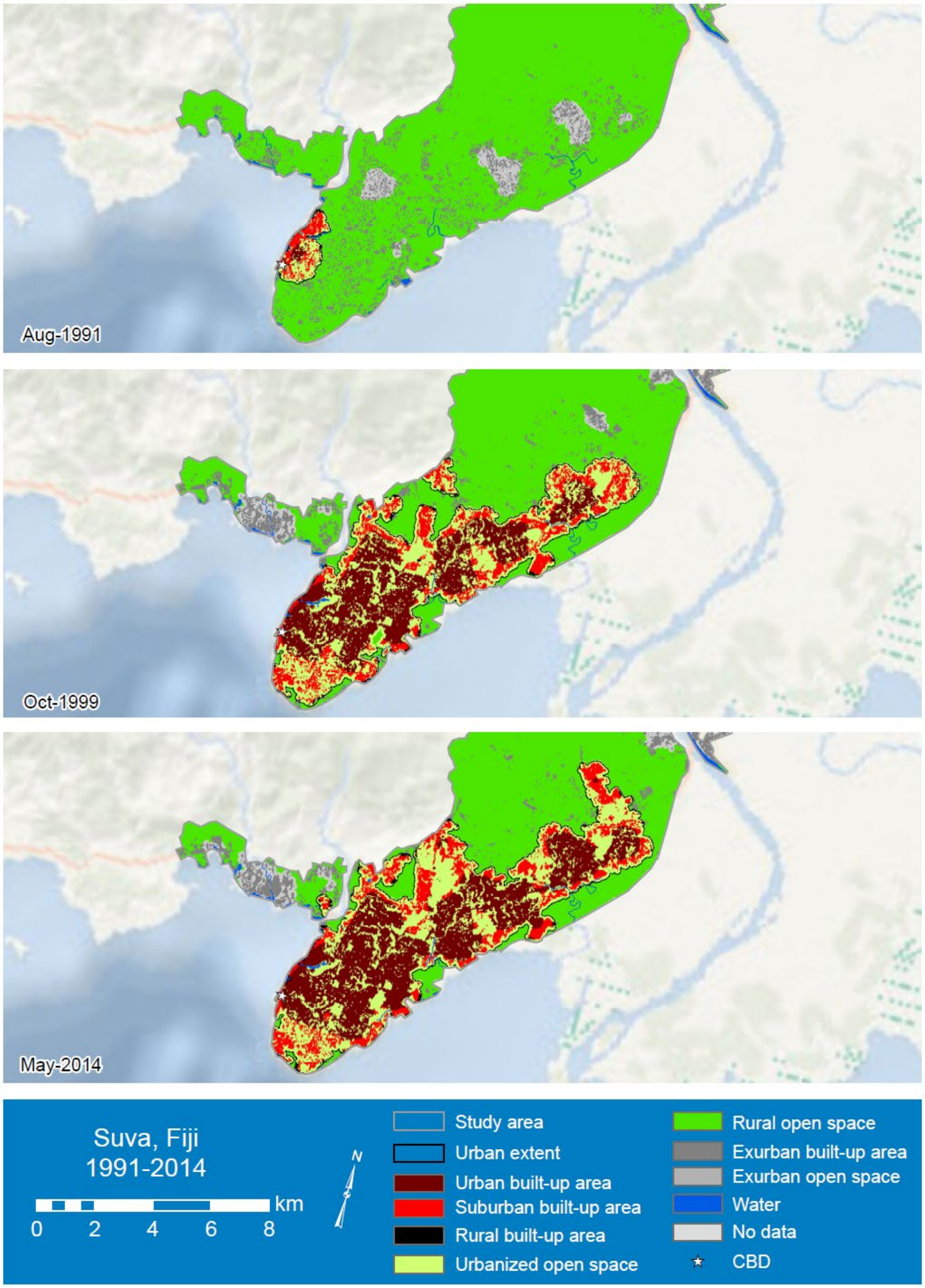
Descriptive statistics for multiples, less developed country cities

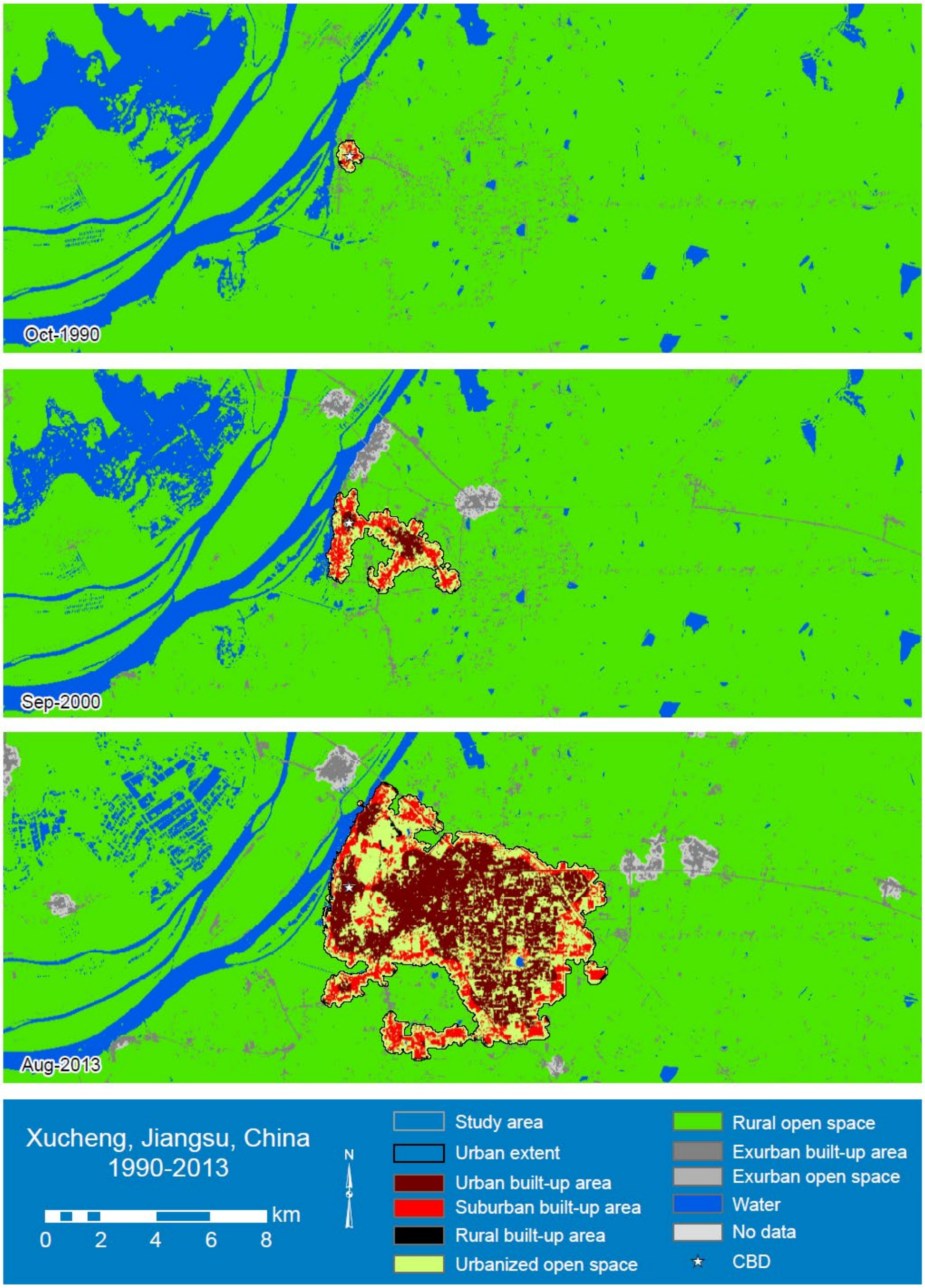
	Urban Extent Multiples			Population Multiples			UE Multiple - Pop. Mutiple		
Statistic	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14
Nbr. of observations	148	148	148	148	148	148	148	148	148
Minimum	0.93	0.94	1.24	0.80	0.79	0.90	-53.01	-1.04	-141.60
Maximum	29.75	12.38	97.35	62.48	6.68	187.34	20.14	10.27	80.20
1st Quartile	1.34	1.30	2.21	1.23	1.25	1.68	-0.11	0.09	0.05
Median	1.74	2.21	4.25	1.52	1.64	2.56	0.20	0.28	1.32
3rd Quartile	2.80	3.39	8.53	1.96	2.10	5.11	0.85	1.55	4.56
Mean	2.71	3.05	9.52	2.49	1.85	4.88	0.23	1.20	4.64
Lower 95 CI	2.24	2.64	6.83	1.87	1.71	3.25	-0.34	0.85	2.08
Upper 95 CI	3.19	3.46	12.21	3.11	2.00	6.51	0.79	1.54	7.20
Variance (n-1)	8.47	6.42	274.67	14.59	0.80	100.69	12.05	4.48	248.73
Standard deviation (n-1)	2.91	2.53	16.57	3.82	0.90	10.03	3.47	2.12	15.77

Descriptive statistics for multiples, more developed country cities

	Urban Extent Multiples			Population Multiples			UE Multiple - Pop. Mutiple		
Statistic	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14	'90 - '00	'00 - '14	'90 - '14
Nbr. of observations	52	52	52	52	52	52	52	52	52
Minimum	-0.16	-0.19	-0.37	0.88	0.80	0.74	-0.16	-0.19	-0.37
Maximum	1.67	1.18	2.44	2.70	1.96	4.87	1.67	1.18	2.44
1st Quartile	0.06	-0.01	0.13	1.00	0.90	1.04	0.06	-0.01	0.13
Median	0.15	0.14	0.29	1.14	1.10	1.35	0.15	0.14	0.29
3rd Quartile	0.22	0.25	0.59	1.25	1.22	1.58	0.22	0.25	0.59
Mean	0.21	0.19	0.48	1.22	1.13	1.42	0.21	0.19	0.48
Lower 95 CI	1.30	1.22	1.64	1.11	1.06	1.22	0.13	0.10	0.32
Upper 95 CI	1.56	1.42	2.15	1.33	1.21	1.61	0.29	0.27	0.64
Variance (n-1)	0.08	0.09	0.32	0.16	0.07	0.48	0.08	0.09	0.32
Standard deviation (n-1)	0.29	0.30	0.57	0.40	0.27	0.69	0.29	0.30	0.57

Appendix C: Extreme Urban Extent Multiple Examples





Oct-1990
Sep-2000
Aug-2013
Xucheng, Jiangsu, China
1990-2013
km
0 2 4 6 8
N
Study area
Urban extent
Urban built-up area
Suburban built-up area
Rural built-up area
Urbanized open space
Rural open space
Exurban built-up area
Exurban open space
Water
No data
CBD