LINCOLN INSTITUTE
OF LAND POLICY

The Shape Compactness of Urban Extents

Working Paper WP18SA1

Shlomo Angel
Marron Institute of Urban Management at New York University

Sara Arango Franco
Marron Institute of Urban Management at New York University

Yang Liu
Marron Institute of Urban Management at New York University

Alejandro M. Blei
Marron Institute of Urban Management at New York University

Xinyue Zhang
Marron Institute of Urban Management at New York University

September 2018

Abstract

The shape compactness of urban extents matters, just like urban density matters. Other things being equal, both metrics determine the average travel distances in cities, and hence their energy consumption and their greenhouse gas emissions. They also affect the length of infrastructure lines and the length of commutes, and hence also labor market integration and overall productivity. In principle, therefore, increasing either the shape compactness or the density of cities can contribute—in different yet equal measure—to mitigating climate change. There are strong forces that push urban extents to become more compact, circular or near circular in shape, and these forces have evolved over time. There are also key forces that have pushed urban extents to become less compact over time. We introduce these key forces and illustrate their effects on particular cities. We then define a set of metrics for measuring the shape compactness of cities. We use them to measure urban extents obtained from satellite imagery in a stratified global sample of 200 cities in three time periods: 1990, 2000, and 2014. We find that the shape compactness of cities the world over is independent of city size, area, density, and income and that, not surprisingly, it is strongly affected by topography. We also find that it has declined overall between 1990 and 2014 and explain some of the sources of this decline. We conclude the paper by assessing the ways in which the shape compactness of cities can be increased to make them more productive, more inclusive, more sustainable, and more climate-resilient in decades to come.

About the Authors

Shlomo 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.

The Marron Institute of Urban Management
60 5th Avenue, 2nd Floor
New York, NY 10011
Telephone: 212-992-6872
sangel@stern.nyu.edu

Sara Arango Franco is a data scientist at the Marron Institute of Urban Management at New York University. She can be reached at saf537@nyu.edu.

Yang Liu is a research scholar and statistician at the Marron Institute of Urban Management at New York University. He can be reached at yl3371@nyu.edu.

Alejandro M. Blei is a research scholar at the Marron Institute of Urban Management at New York University. He can be reached at ablei@stern.nyu.edu.

Xinyue Zhang is a Masters student at the NYU Center for Data Science and a graduate assistant at the Marron Institute of Urban Management at New York University. She can be reached at xz2139@nyu.edu.

Contents

1. The Conceptual Framework

This essay focuses on the shape compactness of urban extents. We can identify the physical footprint of a city—its urban extent—by examining remote sensing images of that city with a view to determining its outer edges, essentially what the ancient Romans referred to as its *extrema tecturum*, the outer limits of its built-up area. If we can determine its outer edges, then any given city can be said to have a two-dimensional shape—a geographical signature, so to speak—at any point in time. We can then focus on the geometric properties of that shape. Chief among those properties is its compactness, the degree to which it resembles a circle, arguably one of the most, if not the most, important characteristic of geographic shapes. It should not come as a surprise that the study of shape compactness in geography is almost two centuries old (Ritter 1822).

In this first section of this essay, we present the theoretical framework for studying the shape compactness of urban extents. In the second section, the methodology section, we focus on the way we obtained the global sample of 200 cities, the way we obtained their urban extents in three time periods—1990, 2000, and 2014—and the way we defined and measured the compactness properties of their urban extents. In a third section, we focus on findings associated with the key forces that act on urban extents to make them more or less compact, using specific examples of cities that illustrate the action of each one of those forces. I.n the fourth section, we present a set of statistical results that seek to answer three questions: (1) How do we account for and explain the variation in shape compactness among cities? (2) Have cities become significantly more or less compact in recent years? And (3) How do compactness and density affect the average distance traveled in cities, once we account for differences in their populations? In the fifth and concluding section, we discuss the policy implications of the foregoing analysis, suggesting that there are clear benefits for making cities more compact and exploring a set of pragmatic tools than can increase the compactness of cities over time.

The conceptual framework for the study of shape compactness in geography was articulated by Angel, Parent, and Civco in an article titled "Ten Compactness Properties of Circles: Measuring Shape in Geography" (Angel Parent and Civco 2010). The key insight in that article is that the circle—which, everyone agrees, is the most compact of two-dimensional shapes—has at least ten, if not more, different compactness properties, and that when studying the compactness of particular geographic shapes it is important to choose the appropriate properties for studying and measuring their compactness. This insight can be best illustrated by example.

One of the properties of the compactness of circles is their *perimeter compactness*: Among all shapes of a given area, the circle is the shape that has the minimum perimeter. Therefore, in designing a small, fortified city, for example, given that its walls are its most expensive public works, it is useful to make its shape approximate a circle. For then, the wall encompassing the city will be of minimum total length and thus of minimum cost. More generally, when we wish to study the shape compactness of fortified cities, perimeter compactness is the appropriate metric for comparing these cities to each other.

A second and entirely different property of the compactness of circles is their *proximity compactness*: Among all shapes of a given area, the circle is the shape with the minimum distance from all its points to a given point. In designing circus tents, for example, given that every spectator wants to be as close to the action as possible, it is useful to make their shape approximate a circle. For then, the distance of all spectators to the center of the tent will be at a minimum. Similarly, the extents of large nineteenth century cities, where most jobs were concentrated in their Central Business District (CBD), tended to be circular or near circular in shape, as workers sought to locate their homes as close to the CBD as possible. More generally, when we wish to study the shape compactness of the urban extents of largely monocentric cities, proximity compactness is the appropriate metric for comparing these cities to each other.

A third and also entirely different property of the compactness of circles is their *depth compactness*: Among all shapes of a given area, the circle is the shape with the maximum average distance from its points to its outer periphery. A circular floor plan of a large resort hotel in Las Vegas, for example, will admit the minimum natural light and air into guestrooms because most rooms will be away from its outer walls. In comparison, a floor plan with narrow wings, with rooms opening off double-loaded corridors, will admit natural light into every room. More generally, when we wish to study the shape compactness of floor plans of large hotels, depth compactness is the appropriate metric for comparing these floor plans to each other. Similarly, when we wish to study access from cities to their rural periphery, for example, we can focus on their depth compactness. We can then observe that depth compactness is much higher in near-circular cities surrounded by green belts, like Seoul, for example, than in tentacle-like cities penetrated by green fingers, like Copenhagen. In cities like Copenhagen, green fingers are, on the whole, closer to urban residents than in cities like London with greenbelts.

A fourth compactness property of shapes is their *Exchange Compactness*. It compares a given shape to an equal-area circle centered at its centroid, seeking to determine what share of its area is within that circle. If the shape were circular too, all its area would be within its equal area circle and it would have perfect Exchange Compactness. Alternatively, if a large share of the shape were outside that circle, it would have low Exchange Compactness. We could say that a share of the area within the circle was exchanged with an equal area outside that circle to create that shape. The essence of political gerrymandering, for example, is the creation of electoral district boundaries that jettison nearby voters—those living near the center of a given district—in exchange for voters living further away, in order to affect election outcomes for political gain. Exchange Compactness is thus a natural metric for detecting aggressive gerrymandering.

The key insight in Angel *et al*'s article is that the choice of compactness metrics should be appropriate to the shape being studied, the forces acting on that shape, and—when the shape is a human creation—the function that the shape seeks to fulfill. It should be rather obvious that proximity, depth or exchange compactness are of little importance in designing fortified cities; that perimeter, depth or exchange compactness are of little importance in designing circus tents or monocentric cities; that perimeter, proximity or exchange compactness are of little importance in designing floor plans for large resort hotels; and that perimeter, proximity, or depth compactness are not helpful in detecting the gerrymandering of election districts. The compactness of each one of these shapes has one or more appropriate measure of compactness associated with it. Associating an inappropriate measure of compactness for studying it, let alone for acting on it, is typically erroneous and misleading.

This insight raises two interesting research questions regarding the shape compactness of the spatial extents of contemporary cities and metropolitan areas. First, what are the forces, tendencies or intentions acting to make these cities more or less compact? And second, given these forces, tendencies, or intentions, what are the appropriate metrics for measuring the shape compactness of the spatial extents of cities and metropolitan areas?

It also raises a third question, a policy-related one: Is compactness a desirable attribute of the spatial extent of cities and, if so, how can we make cities more compact? This third question, raised in a number of studies in recent years (e.g. Harari 2016 and Saiz 2010), should now come to the fore given the recent debates on urban form, debates that have focused renewed attention on urban population density. The recent literature on climate change advances the proposition that increased density can help mitigate climate change, at least in part because higher-density cities reduce overall travel distances and hence reduce energy consumption and greenhouse gas emissions. If this is indeed the case—and the available evidence, though anecdotal, suggests that it is—then it can be argued that more compact cities have the same effect on greenhouse gas emissions as denser cities: Other things being equal, the more compact the city, the shorter the travel distances within it, and thus the lower overall energy consumption, and the lower greenhouse gas emissions. Indeed, one of the findings of this essay is that higher shape compactness has a similar effect on shortening travel distances in cities to that of higher density. In other words, making cities more compact can have similar impacts on climate change to making cities denser. This realization lends a new urgency to the study of the shape compactness of urban extents, beyond that advanced earlier by Harari (2016), for example, who showed that shape compactness has a measurable effect on the productivity of Indian cities, or by Saiz (2010), to take another example, who showed that the distortions of shape compactness brought about by topographical barriers have a measurable effect on the elasticity of housing supply in American cities.

In two earlier articles (Angel and Blei 2016a and 2016b), Angel and Blei argued that the productivity of cities is a function of their agglomeration economies, and that those agglomeration economies are metropolitan in scale. Chief among them are the economies yielded by creating integrated metropolitan labor markets, where firms have access to all workers and workers have access to all jobs, allowing firms to pick the best workers and workers to pick the best jobs, thus increased overall productivity. Indeed, it can be demonstrated that the productivity per worker increases as the size of the metropolitan labor market increases. In American cities, the focus of their study, jobs are now highly decentralized, with only a small share of jobs, less than 15 percent on average, still located in the Central Business Districts (CBDs) of metropolitan areas. This suggests that the productivity of metropolitan areas gains from improved overall access to jobs, in turn suggesting that reducing the distances between all locations increases productivity, over and above reducing greenhouse gas emissions. Still, in metropolitan areas in other countries, especially in less-developed countries, job decentralization may not be as pronounced and a larger share of jobs are still concentrated in CBDs, suggesting that reducing the distances between all locations to the CBD will increase productivity, over and above reducing greenhouse gas emissions.

This suggests that there are at least three appropriate compactness properties of geographic shapes that should be employed in studying the compactness of contemporary urban extents, *Proximity Compactness* and *Exchange Compactness* introduced earlier, as well as *Cohesion*

Compactness. *Proximity compactness* compares the average distance from all points in the city to its CBD to the average distance from all points in a circle of an equal area to its center, noting that among all shapes of a given area, the circle is the shape with the minimum distance from all its points to a given point. Proximity Compactness is thus an appropriate measure of overall access to the CBD.

Exchange Compactness measures the share of the equal area circle centered at the centroid of a given urban extent that could have been built upon but was left open. That share was exchanged for an equivalent share of the urban extent built outside the equal area circle, thus making the city less compact than it could have been. Exchange Compactness introduces the possibility of isolating the effect of topography on the compactness of cities. We can distinguish between buildable land—dry land with a slope lower than, say, 15 percent (8.53°)—and steep slopes and water bodies that cannot be built upon. We can then construct the Buildable Land Circle about the centroid of a given urban extent, a circle that contains the same amount of *buildable* land as the urban extent. *Buildable Land Compactness* is then the Exchange Compactness of the urban extent with the Buildable Land Circle. We may find, as we shall see below, that there are cities that have a low value of Exchange Compactness but a rather high value of Buildable Land Compactness. In other words, the urban extent of these cities is not compact in some absolute sense, but it may be as compact as can be given their topographic limitations.

Cohesion compactness, the third type of compactness that is appropriate for studying urban extents, compares the average distance between two random points in the physical extent of a city to the average distance between two random points in a circle of an equal area, noting that among all shapes of a given area, the circle is the shape with the minimum distance between a random pair of points within it. Cohesion compactness is thus a measure of the overall accessibility of all locations to all other locations in the city, a measure particularly suitable to metropolitan areas with highly decentralized job locations.

In the present study, we do not measure the *Depth Compactness*, although that property of urban shapes may be of some interest. As noted earlier, cities that have open space fingers penetrating them increase overall access to open space. Indeed, they may do so at the expense of making the city less compact. This effect will be more pronounced if the fingers are very wide, but rather small if the fingers are thin compared to the overall shape of the city. A more detailed discussion of Depth Compactness is left for further study.

Metrics associated with the key compactness properties of cities allow us to calculate the shape compactness of the urban extents of the 200 cities in our global sample in three time periods, to be described in greater detail in the following section. It also allows us to begin to answer the three key questions about the shape compactness of cities listed earlier, questions that could not be answered before, using statistical techniques. The dataset created by determining the urban extents of the 200 cities in the global sample, calculating compactness indices for each one of them for three time periods—1990, 2000 and 2014—and testing the resulting values using established statistical methods allows us, for the first time, to provide provisional answers to these important questions.

People flock to cities to be closer to each other. Indeed, we can characterize ‘the urbanization project’—the great migration of people into cities that has started in earnest at the end of the eighteenth century and is likely to come to a halt by the end of the twenty-first century—as the movement of people from being closer to the land to being closer to each other. Cities should be naturally compact because greater compactness—both that brought about by greater urban density and that brought about by rounder urban extents—increases the access of people to each other, facilitating all manners of contact and exchange between them. If access and connectivity would be the only forces acting on urban extents then we should expect urban extents to be very compact. In reality, however, the extents of most cities are by no means perfect circles and do not even resemble circles. The question is why. In this essay, we focus on the interplay between the forces, tendencies, and intentions that render cities more compact or less compact and seek to gain a greater understanding of their effects.

2. Methodology

In this second section, we describe the methodology for selecting the global sample of cities; identifying the urban extents of the cities in the sample in three time periods; and measuring the compactness properties of these urban extents in these three time periods.

2.1. The Global Sample of Cities

The analysis of shape compactness focuses on a 200-city sample featured in the *Atlas of Urban Expansion – 2016 Edition* (Angel *et al* 2016). These 200 cities represent a 4.7 percent sample drawn from a universe of 4,231 cities identified by the research team. The sample was carefully selected to be representative of the distribution of the universe of cities by world region, by city population size, and by the number of cities in a given country.

The 4,231 cities in the universe of cities are all contiguous or near-contiguous built-up areas of settlements that had populations of 100,000 or more in the year 2010. By the geographical extent of the built-up area we refer to the relatively contiguous built-up area extending out of a historical city center that is visible to the naked eye from high resolution satellite imagery, such as that which can be viewed on Google Earth or Bing Maps. A contiguous built-up area may include several municipalities and is neither constrained nor defined by administrative boundaries. A single observation in the universe of cities may therefore represent a number of adjacent municipalities.

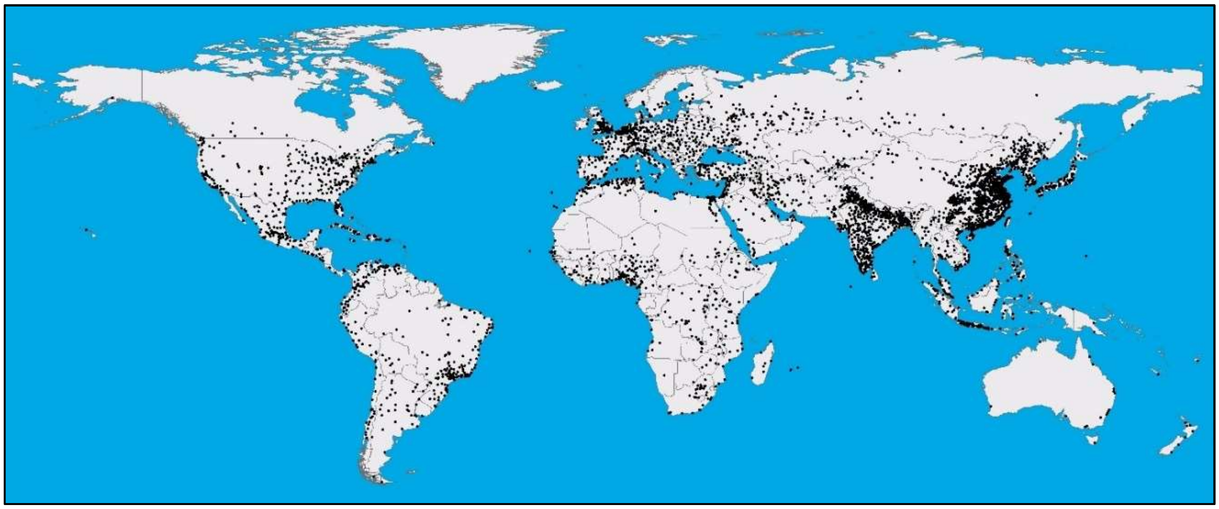
To construct the universe of cities it was necessary to first identify candidate cities from lists of cities and towns, municipalities, metropolitan areas, and urban agglomerations with a reliable population estimate for 2010 or for which a population value at 2010 could be estimated. The three main data sources for this exercise were the UN Population Division, which provided data for settlements with populations of at least 300,000, the website www.citypopulation.de, which reproduces census data and census maps for all countries, and the Chinese Academy of Sciences which provided information for Chinese settlements.

Google Earth satellite imagery was used to inspect each candidate city, both to confirm its existence and to determine whether it should be merged with neighboring observations as part of a larger urban extent. Candidate cities below the population threshold that were not part of a larger extent were excluded from the analysis. In a small number of cases, those associated with cities that are part of larger metropolitan conurbations—such as the Northeast Corridor in the United States—the locally-defined administrative boundary of the city was used to differentiate one built-up extent from another, resulting in the separation of the New York and Philadelphia built-up areas, for example. Similar divisions were applied in China's Pearl River Delta region and in the Tokaido corridor in central Japan, as well as in a few other large conurbations where it was difficult to discern the boundaries of individual cities. In applying administrative boundaries as edges of cities—rather than applying the *Extrema Tectorum*, the edge of their built-up area as their boundary—we acknowledge that a city's extent cannot extend endlessly; it should roughly correspond to a commuting area or labor market area; in other words, the area linked together by social and economic spatial interaction.

It should be noted in passing here that in these cases, admittedly only a few, the calculation of compactness metrics for individual metropolitan areas could be misleading. The compactness of geographic shapes can only be calculated for contiguous or near-contiguous shapes that are complete, namely surrounded by an area that does not belong to the shape. Limiting a shape by one or more arbitrary lines—and administrative boundaries are indeed arbitrary lines—will typically render it more compact than it would be when considered a part of a larger chain of settlements.

The construction of the universe of cities lasted approximately one year during 2014-2015. While great efforts were taken to ensure an exhaustive review of available data, errors of omission or commission are possible, especially in countries with poor data programs, where information on settlement locations and their populations is unreliable. The locations of the 4,231 cities are shown in **figure 1** below. Details regarding the statistical properties of the universe of cities can be found in a companion working paper (Galarza *et al* 2018).

Figure 1: The universe of all 4,231 cities that had 100,000 or more in 2010.



The universe of cities was organized along three strata with a view to selecting a representative sample. The first stratum organized cities by eight world regions: (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.* Cities were sampled from the eight regions in proportion to the population of the universe of cities in these regions.

The second stratum organized cities by city population size, of which there were four categories, 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 from each of the four population size categories.

A third stratum was included in the sampling framework 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.

When combined, the eight regions, four population size categories, and three 'number of cities in the country' create 96 subcategories (8 × 4 × 3 = 96), or boxes, to which an observation in the universe of cities must belong. After distributing all 4,231 observations, 71 non-empty boxes remained. Sample cities were randomly drawn from these non-empty boxes in accordance with the sampling strategy. Although the sample is representative by design, we can adjust a city's representativeness by using information associated with that city's sampling box. Since each sampling box contains a unique number of cities and a unique population total, the findings for a particular city may be weighted to reflect the number of cities that city represents, using a city-based weight, or the total number of people that city represents, using a population-based weight. Which weight to use, or whether to apply weights at all, is a discretionary judgment that largely depends on the metric in question and on the question being asked. When it comes to compactness metrics, for example, the appropriate weights are city-based weights, particularly as we find that the shape compactness of cities is independent of their size. To obtain results for the universe of cities—say, to determine whether compactness has been increasing or decreasing over time, or whether compactness is affected by city income levels—each city in the sample is weighted by the number of cities it represents and a weighted average is obtained for the universe as a whole.

The locations of the 200 cities in the global sample are shown in **figure 2** below.

Figure 2: The global sample of 200 cities

2.2. Identifying the Urban Extent

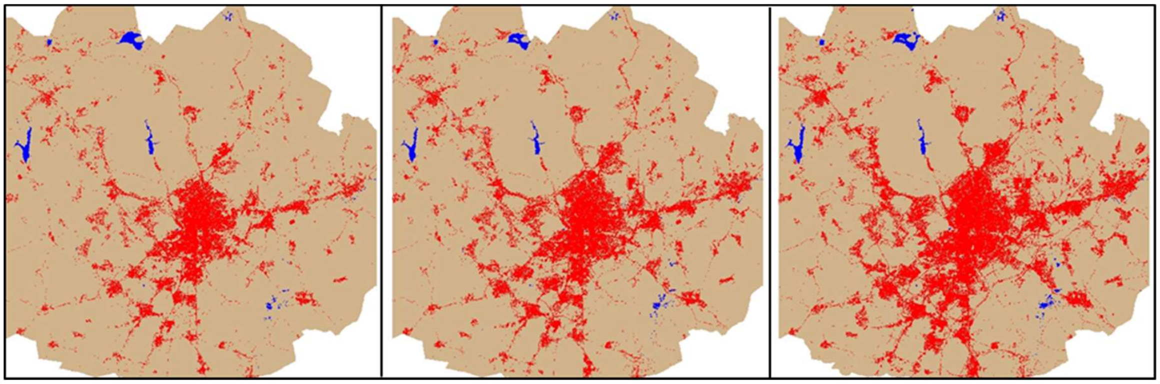
Each of the 200 sample cities was the focus of a detailed spatial analysis to determine its urban extent, or the combined built-up area and open space we associate with the city at a given time period. The urban extent was derived using a consistent methodology developed for the *Atlas of Urban Expansion*. It defines the boundary of the city used for the calculation of various spatial metrics, including its shape compactness.

The first step in the urban extent processing chain was to identify a city's study area. This is the area over which *Landsat* satellite imagery and spatially explicit population data, the two fundamental inputs required to complete all analysis for cities in the *Atlas of Expansion–Vol. 1: Areas and Densities*, would be collected. The study area needed to be large enough to completely contain the relatively contiguous built up area surrounding the city. Global nightlights data was initially used to identify this built-up area, as it is known to overestimate built-up area extent. Inspection of global nightlight data and the verification of these areas on *Google Earth* helped determine an initial study area. The research team then created revised study areas by identifying the set of spatially explicit enumeration districts—districts for which population data were available—that completely contained the initial study area. When enumeration districts completely contain the expected built up extent, or the initial study area, we can ensure that the total population of a zone will be apportioned to all the built-up area within it, and we can improve the estimate of the population associated with a given urban extent. To calculate that population, which may extend across several enumeration districts, we sum the district populations apportioned to its built-up areas.

Using the revised study area boundaries, we downloaded *Landsat* imagery from the United State's Geological Survey's *Earth Explorer* website. Images with cloud free areas of interest were downloaded for dates circa 1990, 2000, and 2014. A typical *Landsat* scene measures 185-by-185 kilometers and its basic building block is a 30-meter-square pixel. We superimposed the revised study areas on the *Landsat* scenes, extracted the intersecting areas, with an additional 1-kilometer buffer, and conducted a land cover classification over this area.

Our objective was to extract three land cover categories from each image corresponding to (1) water, (2) built-up, and (3) other/open space (not water). All *Landsat* pixels in the analysis area were assigned to one of these three classes by way of unsupervised classification techniques. The three-way classification of the Madrid study area in 1991, 2002, and 2010 is shown in **figure 3** below.

Figure 3. The three-way classification of Madrid into water (blue), built-up (red), and open space (brown), at the T1 (1991), T2 (2002) and T3 (2010).

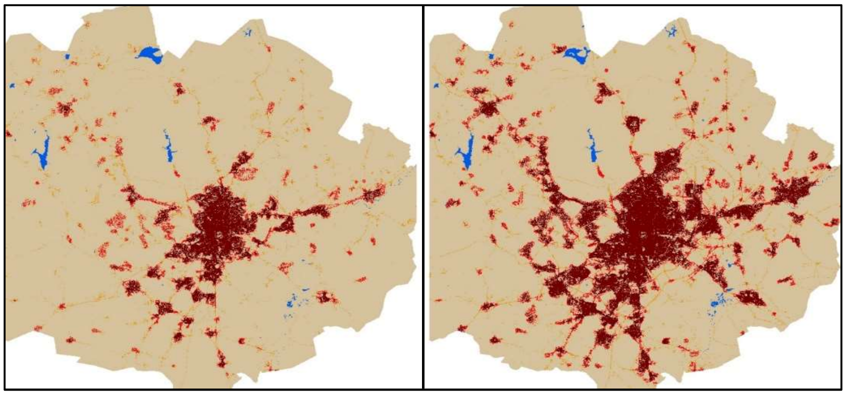


The three-way classification into water, built-up, and open space was the input into a secondary analysis. This secondary analysis, or landscape analysis, sub-classified built-up and open space pixels into three categories each, allowing us to differentiate among different types of built-up and open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

1. *Urban* pixels, where the majority (> 50 percent) of pixels within the Walking Distance Circle are built up;
2. *Suburban* pixels, where 25-50 percent of pixels within the Walking Distance Circle are built-up; and
3. *Rural* pixels, where < 25 percent 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 to identify areas that 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, of course, change the proportion of built up pixels in each category. We settled on these particular cutoffs after experimenting with different combinations of values in various cities, examining the output, and determining which combination of values was associated with the most consistent and intuitive results. The sub-classification of the built-up area of cities into urban, suburban, and rural pixels is demonstrated in **figure 4** below.

Figure 4. The sub-classification of built- up area into urban pixels (dark red), suburban pixels (red), and rural pixels (ochre) in Madrid, Spain in May, 1991 (left) and May, 2010 (right).

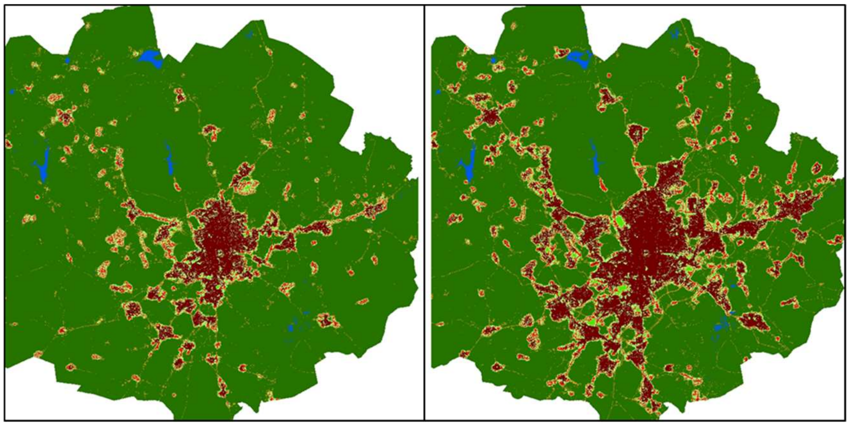


The three categories of open-space produced by the landscape analysis include:

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 that are less than 200 hectares in area; and
3. *Rural* open space pixels, all open space pixels that were neither fringe nor captured.

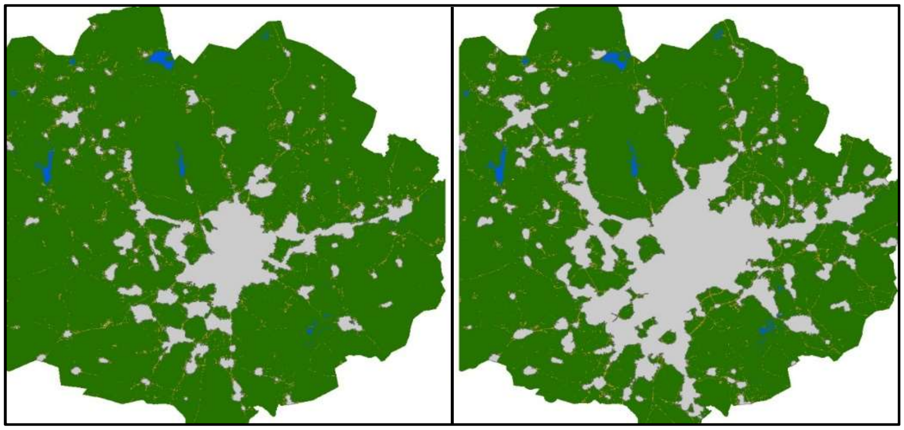
Taken together, the fringe and captured open space within a study area constitute *Urbanized open space*. Urbanized open space and rural open space together make up all of the open space within the study area. This sub-classification is demonstrated in **figure 5** below.

Figure 5. The sub-classification of open space into fringe open space (light green), captured open space (bright green), and rural open space (dark green) in Madrid, Spain in May, 1991 (left) and May, 2010 (right).



The differentiation of the study area into the three classes of built up, three classes of open space, and water facilitates the creation of rules that can be used to identify *urban clusters* across the study area. 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 Madrid, **figure 5** suggests that were dozens of urban clusters in 1991 and 2010. We can see the clusters more clearly in **figure 6** below. As a rule, the cluster containing the city hall location, which is usually indicative of a traditional city center and Central Business District (CBD), 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.

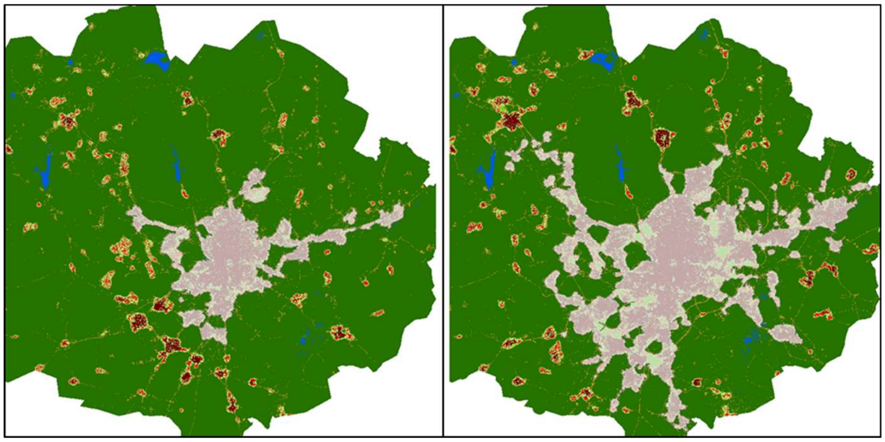
Figure 6: Urban clusters (grey) in Madrid in 1991 (left) and in 2010 (right)



We employed a rule 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 this rule 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 area of the buffer, an area equal to one-quarter the area of the cluster. The inclusion rule unites all clusters whose buffers intersect one another. The new grouping of clusters with overlapping buffers for a given city forms that city's *urban extent*. The urban extent for the city in question is the grouping of clusters that contains the city hall location. **figure 7** shows final urban extent of Madrid in 1991 and 2010.

Figure 7. The Urban Extent of Madrid (grey) in May 1991 (left) and May 2010 (right).



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 to accepted notions of what constituted the spatial extent of a city. In a sense, the task was a form of pattern recognition. The pattern is sometimes easy to discern, say that of a single large cluster completely surrounded by open countryside. Or it may be more difficult to discern, in the case of clusters of varying sizes in different proximities to each other, similar to, but typically more complex than the Madrid example. We apply a single rule to all situations and while it performs quite well, it is not perfect.

It is also important to note here that the choice of a universal inclusion rule, any rule, to determine the urban extent of all cities has an impact on the measurement of their shape compactness. The inclusion rule *a priori* second-guesses what urban clusters belong to the urban extent and what urban clusters do not without resort to local knowledge. If it is too strict, it leaves many urban clusters that are not part of the main cluster around the CBD outside the urban extent. It is it too lenient, it includes many freestanding urban clusters that are quite far away from the CBD. In a small handful of cases we applied local knowledge to make 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. In others, we may have missed outlying residential complexes that, while quite far, are clearly considered part of the city in question.

2.3. Measuring the Compactness of Urban Extents

Given the compactness properties that are appropriate to the study of contemporary urban extents described in Section I, we can define precise indices that can be used to measure these properties. Following Angel, Parent and Civco (2010, 444), we construct the indices to adhere to five rules:

- The index must correspond to **a recognizable property** of the shape that is associated with a recognizable function or set of forces.

- There must be **real-world examples** that illustrate this property—as well as its associated function or set of forces—at both the low end and the high end of the index.

- The index must apply to **all two-dimensional geometric shapes**, including those made up of several non-contiguous patches.

- The index must be **dimensionless** (independent of the size of the shape) as well as *directionless* (independent of its orientation).

- The index must **vary between 0 and 1**, with the value of 1 assigned to the circle as the shape with maximum compactness.

The following intermediate metrics are used to construct, measure, and analyze the four compactness indices used in this essay:

- The **Equal Area Circle** of a city is a circle with an area equal to the urban extent of the city.

- **Buildable Land** is dry land with a slope of less than 15 percent (8.53°).

- **The Buildable Land Circle** of a city is a circle that contains buildable land equal in area to the urban extent of the city.

- The **Buildable Land Ratio** is the area of Equal Area Circle divided by the area of the Buildable Land Circle.

Buildable land was calculated from NASA's Shuttle Radar Topography Mission (SRTM) dataset, which contains a digital elevation model (DEM) and a water file. SRTM data is 30m-resolution and contains elevation data for the entire planet based on information collected in the year 2000. The buildable land threshold was chosen after conversations with builders and real estate professionals that suggested that slope values greater than 15 percent are associated with increased land development costs. It is clearly possible to build on steeper slopes, but building on steeper slopes raises land development costs—e.g. in excavation, in retaining walls, in road building, in water supply, in sewerage, and in drainage—often requiring complex engineering solutions. In the absence of proper structural engineering and adequate investment in land development, buildings on slopes are at risk of damage from landslides. We can say with 95 percent confidence that 5.9±1.4 percent of the urban extent of cities in 2014 were in areas with slopes greater than 15 percent.

The four compactness indices used in this study are defined below.

- The **Proximity Index** of a city is the ratio of the average distance from all points in the Equal Area Circle to its center and the average distance from all points of the city's urban extent to its Central Business District (CBD) identified by its City Hall.

- The **Cohesion Index** of a city is the ratio of the average distance from all points to all other points in the Equal Area Circle and the average distance from all points to all other points in the city's urban extent.

- The **Exchange Index** of a city is the share of its urban extent within an Equal Area Circle located at the centroid of its urban extent.

- The **Buildable Land Index** of a city is the share of its urban extent within the Buildable Land Circle centered at the centroid of its urban extent.

Finally, we introduce a measure of how much more compact urban extents are when we take physical barriers to urban expansion into account.

- The **Compactness Correction Factor** is the percentage increase in exchange compactness once the Equal Area Circle is replaced by the Buildable Land Circle.

Given these definitions, we can obtain values for all compactness indices for all cities in the global sample. The descriptive statistics for the four key compactness indices are given in **table 1** below.

Table 1: Descriptive statistics for the four compactness indices circa year 2014 for the universe of cities, weighted by city weights.

Variable	No. of Cities	Weighted Mean*	95 percent Confidence Intervals	Standard Deviation	Minimum	Maximum
Proximity Compactness	200	0.766	[0.749, 0.783]	0.119	0.356	0.964
Cohesion Compactness	200	0.756	[0.740, 0.772]	0.116	0.377	0.957
Exchange Compactness	198	0.639	[0.621, 0.656]	0.127	0.190	0.860
Buildable Land Compactness	198	0.713	[0.695, 0.731]	0.128	0.220	1.000

All of these four compactness indices are correlated, which comes as no surprise. Their correlations appear in **table 2** below. Of particular interest to us is that the three compactness indices that do not take buildable land into account are very highly correlated. All their correlations are 0.95 or higher. In a sense, therefore, these three indices can be used interchangeably to describe the compactness of urban extents. Indeed, in the remaining sections of this article, we report on one or the other of these three indices with the understanding that similar results have been obtained for the other two others as well, and that these results were not different in any substantial way.

Table 2: The Pearson Correlation Matrix for the four compactness indices for the year 2014.

	Proximity Index	Cohesion Index	Exchange Index	Buildable Land Index
Proximity Index	1			
Cohesion Index	0.996	1		
Exchange Index	0.965	0.950	1	
Buildable Land Index	0.628	0.625	0.635	1

As we saw in **table 1** above, there is considerable variation in compactness values among cities in the global sample. This variation is difficult to envision without looking at maps of the extents of cities and comparing them. The four figures below present the variation in compactness indices in the global sample.

Figure 8 shows cities with the highest 16 and lowest 16 Cohesion Index values in the global sample. The orange circle is the Equal Area Circle centered at the centroid of their urban extent in 2014. In the top left corner of each image, values are given for the Cohesion Index (COH), for the Proximity Index (PRO), and for the Exchange Index (EXC). The maps of the urban extents of cities are shown in declining order of their Cohesion Index, from the highest in the sample, Shanghai, with a Cohesion Index of 0.96, to the lowest in the sample, Cabimas, Venezuela, with a Cohesion Index of 0.36.

Figure 9 shows cities with the highest 16 and lowest 16 Buildable Land Index values in the global sample. The orange circle is the Buildable Land Circle centered at the centroid of their urban extent in 2014. In the top left corner of each image, values are given for the Buildable Land Index (BLD). The maps of the urban extents of cities are shown in declining order of their Buildable Land Index, from the highest in the sample, Caracas, Venezuela, with an Index of 1.00, to the lowest in the sample, Beira, Mozambique, with a Buildable Land Index of 0.22.

Figure 10 shows cities with the highest 28 and lowest 4 ratios between the Buildable Land Index and the Exchange Index—i.e. the Compactness Correction Factor—in the global sample. The dark orange circle is the Equal Area Circle and the light orange circle is the Buildable Land Circle, both centered at the centroid of their urban extent in 2014. In the top left corner of each image, values are given for the Compactness Correction Factor (CCF), for the Buildable Land Index (BLD), and for the Exchange Index (EXC). The maps of the urban extents of cities are shown in declining order of their Compactness Correction Factor (CCF), from the highest in the sample, Caracas, Venezuela, with a Factor of 1.63, to the lowest in the sample, Modesto, USA, with a Factor of 1.00. A factor of 1.63 in Caracas means that its compactness increases by 63 percent when we consider the physical barriers surrounding it. A factor of 1.00 in Modesto means that its compactness remains unchanged when we consider the physical barriers, namely there are no physical barriers around it that prohibit it from becoming more compact.

Finally, **figure 11** shows 16 cities with the highest increase and 16 cities with the highest decline in the Cohesion Index between 1990 and 2014 in the global sample. In the top left corner of each image, values are given for the Cohesion Index in the earlier period (COH.T1) and for the Cohesion Index in the later period (COH.T3). The maps of the urban extents of cities are shown in declining order of their increase in compactness between 1990 and 2014, from the city with highest increase in the Cohesion Index in the sample, Qingdao, China, with an increase of 63 percent, to the city with highest decrease in the Cohesion Index in the sample, Beira, Mozambique, with a decrease in cohesion compactness of 51 percent.

Given the maps of the urban extents of all cities in the global sample of cities, we can measure their compactness using the indices and ratios defined here. The above figures allow us to visually observe the great variation in shape compactness among cities as well. The next section seeks to explain and account for this variation, presenting a set of findings pertaining to the compactness of individual cities as well as to the average compactness of the universe of cities as a whole.

Figure 8: Cities with the highest 16 and lowest 16 values for the Cohesion Index in the global sample. The orange circle is the Equal Area Circle centered at the centroid of their urban extent in 2014.

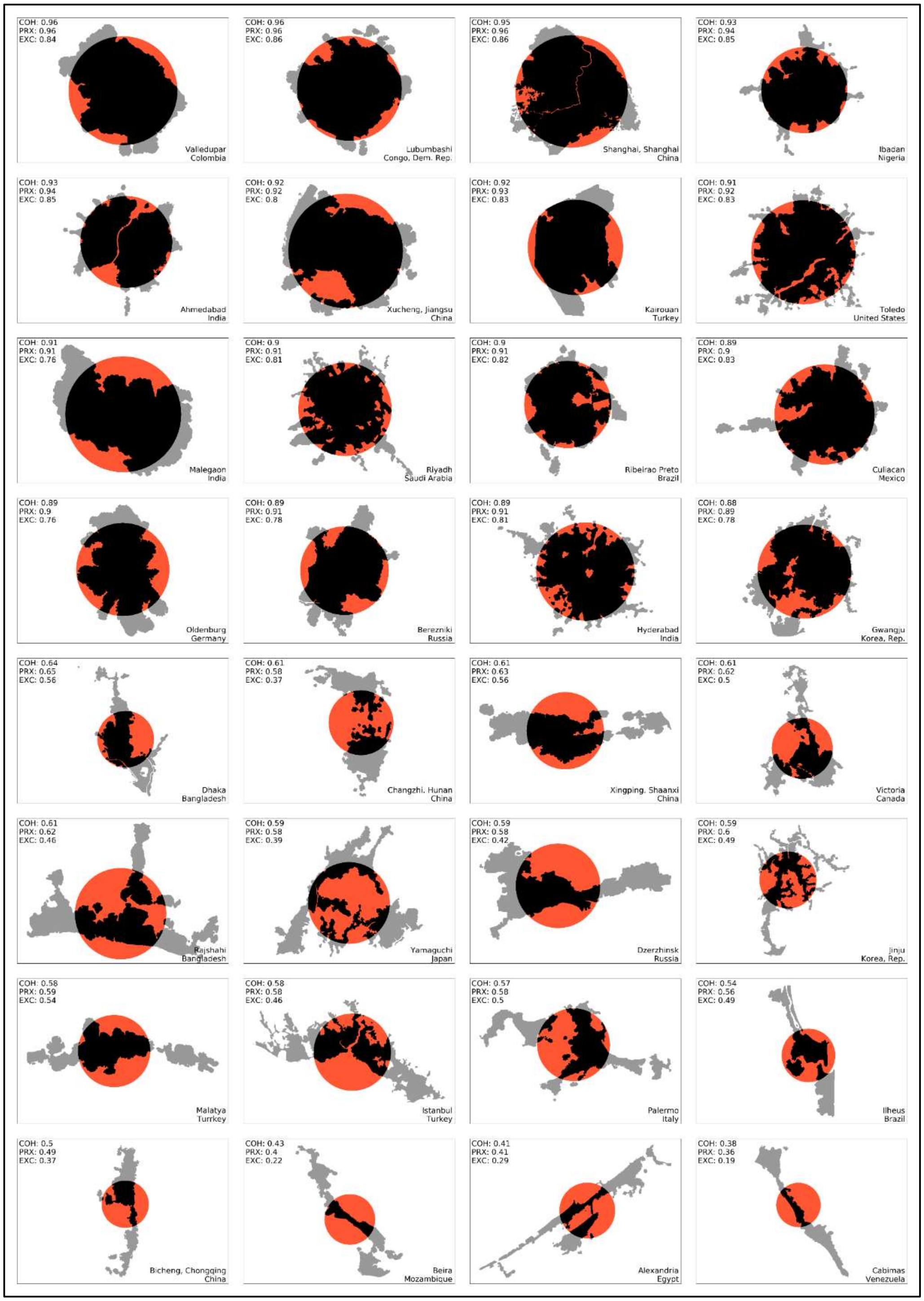


Figure 9: Cities with the highest 16 and lowest 16 Buildable Land Index values in the global sample. The orange circle is the Buildable Land Circle centered at the centroid of their urban extent in 2014.

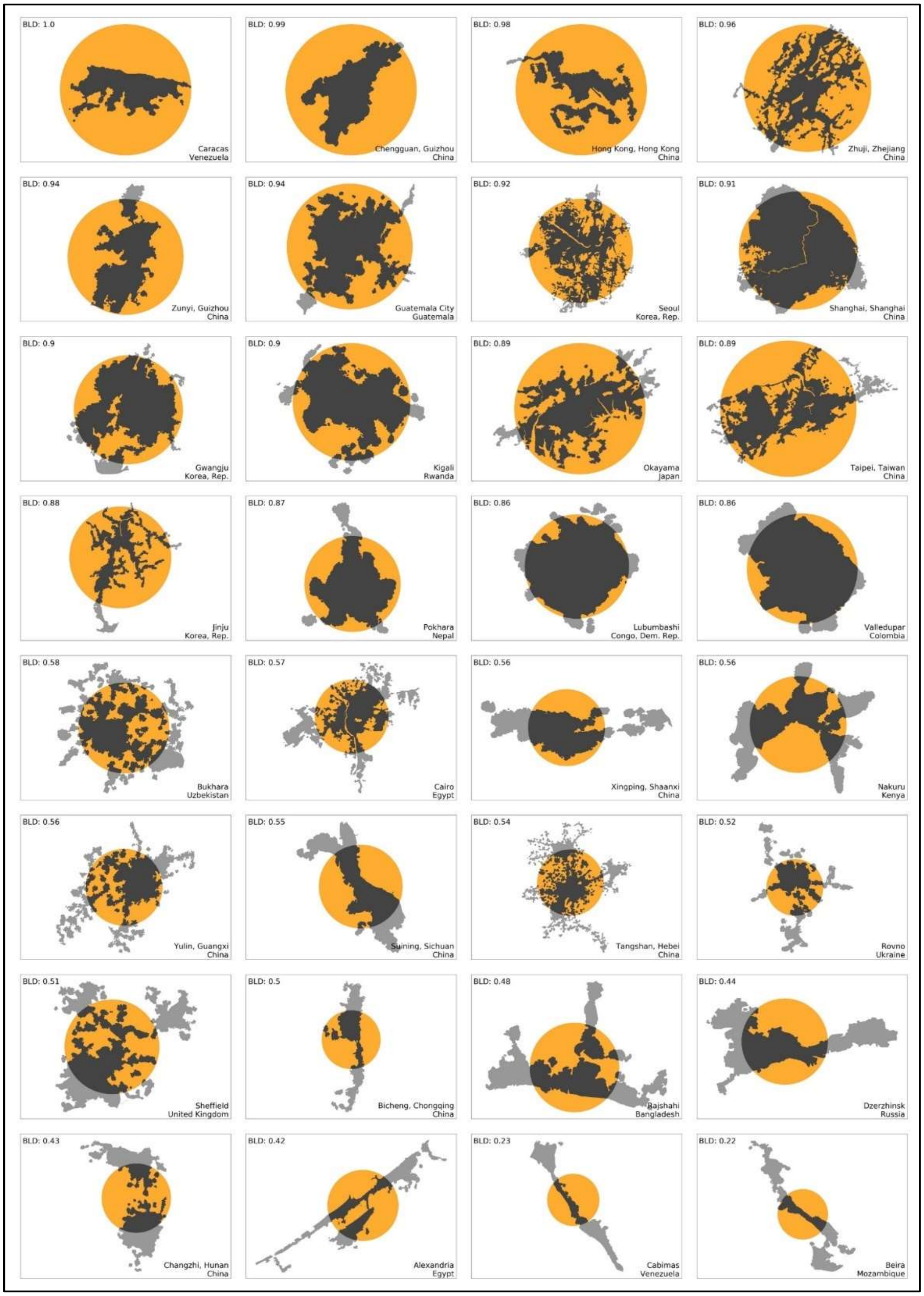


Figure 10: Cities with the highest 28 and lowest 4 Compactness Correction Factors—i.e. the ratios between the Buildable Land Index and the Exchange Index—in the global sample. The dark orange circle is the Equal Area Circle and the light orange circle is the Buildable Land Circle, both centered at the centroid of their urban extent in 2014.

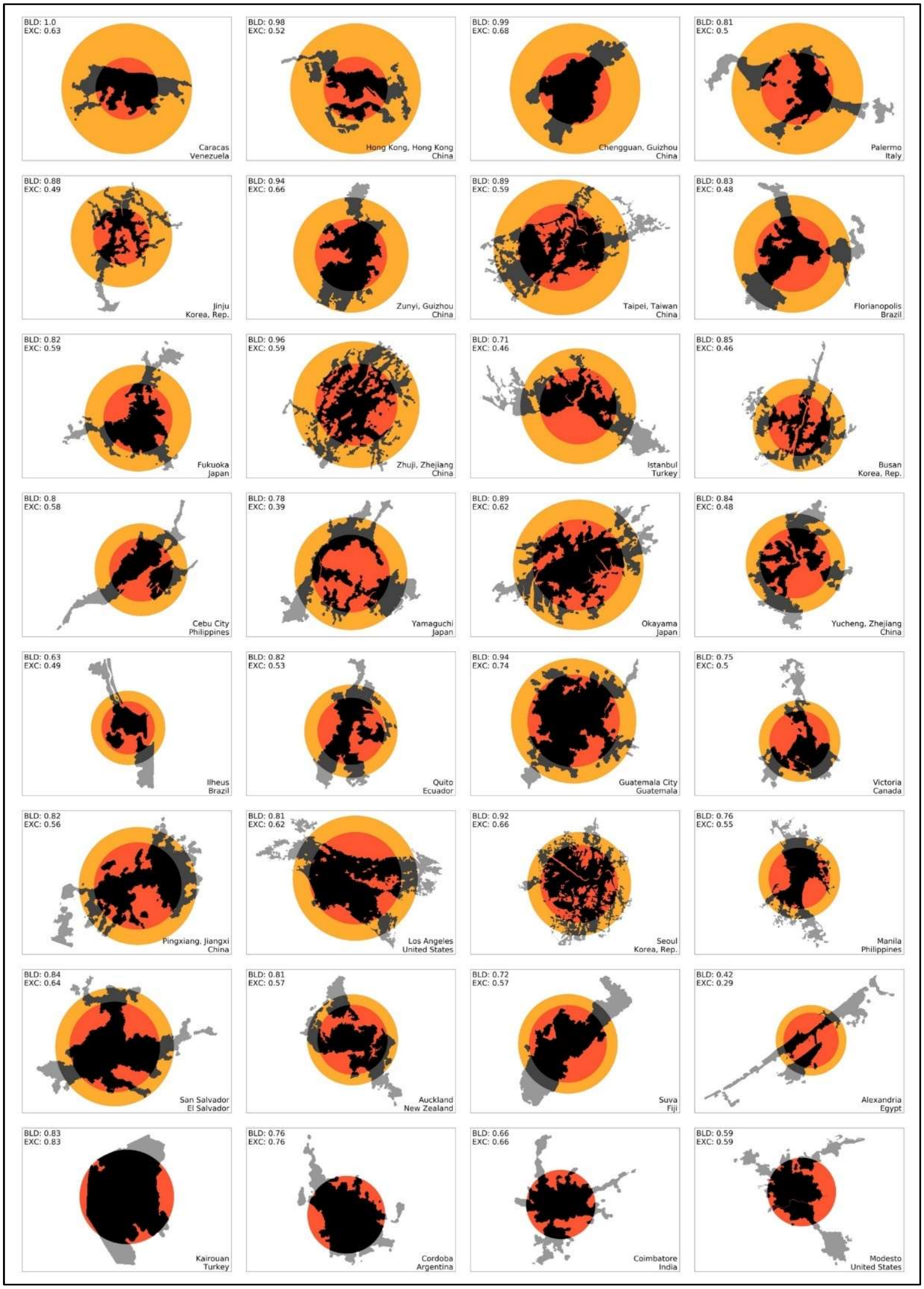
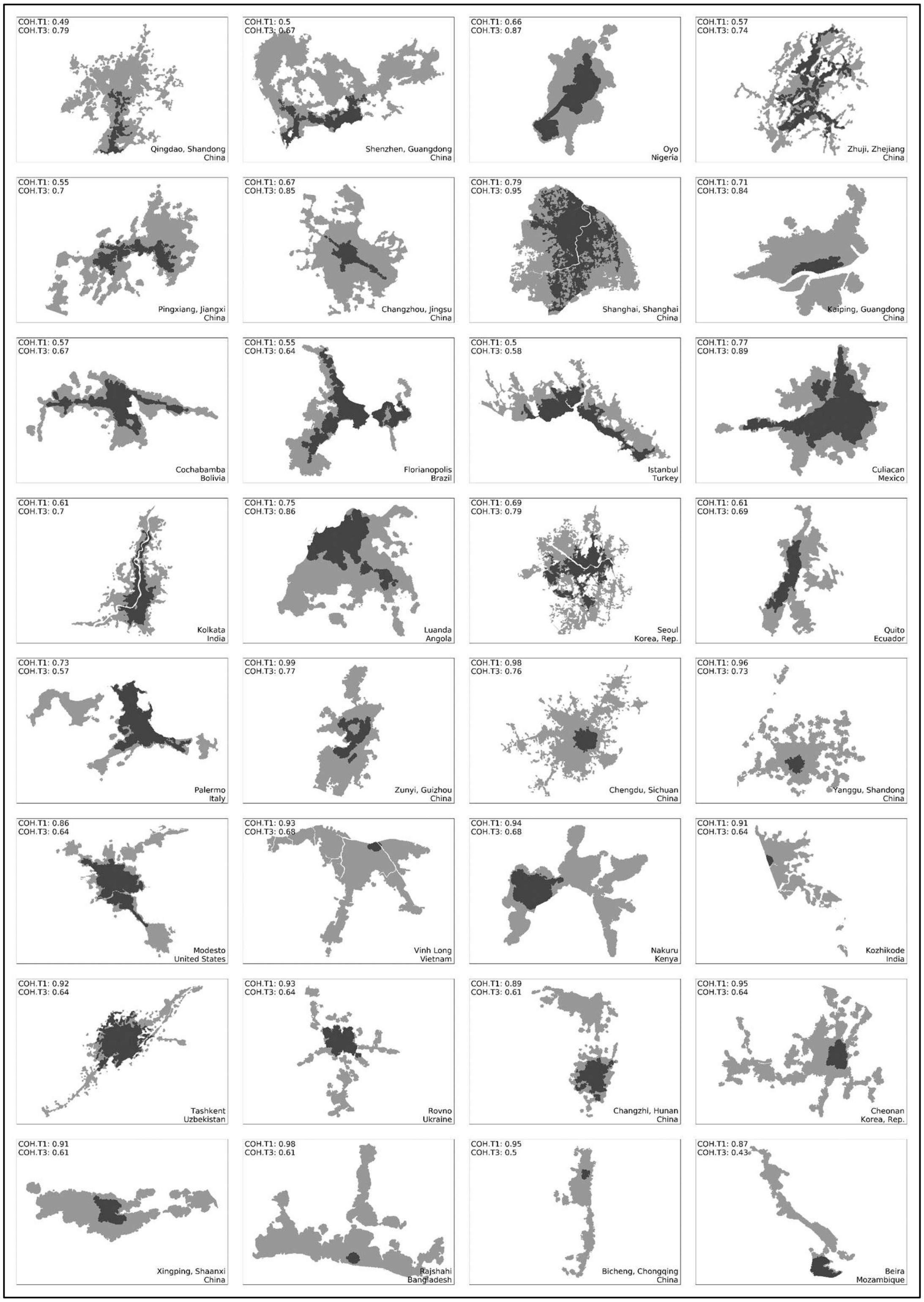


Figure 11: 16 cities with the highest increase and 16 cities with the highest decline in Cohesion Index values between 1990 and 2014 in the global sample.



3. Findings

In this section of the paper, we seek to provide answers to the three key questions raised earlier:

- How do we account for and explain the variation in shape compactness among cities?
- Have cities become significantly more or less compact in recent years?
- How do compactness and density affect the average distance traveled in cities once we account for differences in their populations?

3.1. Explaining the Variation in Shape Compactness Among Cities

A key finding of this study is that the shape compactness of cities is independent of city population size, city area, city population density, and city per capita income. In 2014, for example, the correlation coefficients between each of the compactness indices with city population size were not statically significant. There was no significant difference in the compactness values for large and small cities. There was no significant direct correlation between these three indices with city area, with city population density, and with average city per capita income either, and this was true for the 1990 and the 2000 periods as well.

We noted earlier that people come to cities to be closer to each other, so as to facilitate the exchange of goods, services, and information between them and so as to make possible more extensive and more diverse human contact among them. Other things being equal, the strong forces, tendencies and intentions attracting people to each other in cities should make the shapes of their urban extents compact. The more compact their urban extent, the closer people will be to each other. In other words, we can take it as a given that cities will seek to be compact in shape if they are not prevented from becoming compact by forces, tendencies and intentions that pull them apart, making them less compact. Explaining the variation in shape compactness among cities must thus focus on the drivers of non-compactness in cities, for it is these drivers that can explain why cities are not as compact as expected.

We have identified six main drivers of non-compactness in cities:

(1) Physical barriers;

(2) Merging of adjacent settlements;

(3) Inter-city roads and rail lines;

(4) Land use restrictions;

(5) Beachfront preferences; and

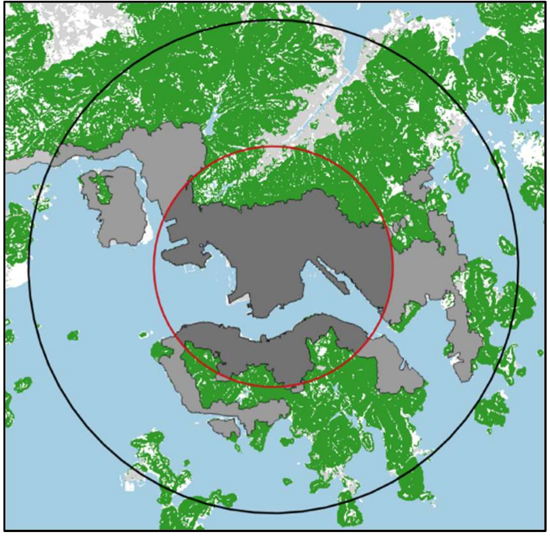
(6) Land market distortions.

It goes without saying that it has not been possible to obtain good data on all of these six drivers for all the cities in the global sample. Each of these drivers of non-compactness will be discussed below with an elaboration of one or more specific examples of cities that illustrate the action of this particular driver on their extents. Where possible, we shall present statistical data pertaining to these drivers of non-compactness for the global sample of cities as well.

3.1.1. Physical barriers

Cities need land to expand, and that land needs to be generally flat. Steep slopes, in our definition slopes greater than 15 percent (8.53°), typically prohibit city building. As we noted earlier, only a very small share of the urban extents of cities in 2014, for example, (5.8±1.2 percent, weighted average with 95 percent confidence interval), was on slopes exceeding 15 percent. And like steep slopes, bodies of water also prevent construction. Cities that are surrounded by steep slopes *and* bodies of water, like Hong Kong, for example, cannot have a very high level of shape compactness. Indeed, the Proximity, Cohesion and Exchange Compactness values for Hong Kong in 2014 were all in the lower fifth of the global sample of 200 cities. By all three measures Hong Kong was by no means compact. But since we are interested in cities becoming more compact—so as to reduce their greenhouse gas emissions from ground transport, for example—we can legitimately ask whether Hong Kong could be made *more* compact. The answer to that is an emphatic no: Given the physical barriers surrounding it, Hong Kong is very close to being as compact as it can be. Its Buildable Land Index value in 2014 was 0.98. In other words, 98 percent of the area of Hong Kong's Buildable Land Circle—a circle centered at the city's centroid and containing enough buildable land for its entire urban extent—was taken up by its urban extent, while only 2 percent of its urban extent was outside that circle (see **figure 12**).

Figure 12: The 2014 Urban Extent of Hong Kong (grey) occupies only 52 percent of its Equal Area Circle (red) and is, therefore, among the least compact cities in the global sample. Yet it occupied 98 percent of its Buildable Land Circle (black), confirming that—given its physical environment—it is as compact as can be (Buildable land shown in white, non-buildable land in green).



More generally, cities surrounded by natural barriers—be they high slopes or bodies of water—tend to be less compact. The Buildable Land Ratio—the area of Equal Area Circle divided by the area of the Buildable Land Circle—is a measure of the degree to which a given city is exposed to natural barriers. The smaller the ratio, the larger the exposure, and where there are no physical barriers, that ratio is 1.0. Are cities with lower ratios significantly less compact? Yes. The Buildable Land Ratio in 2014 and Cohesion Index are significantly correlated to each other at the. For 10 percent increase in Buildable Land Ratio we can expect the Cohesion Index to increase on average by 1.8 percent (Adjusted R-squared = 0.14). Similar results can be obtained for the other two compactness indices and for all three indices in the 1990 and 2000 periods as well. These findings confirm that cities surrounded by natural barriers are less compact than cities on an open plane.

The Compactness Correction Factor—the percentage increase in exchange compactness once the Equal Area Circle is replaced by the Buildable Land Circle—tells us by how much the shape compactness of the urban extent of a city increases when we take buildable land into account. The larger the factor, the less buildable land is available in close proximity and the further out the city must extend in order to find more buildable land. It stands to reason, therefore, that in cities with high Compactness Correction Factors there will be more construction on steeper slopes closer to the city center. Although building on steeper slopes in more accessible locations may be more expensive and possibly riskier, the savings on transport may exceed these extra costs and extra risks. Indeed, we find that the higher the Compactness Correction Factor in a city, the higher the share of its urban extent that is on slopes higher than 15 degrees. We find that in the global sample of cities the share of the area of urban extents on slopes higher than 15 degrees highly correlates with the Compactness Correction Factor, with R-squared of 0.23. For 10 percent increase in the Compactness Correction Factor, we can expect a 1.4 percent increase in the share of the urban extent on slopes exceeding 15 percent. Caracas, Venezuela, is an outlier. It has a Compactness Correction Factor of 0.5, the highest in the global sample of cities. Not surprisingly, 30 percent of its urban extent is in areas with slopes that are steeper than 15 percent (see **figure 13** below). We note here that the strong effect of the Compactness Correction Factor on construction on steeper slopes is an important finding that has serious policy implications: Building on steeper slopes can increase the shape compactness of urban extents. Cities facing serious physical constraints face an important choice: Extending further out and becoming less compact in the process, or building on steeper slopes closer to the city center.

Figure 13: Caracas, Venezuela, has the highest Compactness Correction Factor in the global sample of cities. Given its physical constraints, it is as compact as can be. 30 percent of the urban extent of the city is built on slopes steeper than 15 percent.[1]

Steep slopes are only one kind of barrier to urban expansion that tends to affect the compactness of their footprints. Water bodies are another. Cities built along coastlines tend to be less compact, and for two reasons. The first reason is that they can only expand inland, while cities surrounded by flat, open land can expand in all directions. In a typical city on the coast, the Central Business District (CBD) is situated along the water. If the city were to be built in concentric rings about the CBD, its shape would be that of a half-circle, a shape that is clearly less compact than that of a circle. It can be asserted that the average distance to the CBD in such a city would be 1.41 ($\sqrt{2}$) times larger than the average distance to the CBD in its Equal Area Circle. 67 cities in the global sample of cities are coastal cities. They were found to have significantly lower Proximity Index, Cohesion Index, and Exchange Index values than the remaining cities in the sample in any of the three time periods, 1990, 2000, and 2014. In 2014, for example, the average Proximity Index value with 95 percent confidence interval for the 67 coastal cities was 0.69±0.3, which was significantly lower than the average of the rest of the cities: 0.79±0.2. Similar results were found for the Cohesion Index (0.68±0.3 versus 0.78±0.22), and the Exchange Index (0.56±0.3 versus 0.67±0.2). The second reason that coastal cities are less compact is the common preference of their residents for occupying beachfront properties or for being close to the seashore. This preference will be discussed as a separate driver of non-compactness below.

---

[1] **Source: Wikimedia Commons, online at: https://commons.wikimedia.org/wiki/File:Slums_in_Venezuela,_Caracas.jpg**

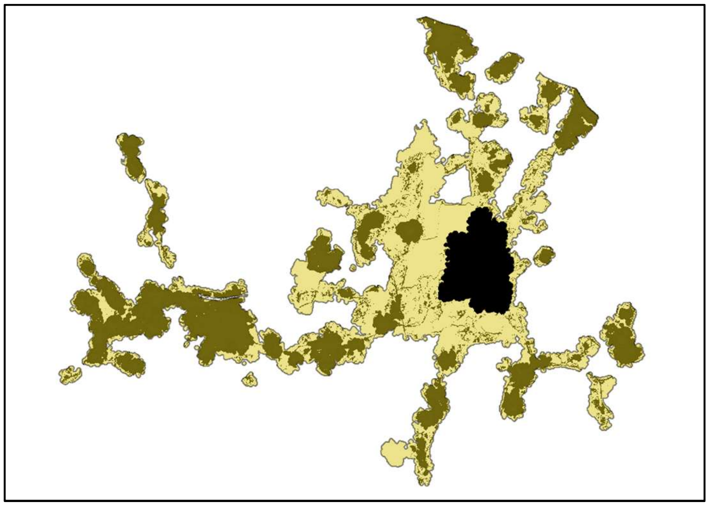
3.1.2. Merging of adjacent settlements

As cities expand outwards, their urban extents come to include settlements—both cities and villages—that hitherto were self-contained, freestanding ones. This process can turn cities that were highly compact and near circular in shape to one long string of connected settlements. By analogy, imagine a drip irrigation pipe where water comes out in drops from holes punched in the pipe at regular intervals, wetting the earth around these holes in expanding circles that eventually blend together into a long, wet stretch of land. Examples of cities that follow that pattern abound. The Rhine-Ruhr area in Germany, the U.S. Northeastern seaboard, the Tokyo-Osaka corridor in Japan, or the Beijing-Tianjin-Hebei agglomeration (BTHA) in China are typical examples. Connecting cities into corridors can occur naturally, but it can also be the result of intentional policy: The Delhi-Mumbai and the Chennai-Bangalore corridor in India or the Northern Corridor in Haiti are recent examples. Merging adjacent settlements typically results in an abrupt decline in compactness because it occurs when two or more separate settlements, each of which can be quite compact, merge into one another.

We calculated the shares of the added areas to the urban extents of the 200 cities in the global sample between 1990 and 2014 attributed to four categories: (1) infill: Building within the urbanized open space of the previous period; (2) extension: building at the edge of and away from the urban extent of the earlier period; (3) leapfrog: building in areas surrounded by rural open space; and (4) inclusion: incorporating settlements built earlier into the urban extent in the later period. The average values for the global sample for the period 1990-2014 were: infill—24 percent, extension—33 percent, leapfrog—1.3 percent, and inclusion—24 percent. Inclusion corresponds to the merging of settlements into common urban extents. We tested the following hypothesis: The greater the share of the added area in 'inclusion', the less compact the resulting urban extent. We tested this hypothesis in a linear regression model with the percent change in the Proximity Index as the dependent variable, using three independent variables that can be associated with a change in the shape compactness of cities: National GDP per capita change during this period, the annual rate of growth of the urban extent during this period, and the share of 'inclusion' in the added area during this period. The model appears in **table 6** on page 35.

A good example of a city that has become less compact over time because of inclusion is Cheonan in South Korea. Between 1990 and 2014, Cheonan has experienced the 6th largest decrease in compactness in the global sample of cities. Its Cohesion Index value, 0.95 in 1991, declined to 0.64, less than two-thirds that value, by 2014, largely due to the inclusion of existing settlements in its urban extent. Between 2000 and 2014, for example, 57 percent of the area added to the city was added through the inclusion of existing settlements (see **figure 14** below).

Figure 14: As Cheonan, South Korea, expanded between 1991 (black) and 2014 (light green), it merged with a large number of settlements surrounding it (light and dark red). 53 percent of the area added to the city's urban extent during this period was attributed to 'inclusion', reducing its Cohesion Index from 0.95 to 0.64.



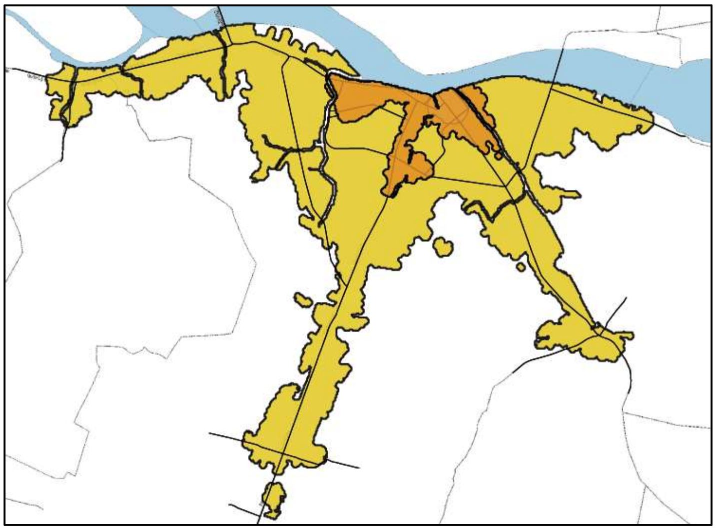
3.1.3. Inter-city roads and rail lines

The compactness of contemporary cities—where trip destinations are distributed everywhere rather than predominantly at the city center—hinges on the ability of people to move in all directions at equal speeds. This, in turn, requires a high density of crisscrossing arterial roads—and, in large metropolitan areas, a high density of crisscrossing rail lines—leading in all directions everywhere.

In theory, it can be safely assumed that urban dwellers will seek to minimize travel time rather than travel distance to their favorite destinations when choosing where to locate their homes. Other things being equal, when destinations are all located at the city center—or, alternatively, when destinations are distributed everywhere—the resulting urban extent will acquire a near circular shape. When the speed on some roads, say all roads running north-south, is higher than the speed of the remaining roads, say the roads running east-west, the urban extent will acquire the shape of an elongated ellipse, with a longer north-south axis. Similarly, when radial roads are generally faster than circumferential roads, the urban extent will acquire the shape of a star. In both cases, that of the ellipse and the star, minimizing travel time will lead to higher average travel distances and—to the extent that travel cost, energy expended, and greenhouse gas emissions are dependent on distance traveled—to higher personal and social costs.

Typically, the density of arterial roads falls dramatically at the urban edge, and the only transportation corridors that extend away from the urban extent into the rural periphery are inter-city roads and rail lines that lead away from the city center, or rural roads and lanes leading to nearby villages on the urban periphery. Naturally, where the provision of public works—and, particularly, arterial roads—at the urban periphery lags behind the demand for peripheral land with good access to the city, urban development takes place along existing inter-city roads, as well as around freeway intersections and railway stations located along inter-city rail lines. When this happens, cities expand in tentacles along transportation corridors, becoming less compact in the process. Vinh Long, Vietnam, with an Exchange Index of 0.52 in 2014, clearly falls into this category (see **figure 15** below). It has experienced a significant reduction in its compactness as it extended along inter-city roads leading out of its center. Its Cohesion Index, for example, declined from 0.93 in 1989 to 0.68 in 2014.

Figure 15: By 2014, Vinh Long, Vietnam, has expanded along four intra-city roads leading away from its center becoming less compact than it was in 1990 in the process.



This form of urban expansion leads to building at further distances from the city center, while leaving areas closer to the city center undeveloped, simply because they are less accessible than areas located further away. This, in turn, as noted earlier, increases the average travel distance in the city. Areas on the urban periphery that are immediately adjacent to the built-up urban extent and closer to the city center do get built upon eventually, but possibly at a slower rate, as the local street network—where travel may be slower than on inter-city roads—is slowly extended outwards. Again, when expansion areas in some directions can only be reached at slower road speeds than expansion areas in other directions, the city may retain its star shape. Clearly, all cities are connected to other cities by inter-city roads and, more often than not, rail lines as well. But not all urban extents have tentacles extending outwards along these inter-city transportation arteries. Why some cities have such tentacles while others do not is a question that must await further study.

In an important sense, however, the formation of such tentacles or their absence can and should be attributed to policy because, as we noted, it does affect average travel distances. The process of rendering the urban extent more compact by building in areas that are closer to the city center—rather than in areas located further away along inter-city roads—can, of course, be accelerated by preparing an efficient arterial road grid on the entire urban periphery in advance of development, an effective planning initiative that is rarely, yet occasionally, implemented and will be discussed further in the conclusion of this essay. In the absence of such an initiative, urban extents may tend to become and remain less compact as urban expansion proceeds further outwards along intercity roads and rail lines.

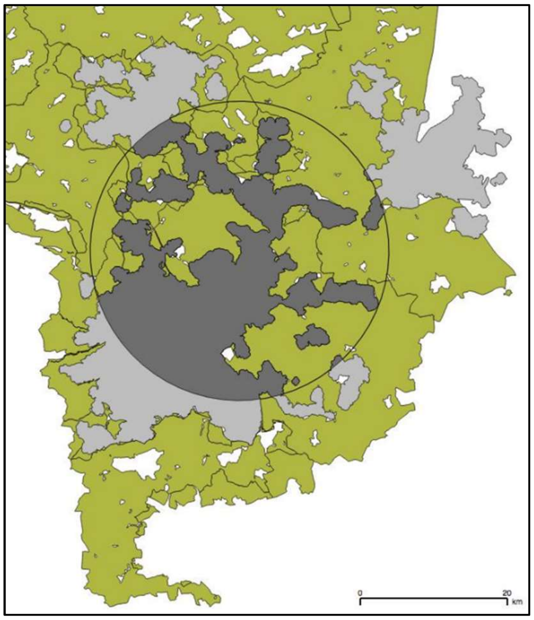
3.1.4. Land Use Restrictions

A number of countries—China, Egypt, and the United Kingdom, to take a few examples—place strict limits on the conversion of rural lands to urban use. In China, for example, there are laws that mandate that the amount of cultivated land in each province must remain fixed, requiring provincial governments to replace cultivated land converted to urban use with new cultivated land, a requirement they find difficult if not impossible to meet. Urban expansion plans are reviewed by the central government and are often required to restrict the amount of cultivated land lost to expansion (Angel, Valdivia and Lutzy, 2007). This often results in highly fragmented urban extents—i.e. a smaller level of saturation of urban extents by built-up areas—and hence in larger urban extents, but not necessarily in less compact ones. Indeed, the urban extents in the Chinese cities in the global sample of cities are significantly less saturated by built up areas—i.e. they contain more open space and vacant land—than non-Chinese cities in the sample, but they are not less compact than other cities in the global sample.

In the United Kingdom, there are expansive green belts surrounding and fragmenting all major metropolitan areas, and there are strict regulations limiting or altogether preventing construction within these green belts: "The extent of the designated Green Belt in England as at 31 March 2017 was estimated at 1,634,700 hectares, around 13 percent of the land area of England. Overall there was a decrease of 790 hectares (less than 0.05 percent) in the area of Green Belt between 31 March 2016 and 31 March 2017." (U.K. Department of Communities and Local Government 2017).

Sheffield, England, is in the global sample of cities. Its Cohesion Index in 2014 was quite low, ranked the 32nd lowest in the global sample of 200 cities, largely because of its green belt. The city is situated in relatively flat land and could, in principle, be very compact. Yet 49 percent of its Buildable Land Circle is occupied by its greenbelt (see **figure 16** below). As a result, in 2014 its Buildable Land Index was 0.51, the 9th lowest value in the global sample of cities. In other words, policy decisions, in this case land use restrictions can have a major on the compactness of cities. Sheffield was only half as compact as it could be if its urban expansion was not constrained by its green belt. Its expansion now takes place mostly towards the Northeast, beyond the greenbelt. Again, while the green belt provides a high level of amenity value in Sheffield, it increases average trip length by an average of some 50 percent, substantially increasing energy use, greenhouse gas emissions, commute times, and infrastructure line length. Unfortunately, detailed maps on the location of lands that cannot be built upon because of land use restrictions of one kind or another are not available for all the cities in the global sample. Hence the overall effect of land use restrictions on the shape compactness of urban extents cannot be investigated in sufficient detail at the present time.

Figure 16: The greenbelt in Sheffield, England, occupies half of its Buildable Land Circle, making Sheffield one of the least compact cities in the global sample of cities. The urban extent of the city is shown in dark grey inside the circle and in light grey outside it.



3.1.5. Beachfront Preferences

As we noted earlier, the location of cities along coastlines necessarily makes their urban extent less compact because the coast acts as a constraint to urban expansion. Over and above that, however, cities expand further along coastlines than they would need to expand because of that constraint. There is an amenity value to occupying beachfront properties, to having lake, sea, or ocean views, and to being in close proximity to the water. As a result, we find coastal cities that are considerably more elongated than they would be if their shape were only dictated by physical constraints to their development. Preference for locating in close proximity to water bodies—the ocean, a lake, or a wide river—tend to extend the built-up areas of cities in linear form and away from a more circular form. This is clearly observable in cities in the global sample like Alexandria along the Mediterranean Coast in Egypt, Cabimas along the shores of Lake Maracaibo in Venezuela, and Cebu City in the Philippines. It is also evident in cities not in the global sample like Miami, Florida, or Montevideo, Uruguay.

Alexandria, to take one example, could have expanded in a southeasterly direction into the Nile delta, becoming more compact in the process. Instead it has expanded to the southwest along a thin sliver of land along the water, becoming less compact in the process. Its Buildable Land Index in 2014, 0.42, was the third lowest in the global sample of 200 cities (see **figure 16** below).

Figure 16: The city of Alexandria in Egypt, expanded further along the Mediterranean shore between 1990 (black) and 2014 (gray), becoming less compact in the process.

3.1.6. Land market distortions

When land markets function properly, there is very little leapfrogging as cities expand outwards. Most leapfrogging is 1km or less away from the built-up areas of cities (Burchfield *et al* 2005). This is not the case, however, when land markets are distorted by government action. A classic case is that of Mexico, where the government's housing finance agency, INFONAVIT (Instituto del Fondo Nacional de la Vivienda para los Trabajadores, or the National Fund for Workers' Housing), founded in 1972, accounts for almost three-quarters of all housing loans and solicits housing directly from large developers for allocation to its clients.

> Over the last several decades, INFONAVIT has encouraged developers to build housing with a price point as the main guiding criteria and provided those developers an almost guaranteed client base. While this practice may have offered a larger share of the population access to housing, it meanwhile led to the construction of thousands of houses for which there was very limited demand in subdivisions far from city centres, job opportunities, and in some cases without adequate infrastructure. This problem would not have been as severe in a more market-based system in which developers that built unwanted houses would have gone out of business quickly; given the close ties that were established between INFONVIT and a handful of large homebuilding firms in the late 1990s and formalized through the Housing Commitment in 1998, the housing finance system continued to support an ultimately suboptimal housing model that had important implications for the country's urban development outcomes. (OECD 2015, 136-37).

This practice has resulted in the location of large housing estates in outlying ex-urban areas, rendering numerous Mexican cities less compact than they would have been in the absence of such interventions (see **figure 17**). Locating housing in distant locations also resulted in increased levels of abandonment. The overall rates of abandonment are difficult to calculate and their attribution to distant locations is difficult to prove yet it is clear that they are very high. "INFONAVIT reported that between 16 percent and 20 percent of INFONAVIT credits originated between 2006 and 2010 were for homes that were ultimately uninhabited" (OECD

2015, 131). Again, data on land market distortions in other cities in the global sample are difficult to obtain and assess and the evidence from Mexican cities can, at best, be considered only anecdotal and illustrative. A more elaborate study of the effects of land market distortions on the compactness of urban extents must await the collection of better and more extensive global data.

Figure 17: INFONAVIT Housing on the urban periphery in Mexico (Photo Credit: Habitat D.F.)

To conclude, in this section of the paper we have presented evidence, some of it pertaining to individual cities and some of it pertaining to the global sample of cities, that seeks to explain the observed variations in the shape compactness in cities the world over. While the explanations given and the statistical results presented are only preliminary in nature, a broad perspective on the variations of shape compactness of urban extents does begin to emerge. We can begin to distinguish some of the key forces—there may be others, yet to be discovered—acting on the shape compactness of cities and to see which ones are subject to policy intervention and which ones are not. In this context, it is interesting to explore whether the totality of forces now acting on the shape compactness of cities is making them more or less compact. This question is addressed in the following section.

3.2. Have cities become significantly more or less compact in recent years?

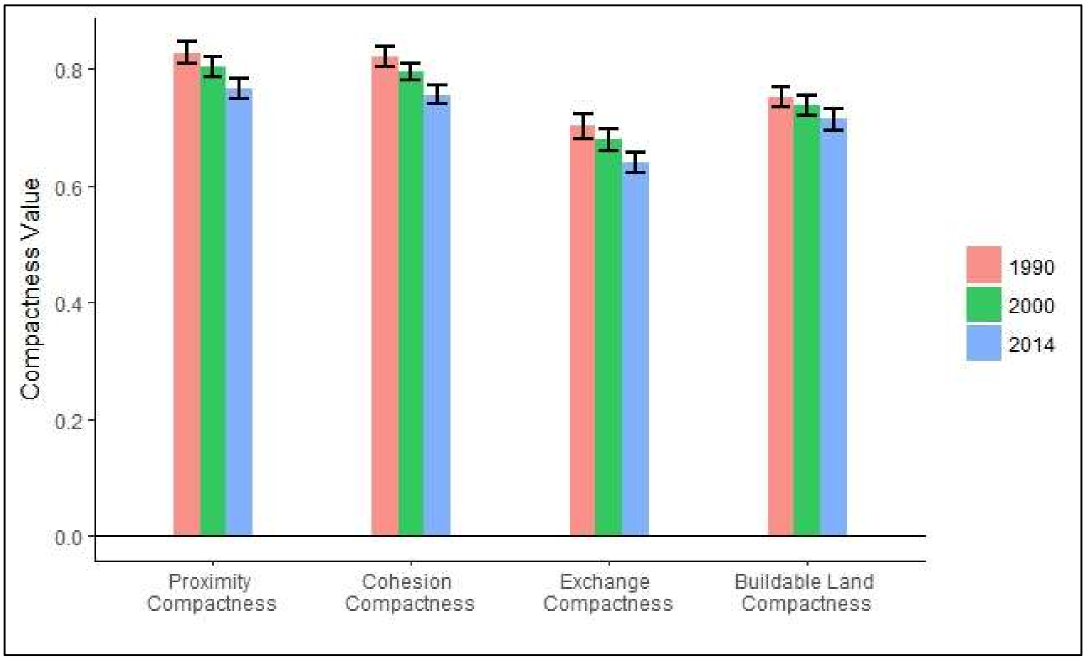
The global sample of cities is representative of the universe of cities. As we explained earlier, the data on the compactness indices for the global sample of cities can be weighted to obtain results for the universe of cities as a whole.

Table 3 and **figure 18** below shows the average values of the four compactness indices defined earlier for the universe of cities as a whole, all 4,231 cities that had 100,000 people or more in 2010. The results were obtained as weighted averages of the sample of cities, weighted by the number of cities in the universe of cities represented by each city in the sample. The table and the figure also show the 95 percent confidence intervals for these indices. The confidence intervals for the year 1990 do not overlap with those of 2014, suggesting that the decline in the average value in the 1990-2014 period as a whole was indeed statistically significant at the 95 percent confidence level. This allows us to conclude that over the 1990-2014 period the shape compactness of cities has been in significant decline.

Table 3: Means and 95 percent confidence intervals for the four compactness indices in 1990, 2000, and 2014, weighted by city weights.

Variable	**1990**	**2000**	**2014**
Proximity Index	0.828 [0.810, 0.845]	0.803 [0.787, 0.820]	0.766 [0.749, 0.783]
Cohesion Index	0.820 [0.803, 0.838]	0.795 [0.779, 0.811]	0.756 [0.740, 0.772]
Exchange Index	0.702 [0.680, 0.723]	0.679 [0.661, 0.696]	0.639 [0.621, 0.656]
Buildable Land Index	0.752 [0.734, 0.769]	0.737 [0.721, 0.753]	0.713 [0.695, 0.731]

Figure 18: All four types of compactness indices decreased over the three periods. The decline in compactness between 1990 and 2014 was significant at the 95 percent confidence level.



We can obtain a stronger result by looking at the weighted average of the *change in compactness* in individual cities between each of the two time periods. We applied a weighted paired t-test to compare the compactness values in the two periods for individual cities. It tests if the differences in compactness indices between two periods are significantly below or above zero. **Table 4** below displays the results of this test. Since the 95 percent confidence intervals of all the comparison pairs are below zero, we can infer that the weighted averages of all the four indices decreased significantly across all periods: between 1990 and 2000, between 2000 and 2014, and from 1990 to 2014.

Table 4: Weighted means and 95 percent confidence intervals of the differences in the four compactness indices between periods. The change between periods was the difference in index values between the two periods.

Variable	1990 to 2000	2000 to 2014	1990 to 2014
Proximity Index	-0.024 [-0.036, -0.013]	-0.038 [-0.052, -0.024]	-0.062 [-0.080, -0.045]
Cohesion Index	-0.026 [-0.037, -0.014]	-0.039 [-0.053, -0.026]	-0.065 [-0.082, -0.048]
Exchange Index	-0.023 [-0.037, -0.009]	-0.040 [-0.055, -0.025]	-0.063 [-0.083, -0.043]
Buildable Land Index	-0.014 [-0.027, -0.001]	-0.024 [-0.039, -0.010]	-0.038 [-0.058, -0.019]

Given this stronger result, we can conclude that the shape compactness of cities the world over has been in significant decline during the 1990-2000 and the 2000-2014 periods. The overlapping confidence intervals in **table 4** do not allow us to determine whether the decline in compactness during the 2000-2014 period was more pronounce than the decline in the 1990-2000 period.

We can also ask ourselves whether the decline in compactness is of the same magnitude in different countries and world regions. The small size of the global sample of cities does not allow us to arrive at statistically significant results for countries and regions, and it does not allow us to differentiate between cities in more developed countries and cities in less developed countries. We noted earlier that the shape compactness of cities does not vary significantly with income. Indeed, there is no difference in the weighted average Proximity Index in 2014, for example, between cities in more developed countries, 0.78±0.02 (95 percent confidence interval) and cities in less developed countries, 0.77±0.03. Index values were not statistically different in the other two periods. The other compactness indices also showed no difference in any of the three periods.

We did detect a difference between cities in more developed countries and cities in less developed countries in the decline in shape compactness over time. As **table 5** below shows, that decline was more pronounced in cities in less developed countries. The table shows that the weighted average decline in compactness index values was greater in the cities in less developed countries. The association is weak, as only the Proximity Index and Cohesion Index show a significant difference (at 95 percent confidence level) in the magnitude of decline between cities in more developed countries and cities in less developed countries.

Table 5: the weighted means and confidence intervals for the four compactness indices. P-values were obtained using a weighted two-sample t-test comparing, the magnitude of the decrease in compactness values for cities in less developed countries against cities in more developed countries.

Variable	All Cities	Less Developed Countries	More Developed Countries	p-value
Observations	200	148	52	
Proximity Index	-0.061 [-0.082, -0.040]	-0.072 [-0.098, -0.047]	-0.030 [-0.061, 0.002]	0.038 *
Cohesion Index	-0.066 [-0.086, -0.046]	-0.078 [-0.103, -0.053]	-0.033 [-0.064, -0.002]	0.029 *
Exchange Index	-0.054 [-0.086, -0.022]	-0.072 [-0.110, -0.035]	-0.004 [-0.065, 0.056]	0.056
Buildable Land Index	-0.031 [-0.059, -0.003]	-0.036 [-0.071, -0.001]	-0.017 [-0.056, 0.022]	0.47

Unfortunately, we do not have enough data on all of the possible determinants of the decline in the shape compactness of cities during the 1990-2014 period. Given the available data, we formulated three hypotheses. The first two posit that when urban incomes rise rapidly or when the cities expand quickly, urban planning cannot catch up with the rate of expansion and, as a result, cities become less compact. The third one posits that when cities merge together with settlements in their vicinity, they become less compact:

- The faster the rate of economic growth in the city, the faster the decline in its shape compactness.
- The faster the rate of expansion of the urban extent of a city, the faster the decline in its shape compactness.
- The greater the share of the added expansion area in 'inclusion', the faster the decline in its shape compactness.

We tested these hypotheses in a multiple regression model with the percent change in the Proximity Index during the 1990-2014 period as the dependent variable, using national GDP per capita change during this period as a proxy for the rate of economic growth in the city, the annual rate of growth of the urban extent during this period, and the share of 'inclusion' in the added area during this period. The model appears in **table 6** below. The coefficients of all three independent variables are significant at the 95 percent confidence level, but the sign for the first independent variable, the rate of economic growth, is reversed. The first of the three hypotheses listed above is thus not confirmed. The opposite is true. Other things being equal, cities in countries whose economies grew rapidly during the 1990-2014 period became more compact, not less compact, during this period. The model is robust, with an Adjusted R-squared of 0.236.

The results also pertain to the other two compactness indices, and to both the 1990-2000 and the 2000-2014 period.

Table 6: multiple linear regression model with percentage change of Proximity Compactness during 1990 to 2014 period as the dependent variable.

Independent Variable	Coefficient B	Confidence Interval	P-value
National GDP per capita percentage change	0.0065	[0.002, 0.011]	0.003
Urban Extent Growth	-0.0050	[-0.007, -0.003]	< .0001
Share of Inclusion in Added Area	-0.411	[-0.554, -0.269]	< .0001

To conclude this section, we note that the shape compactness of cities has declined significantly in recent years and that the rate of decline was significantly faster in less developed countries. In other publications (Angel 2012, 171-185) we have shown that a similar pattern prevails with regard to urban population densities. Those too have been in decline in recent years. More recent data from the *Atlas of Urban Expansion—2016 Edition* (Angel *et al* 2016) confirms that the average annual rate of decline of urban extent densities in less developed countries between 1990 and 2014, 2.0±0.4 percent, was significantly faster than the annual rate of decline in density in more developed countries during that period, 1.3±0.3 percent. The implications of these findings are highlighted in the following section.

3.3. How do compactness and density affect the average distance traveled in cities?

In a previous section, we defined the Proximity Index and the Cohesion Index as follows:

- The **Proximity Index** of a city is the ratio of the average distance from all points in the Equal Area Circle to its center and the average distance from all points of the city's urban extent to its Central Business District (CBD) identified by its City Hall.
- The **Cohesion Index** of a city is the ratio of the average distance from all points to all other points in the Equal Area Circle and the average distance from all points to all other points in the city's urban extent.

It can be ascertained that in a circular city of radius R, assuming that all jobs are concentrated in the Central Business District (CBD), located at the center of the circle, and that travel takes place at equal speed in all directions and at all locations, the average commuting distance will by $\frac{2}{3}R$ (The Math Forum, n.d.). Similarly, In a circular city of radius R, assuming that jobs are randomly distributed throughout the city, and that travel takes place at equal speed in all directions and at all locations, the average commuting distance will be $128R/45\pi$, or $0.9054R$ (Garcia-Pelayo, 2005, 2477). In both cases, commute distance will be proportional to R, the radius of the circle circumscribing the urban extent of the circular city in question.

In calculating the Proximity Index for a given city, we calculate the radius *R* of its Equal Area Circle and we calculate the average beeline distance from random points within its urban extent to its CBD. The Proximity Index is the ratio of the two. In calculating the Cohesion Index for a given city, we calculate the radius *R* of its Equal Area Circle and we calculate the average beeline distance between random points within its urban extent. The Cohesion Index is the ratio of the two. A Proximity Index of 0.25 thus means that the average distance to the CBD in the city is 4 times the average distance from a random point in its Equal Area Circle to its center. A Cohesion Index of 0.25 means that the average beeline distance between random points in the city is 4 times the average distance between all points in the Equal Area Circle. In a city with an urban extent of a given area, therefore, a doubling of the Proximity Index will amount to halving the average distance to its CBD. The same will be true in the case of the Cohesion Index: A doubling of the Cohesion Index in that city will amount to halving the average distance between random locations in the city.

A similar observation can be made about urban population density. Imagine a circular city of Radius *R* and a population *P*. Its average population density will be $P/\pi R^2$. Now imagine that its population remains the same and its density doubles. This would amount to shrinking its area to half its previous area. Correspondingly, its radius R' will shrink by a factor of $\sqrt{2}$, $(R' = R/\sqrt{2})$. And since the average distance to the CBD and the average distance between two random points in the city are proportional to the radius R, they too will shrink by a factor of $\sqrt{2}$.

We can thus see that both shape compactness and density have similar effects on average travel distances in cities. Other things being equal, the more compact the urban extent of a city, the shorter the travel distances within that city will be, and the denser the city, the shorter will travel distances within that city will be as well. We can indeed calculate the average beeline distance between random points for all the city in the sample for all time periods and use it as a proxy for actual travel distances within that city. We can use these values to model the effects of population, density, and shape compactness on travel distances in cities in the sample and in the universe of cities as a whole.

We constructed a multiple regression model with the natural logarithm of the beeline distance between random points in a city as a dependent variable and the logarithms of the city's population, its average population density, and its Exchange Index as independent variables. The model is presented in **table 7** below.

Table 7: multiple linear regression with the log of the average distance between points within the urban extent as the dependent variable. Adjusted R-squared = 0.996, N = 198.

Variable	Coefficient B	Confidence Interval	P-value
Log Population Size	0.502	[0.497, 0.507]	<.0001
Log Population Density	-0.504	[-0.513, -0.495]	<.0001
Log Exchange Index	-0.655	[-0.688, -0.622]	<.0001

The model is robust and explains almost all of the variation in the average travel distance in cities with the three independent variables employed in the model; and because the model is using logarithms, its coefficients can be interpreted as elasticities. The model predicts that, other things being equal, a 10 percent increase in city population is associated with a 5 percent increase in the average distance among point locations. Similarly, a 10 percent increase in the average population density is associated with a 5 percent decline in the average distance among point locations, and a 10 percent decrease in the Exchange Index is associated with a 6.5 percent decline in the average distance among point locations.

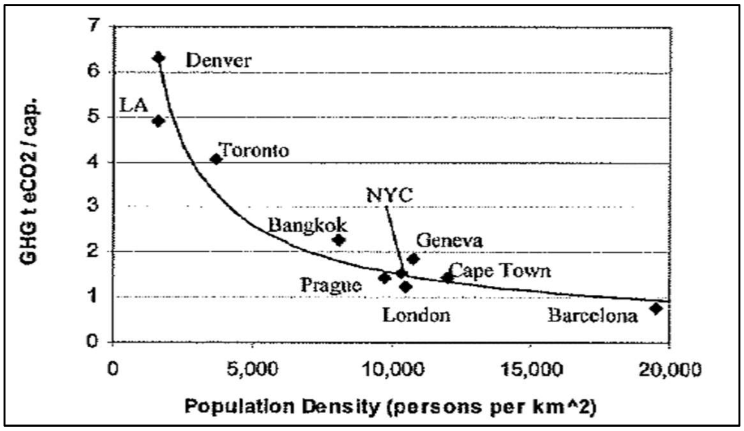
We can conclude, therefore, that both population density and compactness affect travel distances in cities. To the extent that the reduction of travel distances in cities could have a positive effect on reducing the energy spent in travel and hence on reducing greenhouse gas emissions, to the extent that they could also reduce commute time and thus improve labor market performance, and to the extent that they can reduce the overall length of infrastructure networks, we should seek to employ policies that increase the shape compactness of urban extents as well as those that increase the average densities of urban extents. We discuss the implications of this finding in the concluding section of this paper.

4. Conclusions and policy implications

The first conclusion and policy recommendation of this paper is that the shape compactness of urban extents must enter the discussion of the relationship between urban form and climate change, a discussion that until now has been dominated by a singular attention to urban density. Urban density, measured simply as the ratio of the total population of a city or metropolitan area and its urban extent, has emerged as the key attribute of urban form that drives climate change. According to the 5th Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC), “key urban form drivers of energy and GHG emissions are density, land use mix, connectivity, and accessibility” (Seto *et al* 2014, 927). Günerlap *et al* (2017, 8945) assert that “Systemic efforts that focus on…urban density…can improve the well-being of billions of urban residents and contribute to mitigating climate change by reducing energy use in urban areas”. In other words, quite apart from GHG emissions from energy generation in or near cities and quite apart from the inefficient use of energy in urban industry or in the urban building stock, the territorial organization of cities, in and of itself, drives energy use and GHG emissions in cities. Higher-density cities make for shorter trips and therefore have lower Vehicle Kilometers Traveled (VKT). They also make public transit more feasible, both leading to lower energy use and lower GHG emissions. The shape compactness of urban extents has not yet entered this discussion.

Newman and Kenworthy (1999) already noted the inverse relationship between urban density and energy use and, by implication, on GHG emissions. A recent article by Kennedy *et al* (2009) reproduces and confirms their results for energy use from ground transportation fuels for ten global cities (see **figure 19**). Although the graph does not account for differences in per capita income between the cities, the relationship of GHG emissions to density is unmistakable.

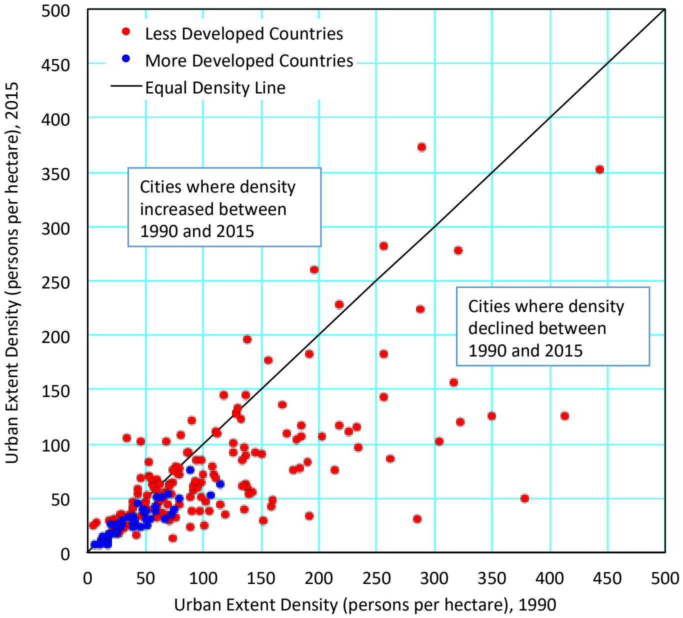
Figure 19: GHG emissions from ground transportation fuels are inversely related to population density (reproduced from Kennedy et al (2009, 7299).



It stands to reason, therefore, that if cities are to contribute to global efforts to mitigate greenhouse gas emissions, they must make serious efforts to increase their urban densities as well as to increase their shape compactness. Density increases in cities over time can occur in one of two ways: First, by densifying their existing footprints and second, by building at higher densities in their expansion areas. The monocentric city model (e.g. Alonso 1964) postulates that densities, hand-in-hand with land prices, decline exponentially with distance from city centers and this observation has been shown to hold in the great majority of cities with functional land markets. Moscow and Johannesburg, as reported by Bertaud and Renaud (1996), may have been singular exceptions. Since the expansion areas of cities are by definition on their periphery, land prices there are typically lower than land prices in their existing urban footprints. Expecting urban peripheries at large to be built at higher average densities than existing urban footprints thus makes little economic sense.

We must conclude, therefore, that if we are to increase urban densities over time, we must *densify* existing urban footprints. Not surprisingly, all the studies reporting on the relationship between density and climate change are cross-sectional studies, focusing on comparing cities with different densities to each other, but telling us little or nothing about *how* to increase urban densities over time. This information is of particular significance in the light of our previous findings (e.g. Angel 2012, 170-185) regarding the persistent and statistically significant *decline* in urban densities over time (**figure 20**), a decline associated with increasing urban incomes and the increased availability of affordable urban transportation. In this paper we have also reported on the decline in the compactness of urban extents. Both trends do not bode well for actions aimed at combating greenhouse gas emissions through urban form.

Figure 20: In the global sample of 200 cities, average urban extent densities declined in 72 percent of cities in less developed countries and in 75 percent of the cities in more developed countries between 1990 and 2015.



Source: *Atlas of Urban Expansion—2016 Edition*, online at www.atlasofurbanexpansion.org, Table 1: Areas and Densities.

Urban densities are oftentimes the outcome of supply and demand pressures for residential living space. They may also be the outcome of consumer preferences for larger homes further away or smaller homes closer to urban centers. Still, there is an ongoing policy debate on the merits of accommodating urban population growth through urban densification as against through urban expansion. Those engaged in this debate claim that unconstrained markets fail to account for air pollution and GHG emissions and thus create lower than expected densities, and that public intervention is necessary to ensure that cities grow at higher densities in a productive, inclusive, and sustainable manner.

Densification, it must be emphasized here, is not an unmitigated good. It is typically the preferred course of action for those concerned with energy conservation, with the mitigation of global GHG emissions, particularly from urban transport, and with excessive public infrastructure costs. Densification is typically resisted by existing communities that prefer the status quo and by established planning regulations that limit what can be built where. Community resistance to densification, or the inability to reform planning regulations that prohibit it, may limit densification and accelerate expansion.

Expansion—and, preferably, orderly expansion—is typically the preferred course for those concerned with overcrowding or with land supply bottlenecks that may lead to unaffordable housing. Urban expansion is typically resisted by homeowners who want to protect their property values and by citizens who want to protect green spaces on the urban periphery. Resistance to urban expansion may compromise preparing for it at the proper scale, failing to put in place adequate public works and to protect public open spaces and areas of high environmental risk in advance of development.

Our position in this policy debate takes the middle road, promoting both acceptable densification and orderly urban expansion that renders cities more compact and seeking a proper balance between the two. Neither acceptable densification nor acceptable expansion, we note, is easy or simple to implement. Both require strong leadership and, more often than not, regulatory reform. And both are indeed substitutes: Preparing for more urban expansion than expected can be seen as a resilience strategy, substituting expansion for densification in case cities fail to densify at the rate expected in their plans.

Orderly expansion can go hand in hand with densification and can seek to make urban extents more compact over time, thus contributing to efforts to reduce greenhouse gas emissions and to mitigate their effects on climate change. We note here that the policy implications reported below paraphrase the conclusions of "Chapter 14: The Pulsating Compactness of Urban Footprints" in Angel's *Planet of Cities* (2012, 223-247). The reader is referred to this chapter for a detailed discussion of the compactness of urban extents, as well as data on the change in compactness of a representative group of 30 cities over a 200-year time period.

When making preparations for expansion in any particular city in the coming decades, we must therefore seek a deeper understanding of the forces making its urban extent more or less compact and to come to terms with their real potential to subvert our best intensions. This is particularly crucial if planning for expansion by public authorities aims to *guide* it into particular lands while seeking to prevent the conversion of other lands to urban use. There is a natural and perfectly understandable desire on the part of public officials drawing up plans for urban expansion to guide built-up areas away from open spaces that need special protection, for example (a) lands that are needed to ensure access to public open space within a reasonable distance from built-up areas; (b) lands with steep slopes that should be left unoccupied because of the danger of landslides; (c) wetlands containing sensitive fauna and flora that should be left undisturbed; (d) watersheds that feed into reservoirs supplying drinking water to the city; or (e) farmlands on rich soils that need to be preserved to protect food supplies. Most, if not all, of these considerations act to make urban extents less compact, decreasing overall access in the urban area, while increasing the length of infrastructure lines. Building on steeper slopes, for example, can increase the shape compactness of urban extents. Cities facing serious physical constraints thus face an important choice: Extending further out and becoming less compact in the process, or building on steeper slopes closer to the city center.

Such tradeoffs need to be properly considered when making plans for urban expansion, of course. What is more, we should remain fully aware of the possibility that the forces acting to negate and compromise such lofty plans—those forces that seek to make the city more compact by locating as close as possible to job opportunities, for example, or those that seek ocean views, to take another example—may end up having the upper hand.

To conclude, plans for guiding urban expansion cannot and should not be based on wishful thinking. Instead, they should be based on a full recognition of the forces seeking to make city extents more compact, namely the desires of households and businesses to be as close as possible to the city center and to each other, forces that often trump their desire to have access to open space. It should not come as a surprise that the pursuit of urban locations with easy access to jobs, markets and other people fulfills a more basic need in the hierarchy of needs than access to open space. It is a legitimate preference of many families—especially low-income ones—that should therefore be given its due weight in the planning calculus. More generally, open spaces are difficult to protect when households' and firms' preferences result in strong political and economic pressures to occupy them. We must keep in mind that the economic and political costs of effectively protecting open spaces are limited and must therefore be marshaled judiciously. Trying to protect too much open space with too few resources may result in failure to protect any open space at all. As it says in the Talmud: "If you have seized a lot, you have not seized; if you have seized a little, you have seized."

The effects of radial intercity lines that allow for higher travel speeds on urban expansion ─ be they commuter rail lines or freeways ─ should also be taken into account when seeking to guide urban expansion, as these tend to make city extent less compact. Guiding urban development into the interstices between the tentacles of urban development along these lines, so as to makes cities more compact, requires the planning and construction of a dense network of higher-speed arterial roads in these areas, roads that can carry public as well as private transport, that allow for lateral movement of traffic, and can help equalize travel times along alternate routes so as to compromise the advantage of radial travel on intercity lines. Simply marking these areas on land use plans as available for urban use may not be sufficient to direct development there. Planting trees along the future sidewalks of an arterial road grid lay out in the areas of projected urban expansion in coming decades—as currently practiced in Colombian cities (Vasconez *et al* 2015)—may be a more realistic alternative. Guiding urban expansion in a realistic fashion cannot take place in a vacuum. It must be planned and executed in full recognition of the complex interplay of forces now acting to make cities more compact or less compact.

5. References

Alonso, W. (1964). *Location and Land Use* (Cambridge MA: Harvard University Press).

Angel et al, 2016. *Atlas of Urban Expansion—2016 Edition*, online at: www.atlasofurbanexpansion.org.

Angel, S. 2012. *Planet of Cities*, Cambridge MA: Lincoln Institute of Land Policy, 170-185.

Angel, S. and Blei, A. M., 2016a. The productivity of American cities: How densification, relocation, and greater mobility sustain the productive advantage of larger U.S. metropolitan labor markets, *Cities* 51, January, 36-51.

Angel, S. and Blei, A.M., 2016b. The spatial structure of American cities: The great majority of workplaces are no longer in CBDs, employment sub-centers, or live-work communities, *Cities* 51, January, 21-35.

Angel, S. Valdivia I.M., and Lutzy R., 2007. *Urban Expansion, Land Conversion, Affordable Housing, and Energy Use in China: The Case of Zhengzhou, Henan.* Princeton University: The China Policy Workshop Paper. Online at http://wws.princeton.edu/research/pwreports_f07/wws591a.pdf.

Angel, S., J. Parent, and D. L. Civco. 2010c. Ten compactness properties of circles: Measuring shape in geography. *The Canadian Geographer* 54(4): 441–461.

Bertaud, Alain and Bertrand Renaud, 1995. "Cities without Land Markets: Location and Land Use in the Socialist City," *Policy Research Working Paper 1477*, Washington D.C.: The World Bank, June.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner. 2006. "Causes of sprawl: A portrait from space." *Quarterly Journal of Economics* 121(2): 587–633.

Galarza Sánchez, N., Liu, Y., Angel, S. and Thom, K., 2018. The 2010 Universe of Cities: A New Perspective on Global Urbanization (Working Paper), Cambridge MA: Lincoln Institute of Land Policy, March.

García-Pelayo, R., 2005. Distribution of distance in the spheroid, *Journal of Physics A: Mathematical and General* 38 3475-2482, online at: http://iopscience.iop.org/article/10.1088/0305-4470/38/16/001/pdf

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P.L. Fragkias, M., Li, X., and Seto, K.C., 2017. "Global scenarios of urban density and its impacts on building energy use through 2050", in *Proceedings of the National Academy of Science (PNAS)* 114(34) August 22, 8945–8950, online at: www.pnas.org/cgi/doi/10.1073/pnas.1606035114.

Harari, M., 2016. Cities in Bad Shape: Urban Geometry in India (working Paper), The Wharton School, University of Pennsylvania, February.

Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., Havránek, M., Pataki, D., Phdungsilp, A., Ramaswami, A., and Mendez, G. V., 2009. Greenhouse Gas Emissions from Global Cities, *Environmental Science and Technology* 43(19), 7297–7302.

Newman, Peter and Kenworthy, Jeffrey, 1999, *Sustainability and Cities: Overcoming Automobile Dependence*, Washington D.C.: Island Press.

OECD, 2015. OECD Urban Policy Reviews: Mexico 2015 Transforming Urban Policy and Housing Finance, OECD Publishing. Http://dx/doi/org/10.1787/9789264227293-en.

Ritter, C. 1822 *Die Erdkunde im Verhaltniss zur Natur und Geschichte des Menschen, oder allgemeine vergleichende Geographie*, 2nd edition, Part I, Book 1, Berlin, quoted in Y. S. Frolov 'Measuring of Shape of Geographical Phenomena: a History of the Issue' *Soviet Geography: Review and Translation* 16, 1975, 676–687

Saiz, A., 2010. "The Geographic Determinants of Housing Supply", Quarterly Journal of Economics, 125 (3), 1253-1296.

Seto K. C., S. Dhakal, A. Bigio, H. Blanco, G. C. Delgado, D. Dewar, L. Huang, A. Inaba, A. Kansal, S. Lwasa, J. E. McMahon, D. B. Müller, J. Murakami, H. Nagendra, and A. Ramaswami, 2014: "Chapter 12: Human Settlements, Infrastructure and Spatial Planning", in *Climate Change 2014: Mitigation of Climate Change, Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* online at: https://www.ipcc.ch/pdf/assessment-report/ar5/wg3/ipcc_wg3_ar5_chapter12.pdf.

The Math Forum, n.d. Average Radial Distance of Points within a Circle, online at: http://mathforum.org/library/drmath/view/62529.html

U.K. Department of Communities and Local Government, 2017. Local Planning Authority Green Belt: England 2016/17, Statistical Release, 7 September, online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/642684/Green_Belt_Statistics_England_2016-17.pdf

Vasconez, J., Galarza, N. and Angel, S., 2015. Preparando ciudades de rápido crecimiento para su expansión: Informe sobre Valledupar y Montería, Colombia. *Medio Ambiente y*