Progress in Planning xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Progress in Planning

journal homepage: www.elsevier.com/locate/progress

The shape compactness of urban footprints

Shlomo Angel*, Sara Arango Franco, Yang Liu, Alejandro M. Blei

The Marron Institute of Urban Management, New York University, USA

ARTICLE INFO

Keywords:
Compact city
Urban footprints
Shape compactness
Urban expansion
Sample of cities
Climate change
Population density

ABSTRACT

Urban population density has featured in a large body of literature on the Compact City paradigm as the key compactness attribute of cities, yet the shape compactness of urban footprints has hardly deserved a mention. This essay seeks to correct that. We review the literature on the Compact City Paradigm with a special focus on the relationship between urban form and climate change, and focus on twelve physical attributes of cities that make them more or less compact. Other things being equal, both population density and shape compactness help determine the average travel distances in cities, and hence affect their energy consumption and their greenhouse gas emissions. They also affect the length of infrastructure lines and the length of commutes. In principle, therefore, increasing either the shape compactness or the population density of cities can contribute—in different yet similar measure—to mitigating climate change. There are strong forces that push urban footprints to become more compact—that is, circular or near circular in shape—and these forces have evolved over time. There are also powerful forces that have pushed urban footprints to become less compact over time. We introduce these forces and illustrate their effects on particular cities. We then focus on a small set of metrics for measuring the shape compactness of cities. We use them to measure urban footprints 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 urban footprints 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 better able to mitigate climate change in decades to come.

> Of all the attributes that characterize a city, there can be little doubt that proximity is the most crucial.
> Fanis Grammemos, 2011
> The circle is the most compact of shapes because the proximity of all points to all other points within it is at a maximum.
> A yet to be proven mathematical conjecture

1. Introduction

1.1. The conceptual framework

The central objective of this essay is to broaden and deepen our understanding of *the compact city paradigm* by introducing readers to a number of new compactness attributes of cities that have not been extensively discussed in the literature before, essentially those that have to do with the geographical shape of urban footprints, rather than with their densities or with their internal spatial structure. We begin this section by reviewing the emerging interest in compact cities in the last 25 years and then focus on an integrated conceptual framework for understanding, studying, and acting upon the various compactness attributes of urban forms. More specifically we introduce, define, and give real-world examples of twelve compactness attributes of cities and discuss the relationships between them and later measure the correlations between most of them, using data from a global sample of 200 cities. In this manner, we aim to provide the reader with a novel and rigorous understanding of a subset of these attributes, the compactness attributes of urban footprints, and of what can be done and needs to be done to make cities more compact—and in many instances possibly more productive, more inclusive, and more sustainable as well—by making their urban footprints more compact.

Since the Earth Summit of 1992 (United Nations, 1993), there have been worldwide efforts to address environmental challenges—be they the depletion of natural resources, the loss of cultivable lands, air and water pollution, or greenhouse gas emissions—by changes in urban form. Proposed changes have ranged from the design of energy-efficient

* Corresponding author at: The Marron Institute of Urban Management, New York University, 60 5th Avenue, 2nd Floor, New York, NY, 10011, USA.
E-mail addresses: sangel@stern.nyu.edu (S. Angel), saf537@nyu.edu (S. Arango Franco), yl3371@nyu.edu (Y. Liu), ablei@stern.nyu.edu (A.M. Blei).

https://doi.org/10.1016/j.progress.2018.12.001
Received 16 July 2018; Received in revised form 30 November 2018; Accepted 2 December 2018
0305-9006/

Please cite this article as: Angel, S., Progress in Planning, https://doi.org/10.1016/j.progress.2018.12.001

buildings to making room for bicycle lanes, but the great majority of them have focused on making cities more compact (see, e.g. Seto et al., 2014). The various claims that the more compact a city is, the better it is—not only in environmental terms, but in economic and social terms as well—make up the *compact city paradigm*, a paradigm or worldview that is now "enshrined in land use planning policy in many countries" (Burton, 2001, 219). Indeed, most developed countries now pursue policies that implicitly or explicitly promote compact urban form (OECD, 2012).

The compactness of cities has been defined and measured in a number of different ways, as we shall see below, and since the advent of the compact city paradigm, there has been a steady flow of findings—some more robust than others—showing, for example, that more compact cities are more productive and more innovative; that they deliver public services at a lower cost; that they enhance social and economic mobility as well as diversity; that they increase public transit use; that they reduce energy use and greenhouse gas emissions; and that they improve health and well-being (for a comprehensive review, see Boyko & Cooper, 2011). In a recent review of evidence from 300 studies, Ahlfeldt and Pietrostefani show that the measurable effects of compact urban form have tended to become more positive over time, possibly driven by the application of more rigorous research techniques (2018, 19). That said, the compact city paradigm has had a number of detractors as well (e.g. Melia, Parkhurst, & Barton, 2011; Neuman, 2005). Compact city policies that seek to contain urban expansion—such as greenbelts or urban growth boundaries—have been criticized for their adverse effects on housing affordability through the creation of residential land supply bottlenecks (Cheshire & Hilber, 2008). In cities of less-developed countries, compactness has been associated with serious overcrowding and increased compactness has been suspect of exacerbating traffic congestion, water supply, sewerage and drainage shortages, fire risk, and emergency vehicle access. It may also have led to an increase in a range of health risks (Ezeh et al., 2016; Mitlin & Satterthwaite, 2013).

The growing body of literature on the compact city and the emerging global consensus regarding the benefits of compact cities for mitigating climate change require a rigorous and consistent approach to the definition and measurement of the compactness attributes of cities. The literature on the subject uses a wide variety of disparate metrics to measure compactness, sometimes focusing on one aspect of compactness and sometimes on another. Indeed, it stands to reason that a city can be more compact or less compact in a number of distinct ways that are not necessarily correlated with each other, and that acting to make cities more compact requires concerted action on one or more compactness attributes. In the next section we review the key compactness attributes identified in the literature that have a direct bearing on climate change and introduce a number of new attributes that focus attention on the shape compactness of urban footprints, the central theme of this essay. In the following section we discuss the expected relationships among the various compactness attributes of cities and argue that some attributes can be expected to be independent of others, in line with our claim that compact city policies must be conceived as coordinated strategies involving pragmatic efforts to modify several compactness attributes of cities at the same time so as to render cities more compact.

While it has become quite clear that the costs and benefits of rendering cities more compact require a rather complex calculus involving complicated tradeoffs—tradeoffs that may vary considerably from one city to another—we believe that the growing concern with climate change may change and simplify this calculus. The urgent need to reduce greenhouse gas emissions—and, more specifically, the need to reduce greenhouse gas emissions from vehicles—calls for significant reductions in vehicle-kilometers-traveled in cities in the coming years, reductions that will be critical to meeting global greenhouse reduction targets in time. In terms of orders of magnitude, "estimates from a review of published studies of U. S. cities forecasted a 5% to 12% VKT reduction from doubling residential densities and as high as 25% reductions when combined with other strategies, including road pricing" (National Research Council, 2009, quoted in Seto et al., 2014, 948).

The compact city paradigm theorizes that the reduction of vehicle-kilometers-traveled can be done through several logical pathways. We have thus far identified eight such pathways. Three of them pertain to the overall density of cities, no doubt the key building block of the compact city paradigm: (1) higher urban population densities, of both trip origins and destinations, will bring more of them within walking or biking distance of each other, thus reducing the number of vehicle trips made, and possibly the number of vehicles owned as well (Næss, 2005; Saelens, Sallis, & Frank, 2003; Zhou & Kockelman, 2008); (2) higher densities—of both trip origins and destinations—will make public transit operations more viable, increasing the share of trips made by public transport and thus reducing the number of vehicle trips made and possibly the number of vehicles owned as well (Bunting, Filion, & Priston, 2002; Forsyth, Oakes, Schmitz, & Hearst, 2007; Holtzclaw, Clear, Dittmar, Goldstein, & Haas, 2002; Saelens et al., 2003); and (3) higher densities will require smaller urban footprints to accommodate a given population, thus shortening the average travel distance between trip origins and destinations, leading to reductions in the total vehicle-kilometers-traveled (Brownstone & Golob, 2009; Cervero & Kockelman, 1997; Ewing & Cervero, 2001; Frank & Pivo, 1994).

Urban spatial structure can affect the reduction of total vehicle-kilometers-traveled through two additional pathways: (4) for a given overall population density, the greater *concentrations* of populations and jobs in city centers and sub-centers—either in and around the Central Business Districts (CBDs) or in higher-density sub-centers, commonly referred to as Transit-Oriented Districts (TODs)—can increase the share of trips made by public transport and thus reduce the number of vehicle trips made and possibly the number of vehicles owned as well (Bento, Cropper, Mobarak, & Vinha, 2003; Cervero, 1998); and (5) for a given overall population density, the greater the share of people who live and work in the same community—facilitated through the decentralization of jobs and through mixed-use or small grain zoning, and measured, in part, by a job decentralization index or a jobs-housing balance—the more trips will be taken by walking and biking, thus reducing the number of vehicle trips made and possibly the number of vehicles owned as well (Ewing & Cervero, 2010; Kockelman, 1997; Mogridge, 1985; Permana, Perera, & Kumar, 2008).

The transportation network itself—the availability and density of the arterial road and rail network, the frequency of bus and rail service, and the connectivity of the street network—can reduce total vehicle-kilometers-traveled traveled through two logical pathways: (6) the higher the density of the rail and bus network, and the higher the frequency of service, the more competitive will public transit be with private automobiles in terms of travel time, increasing the share of trips made by public transport and thus reducing the number of vehicle trips made and possibly the number of vehicles owned as well (Bento et al., 2003; Hidalgo & Gutiérrez, 2013); and (7) the higher the connectivity of streets—measured, for example, by average block size or by the share of four-way intersections—the more walkable the city, bringing more trip origins and destinations within walking or biking distance of each other, thus reducing the number of vehicle trips made and possibly the number of vehicles owned as well (Gehl, 2010; Salon, Boarnet, Handy, Spears, & Tal, 2012). That said, we should keep in mind that the greater the share of the land in streets, the more energy is expended—and the more GHG emissions created—in paving them (Horvath, 2004; Müller et al., 2013). Moreover, other things being equal, the greater the share of the land in streets, the greater the vehicle-kilometers traveled (Duranton & Turner, 2011; Noland, 2001).

Finally, the shape of urban footprints—although rarely mentioned in the literature on the compact city—can affect a reduction in vehicle-kilometers traveled. Just as higher densities require smaller urban footprints to accommodate a given population, urban footprints that are more circular in shape, rather than elongated or tentacle-like, have the same effect: (8) Other things being equal, more compact urban footprints shorten the average travel distance between trip origins and destinations, thus leading to reductions in the total vehicle-kilometers-traveled and hence to reductions in energy use and greenhouse gas emissions (Bento et al., 2003).

While it remains to be shown that the potential contribution of increasing the shape compactness of urban footprints with the aim of reducing average travel distances in cities—and thus contributing to the reduction of greenhouse gas emissions—can indeed be substantial, and that it, therefore, merits efforts at implementing land use and transportation policies that can help make it happen, it is this contention that motivates this essay.

1.2. Twelve compactness attributes of cities

Twelve compactness attributes of cities are introduced below. All twelve attributes pertain to aspects of urban form, and they all seek to focus on aspects of urban form that, when attaining higher values, contribute to attaining a desirable sustainability objective: maintaining or intensifying contact and connectivity within the city in one way or another, so as to reduce the energy expended in travel between urban locations, with the consequent reduction in greenhouse gas emissions. Spatial attributes of the city that, when attaining higher values, do not contribute to this sustainability objective are excluded from this analysis. From a policy perspective, it is important to keep in mind that to the extent that these individual attributes of urban form are not correlated with each other, each one can make an independent contribution toward meeting this objective. The possible correlations among these variables are discussed in the following section.

In parallel with the logical pathways linking the spatial structure of cities to the mitigation of climate change, the compactness attributes of cities can be conveniently divided into four groups: (1) the density attributes of cities; (2) the compactness attributes of their internal spatial structure; (3) the compactness attributes of their transportation networks; and (4) the shape compactness of their urban footprints.

The first group of the compactness attributes of cities contains three attributes that are associated with aspects of the overall density of a given urban extent:

1 **Density** is the population density of a city, measured as *the ratio of its population and its area (or urban extent)*. For a given population, the lower the population density of the city, the larger its urban extent. The larger its urban extent, the larger the average distance between locations in the city. The larger that average distance, the longer people have to travel and thus the more vehicle-kilometrs-traveled, the longer infrastructure lines have to be, the more energy is consumed in travel, and the more greenhouse gases are released into the atmosphere. In parallel, the larger the urban extent, the more of the urban periphery is disturbed, the more agricultural land on the urban periphery is lost, and the less vegetation is available to convert carbon dioxide to oxygen. Density is the most often used metric in the literature to characterize the compactness of cities. It is related, among others, to productivity (Neuman, 2005), innovation (Jones, Leishman, MacDonald, & Watkins, 2010), job accessibility (Beer, 1994), access to amenities (Churchman, 1999), efficiency of public service delivery (Matsumoto, 2011), pollution reduction (Bechle, Millet, & Marshall, 2011), health (Matsumoto, 2011), and wellbeing (Wilson & Baldassare, 1996). Several authors have attempted to link density to the reduction of greenhouse gas emissions, through the three logical pathways introduced in the previous section (Brownstone & Golob, 2009; Cervero & Kockelman, 1997; Ewing & Cervero, 2001; Frank & Pivo, 1994).

2 **Saturation** is the degree to which the urban extent of the city is saturated by its built-up area, measured as the *ratio of the built-up area of the city and its overall urban extent*. The urban extent of a city typically includes urbanized open spaces within it, be they public or private open spaces or vacant lands awaiting development. Other things being equal, the more saturated the urban extent with built-up areas, the smaller the urban extent and the more compact the city is. In effect, saturation is a factor of density: The higher the level of saturation of a city, the higher its density. Higher levels of saturation thus decrease the distance between locations in the city, with a concomitant decrease in travel distances, in vehicle-kilometers traveled, in transport-related energy expenditures and in greenhouse gas emissions, in the length of infrastructure lines, and in disturbance of the countryside. Burchfield, Overman, Puga, and Turner (2006) for example, use a saturation index highly correlated with the one defined above—the percentage of undeveloped land in a square kilometer surrounding an average residential development—as a sprawl index.

3 **Mass** is the density of residential floor space in the city measured as *the ratio of its total residential floor area and its entire urban extent.* Mass measures the density of the built environment, in contrast with population density. That said, other things being equal—especially the amount of floor area per person in the city—cities with higher mass will also have higher population density, and thus a smaller urban extent. In effect, mass, like saturation, is also a factor of density. The greater the mass of a city, the shorter the distance between locations in the city, with concomitant decrease in travel distances, vehicle-kilometers-traveled, transport-related energy expenditures and greenhouse gas emissions, in the length of infrastructure lines, and in disturbance of the countryside. The density of the built environment is also often used in the literature to characterize the compact city. For some authors (e.g. Boyko & Cooper, 2011), Mass measured in dwelling units per unit area, rather than residential floor area per unit area, is the key measure of density. Mass is associated in the compact city literature with access to amenities (Bonfantini, 2013), pollution reduction (Churchman, 1999), travel mode choice (Thomas & Cousins, 1996), and wellbeing (Burton, 2000).

The second group of the compactness attributes of cities contains two attributes that are associated with the spatial structure of cities. The first attribute in this group is associated with the spatial mix of land uses within urban neighborhoods, an attribute that is not, *a priori*, associated with population density of cities. The second focuses on the distribution of densities within the urban extent of cities, rather than with their overall density:

4 **Mix** is the extent to which residential land use is in proximity to other land uses. It is measured, for example, by *the job-housing mix of urban neighborhoods,* i.e. *the extent to which there is a balance between the number of workplaces and the number of workers in a given community.* Other things being equal, mixed land uses or small-grain land use clusters shorten the distance between residences and desired destinations, such as shops, schools, amenities, or jobs. Hence land use mix has the same effect as density, for example, on reducing travel distances, with the concomitant decline in their negative effects. Mixed land use is also commonly used in the literature as an attribute of the compact city. It is associated, among others with access to services and amenities (Churchman, 1999), pollution reduction (World Health Organization (WHO), 2011), energy efficiency (OECD, 2012), and well-being (Vorontsova, Vorontsova, & Salimgareev, 2016). The decentralization of jobs is also found to be associated with positive environmental outcomes (Gaigné, Riou, & Thisse, 2011). Bento et al. (2003), for example, find a positive relationship between the job-housing mix and the share of walking trips in U.S. cities.

5 **Concentration** is *the share of residences and workplaces that are located in city centers and sub-centers—either in the Central Business Districts (CBDs) or in higher-density sub-centers, commonly referred to as Transit-Oriented Districts (TODs)—rather than dispersed equally*

throughout the urban extent. Other things being equal, concentration increases the potential of serving a substantial share of trips by a public transit network that connects these higher-density centers, while a larger share of travel within them can take place through walking and biking. Bento et al. (2003), for example, found a significant negative relationship between the concentration of jobs in CBDs and the number of private vehicles owned in U.S. cities.

The third group of attributes of the spatial structure of cities is associated with properties of their street, bus, and rail networks. Where possible, these attributes measure the density of rail lines or rail stations, the density of bus lines or bus stations, and the frequency of service as well. The two attributes introduced below focus on the street networks and the extent to which they facilitate travel by creating physical spaces that allow for efficient movement of public transport, on the one hand, and walking and biking on the other:

6 **Walkability** is the extent to which the street network shortens walking distances between locations in the city. It is measured, for example by (1) *the average size of blocks;* or by (2) *the density of four-way intersections in the city*. Small city blocks, formed by streets that connect with four-way intersections allow for shorter trips between locations. Large city blocks or large areas served by streets with three-way intersections require long detours, making walking or biking more difficult. This, in turn, increases actual distances between urban locations, with the concomitant increase in their negative effects. Baruah, Henderson, and Peng (2017) find a positive association between walkability and population density in African cities. Gehl (2010) and Salon et al. (2012) find associations between walkability measures of the street network and the share of walking trips in cities.

7 **Connectivity** is the extent to which the entire urban area is serviced by an arterial inter-city road network. A dense and well-managed arterial road network facilitates the movement of buses, creates opportunities for introducing bus rapid transit, and thus improves the competitiveness of public transport vis-à-vis private vehicles. It is measured, for example, by *the average density of arterial roads in the city*. Cities that may be compact in both shape and density may still require longer commutes, for example, if there is a paucity of arterial roads allowing for efficient intra-city travel to work. In other words, places that may be close by geographically may be quite far away, measured along a sparse arterial road network. When the arterial road network is sparse, distances among urban locations—as well as the time of travel between them—increase with the concomitant increase in their negative effects. To our knowledge, connectivity has not yet been used in the literature as a measure of the compactness of cities. That said, Bento et al. (2003) found a significant reduction in car ownership in U.S. cities associated with the presence of rail and bus transit networks.

The fourth and final group of the compactness attributes of cities contains five attributes that are associated with the geometric shapes of the two-dimensional geographic footprints of cities. Bento et al. (2003), for example, measure the extent to which a given urban footprint is elongated, rather than circular, and give it a City Shape metric associated with its elongation. They report that in a sample of 114 U.S. cities in 1990, City Shape had a small yet significant effect on the number of vehicles per household and on vehicle-miles-traveled per vehicle. Our research reported here expands on the original insight of Bento et al. Indeed, defining and measuring the compactness attributes of the shape of urban footprints in a global sample of 200 cities will be our central focus in this essay. And while we fall short of Bento et al.'s accomplishment in that we do not have the data to associate these attributes with actual vehicle ownership or actual vehicles-kilometers-traveled in the cities studied, we are able to show in this essay that in a sample of 38 U.S. cities, both density and the compactness of urban footprints have significant and comparable effects on reducing average travel distances and thus on lowering vehicle-kilometers-traveled.

We can identify the physical footprint of a city—its urban extent—by examining remotely- sensed imagery with a view to determining its outer edges of the city, 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, therefore, that the study of shape compactness in geography is almost two centuries old (Ritter, 1822).

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, & 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 choose the appropriate properties for studying and measuring their compactness. 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.

Five compactness attributes of cities that are associated with the shape of their urban footprints—and that can be expected to have an impact on travel distances in cities and therefore on vehicle-kilometers traveled, energy consumption, and greenhouse gas emissions—are described below:

8 **Contiguity** is the extent to which the footprint of the city is not fragmented into a large number of freestanding urban clusters, measured as *the share of the urban extent of the city in the largest contiguous urban cluster*. Other things being equal, the more contiguous an urban area is, the shorter the average distance between locations within it, and the less area on the urban periphery is disturbed by urban intrusion into it. As with density, the shorter that average distance, the shorter people have to travel, the shorter infrastructure lines have to be, the less the total vehicle-kilometers-traveled, the less energy is consumed in travel, and the less greenhouse gases are released into the atmosphere. In parallel, the more contiguous the urban footprint is, the less fragmented is the adjacent urban periphery, and the less its agriculture or its wildlife are disturbed. McGarigal and Marks (1994), for example, define Contiguity as the Largest Patch Index, a key measure for studying landscape ecology. To our knowledge, a number of variations on Contiguity have been used in the compact city literature to measure leapfrog development (e.g. Burchfield et al., 2006), but the relationship between Contiguity and vehicle-kilometers-traveled and thus greenhouse gas emissions has not been studied.

9 **Proximity** focuses on the degree to which urban footprints facilitate access to the Central Business District. It is measured by *the ratio of the average distance to the center in the Equal Area Circle (a circle with the same area that of the city) and the average distance between locations in the city and its Central Business District*. Proximity is also used by Harari (2017) as a measure of urban compactness. Clearly, in monocentric or near-monocentric cities, increased Proximity decreases commuting distances to the CBD with the concomitant positive effects on energy savings and reduced travel-related greenhouse gas emissions.

10 **Cohesion** is the extent to which locations in the city are close to one another. It is measured as *the ratio of the average distance between any two locations in the Equal Area Circle (a circle with the same area as*

that of the city) and the average distance between any two locations in the city. In a circle, the average distance between points is at a minimum. To the extent that an urban footprint resembles a circle, average distances between locations in the city will be shorter than in, say, urban footprints that are long and thin. Other things being equal, increased cohesion compactness thus acts much like increased density to reduce the average distance between locations in the city. As with density, the shorter that average distance, the shorter people have to travel, the shorter infrastructure lines have to be, the less energy is consumed in travel, and the less greenhouse gases are released into the atmosphere. Harari (2017), for example, investigates the effects on more compact cities on consumer welfare and firm productivity in Indian cities, using Cohesion as a measure of compactness.

11 **Exchange** focuses on the extent to which the urban footprint fills a circle of the same area centered at its centroid. It is measured by *the share of the urban footprint within the Equal Area Circle (a circle of the same area as that of the city) centered at its centroid*. Exchange has been used by Angel and Parent (2011) to measure the gerrymandering of U.S. election districts, but to our knowledge has not been used in the compact city literature. Higher levels of Exchange compactness go hand in hand with higher levels of Cohesion and Proximity: They reduce travel distances and increase accessibility within urban footprints with the concomitant positive effects on their sustainability.

12 **Fullness** is the extent to which the urban footprint of the city fills the available buildable land (dry land with a slope of less than 15°) in and around it. It is measured—in a similar way Exchange is measured—as *the share of the urban footprint in the Buildable Land Circle (a circle containing buildable land equivalent to the area of the city) centered at its centroid*. The urban footprint of a city can only be as compact as the physical features of the natural landscape in which it is located allow it to be. Cities are typically built on dry land of modest slope. Cities surrounded by steep mountain slopes or bodies of water cannot attain the levels of cohesion compactness, for example, as cities built on alluvial plains. Fullness measures the extent to which an urban footprint is as compact as can be given the constraints imposed on it by the natural landscape. Other things being equal, the higher the Fullness of an urban footprint, the more cohesive it will be, thus reducing the average distances between locations in the city with its concomitant positive environmental effects. Saiz (2010), for example, investigates the impact of landscape constraints on the elasticity of housing supply in U.S. cities, using a calculation that has a similar function to that of the fullness index.

1.3. Expected correlations among compactness attributes

To formulate an effective compact city strategy that can reduce travel distances in cities, we can start by assuming that an increase in compactness of one or more of these twelve attributes will make the city as a whole more compact and thus better able mitigate climate change. An effective strategy would seek to increase the compactness of the city in the most efficient and equitable way, where the benefit of an intervention to change any attribute outweighs the cost associated with that change. This suggests, first of all, that we need to better understand what is involved in increasing the compactness of cities along each one of these dimensions: What can be done, at what cost—be it political, social, or economic—to increase the compactness of a given attribute, by how much, and what its effect will be on attaining sustainability goals. Second, it suggests that we need to better understand the extent to which the twelve attributes introduced here are independent of each other. If they are not, then increasing the compactness of one attribute will likely increase the compactness of the other and vice versa.

Given data on these attributes, we can study the correlations among them. When there is no significant correlation between two attributes, we can suspect that they are independent of each other. This, in turn, would suggest that an effective compact city strategy would benefit from acting to increase the compactness of both attributes, and that acting on one of them cannot be a substitute for acting on the other, but can contribute to the overall goal of mitigating climate change. Data collected for a global sample of 200 cities in the *Atlas of Urban Expansion—2016 Edition* on nine out of the twelve compactness attributes introduced earlier (all except Mass, Mix, and Concentration) can be used to calculate the correlations among them and to determine which ones are indeed significant. But before we engage in calculating correlations, we should ask ourselves whether we expect the various compactness attributes to be correlated to each other *a priori* and, if so, why?

We begin by noting that we should expect that the first three attributes—Density, Saturation, and Mass—would be correlated with each other, because both Saturation and Mass are factors of density. In fact, as we plan to show in a forthcoming paper titled "The Anatomy of Density", Density can be decomposed into factors that, when multiplied together reconstitute density. For example, Density can be decomposed into two factors, Built-up Area Density and Saturation:

(1) Density = Population ÷ Urban Extent;
(2) Built-up Area Density = Population ÷ Built-up Area; and
(3) Saturation = Built-up Area ÷ Urban Extent.

Therefore,

(4) Density = Built-up Area Density × Saturation.

It is clear, from examining equation (4) that Saturation is a factor of Density. In other words, other things being equal, the more saturated a city is, the denser it is. The same is true for Mass. By definition,

(5) Mass = Total Floor Area ÷ Urban Extent.

We define Crowding as the number of people per unit of floor area:

(6) Crowding = People ÷ Total Floor Area.

Therefore,

(7) Density = Crowding × Mass.

In other words, Mass is also a factor of Density. Other things being equal, the greater the Mass in the city, the denser it is. The same is true when we focus on the density of dwelling units rather than on floor area density.

(8) Dwelling unit density = Number of Dwelling units ÷ Urban Extent.
(9) Number of Dwelling Units = Total Floor Area ÷ Average Floor Area of Dwelling Unit.
(10) Dwelling Unit Density = Mass ÷ Average Floor Area of Dwelling Unit.

Dwelling Unit Density should therefore be proportional to Mass and highly correlated with Mass as well. We should therefore expect Saturation, Mass, and Dwelling Unit Density to be significantly correlated with Density.

Next, we must ask ourselves whether there is any reason to suspect that the three density-related attributes—Density, Saturation, and Mass—would be correlated with the second group of attributes, those associated with the shape compactness of urban footprints. Baruah et al. (2017), for example, find less leapfrogging development in former French African colonies is associated with higher average densities than those found in former British African colonies, suggesting that higher

Contiguity values may be associated with higher densities.

Clearly, we can expect the two attributes of the shape compactness of cities that are associated with Cohesion—Proximity and Exchange—to be highly correlated with Cohesion. All three measure the degree to which the shape as a whole approximates a circle, where the circle is taken to be the most compact shape in terms of access between any two locations within it and between locations within it and its center. These three attributes should also be positively correlated, but to a lesser degree, with Contiguity. Contiguity is not measured with respect to a circle, and has less to do with overall accessibility in the city and should therefore be expected, a priori, to have positive yet weaker correlations with Cohesion, Proximity, and Exchange.

Fullness presents an interesting case by itself. It measures the degree to which the urban footprint is compact taking into account the constraints imposed upon it by the physical landscape. Since many cities are not constrained by their physical landscapes, and since the Fullness value in these cities is equal to their Exchange value, we can expect Fullness to be positively correlated with Exchange, and therefore with Cohesion and Proximity as well, but to a lesser degree than their correlations among themselves. We can also look at the increase in the shape compactness of cities once we replace Exchange by Fullness, namely once we take landscape constraints into account. We can define the **Compactness Correction Factor** as the percentage increase in Exchange compactness when Exchange is replaced by Fullness.

(4) Compactness Correction Factor = (Fullness ÷ Exchange) – 1.

Other things being equal, the more stringent the landscape constraints, the higher the Compactness Correction Factor is expected to be. We have reason to believe that cities with stringent landscape constraints are likely to be denser than cities with little or no landscape constraints. We can therefore expect the Compactness Correction Factor to be positively correlated with Density and hence with Saturation and Mass as well.

Finally, we can ask ourselves whether the three compactness indicators associated with the distribution of land uses and the spatial organization of streeets and roads—Mix, Walkability, and Connectivity—are likely to be correlated with the density attributes or the geographic shape attributes of cities. First, it is not clear why cities with higher levels of Mix, and where workplaces are closer to residences, would have higher or lower densities. Second, Baruah et al. (2017) find that gridded cities in more planned former French colonies in Africa—those with higher values of Walkability—are indeed denser than less-gridded cities in less planned former British colonies in Africa. And third, we have no *a priori* reason to suspect that, other things being equal, cities with higher levels of Connectivity will also be found to be more or less dense or that the shape of their urban footprints will be more or less compact.

As noted earlier, we can examine the correlations among all of the compactness attributes discussed here—except Mass and Mix—by looking at data for a global sample of 200 cities. We shall return to his topic after introducing the methodology used to measure these attributes in the following section.

To conclude, in this first section of this essay, we presented the theoretical framework for studying the shape compactness of urban footprints, couching it in the broader study of the compactness attributes of cities and their potential to mitigate climate change. In the following section, the methodology section, we focus on the way we obtained the global sample of 200 cities, the way we obtained their urban footprints in three time periods—1990, 2000, and 2014—and the way we defined and measured the compactness properties of their urban footprints. We then present the results of our measurements of compactness in the global sample of 200 cities, and we introduce and analyze the correlations between them. In a third section, we focus on findings associated with the key forces that act on urban footprints to make them more or less compact, using specific examples of cities that illustrate the action of each one of those forces. In 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 shape 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 if the densification of cities can contribute to the reduction of travel distances and hence to the reduction the greenhouse gas emissions associated with travel, then so can making the shapes of urban footprints more compact. If that is indeed the case, then cities can and should employ a set of pragmatic tools that can increase their shape compactness over time.

People flock to cities to be in closer proximity 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 fizzle 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—that brought about by greater urban density, that brought about by rounder urban footprints, and that brought about by the morphology of land use and transportation networks—increases the access of people to each other, facilitating all manners of contact and exchange between them. If preferences for proximity, access, and connectivity would be the only forces acting on urban footprints then we should expect densities to be very high and urban footprints to be very compact, approaching the shapes of circles. In reality, however, the footprints 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 urban footprints more compact or less compact and seek to gain a greater understanding of their effects on urban form.

2. Methodology

In this second section, we describe the methodology for obtaining empirical results: Selecting the global sample of cities, identifying the urban footprints of the cities in the sample in three time periods, and then measuring the compactness properties of these urban footprints in these three time periods. We then display the results of our compactness measurements and calculate the correlations between them.

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 stratified sample drawn from a universe of 4231 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 4231 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 metropolitan area boundary 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 these 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 4231 cities are shown in Fig. 1 below. Details regarding the statistical properties of the universe of cities can be found in a companion working paper (Galarza Sánchez, Liu, Angel, & Thom, 2018).

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 ($8 \times 4 \times 3 = 96$), or boxes, to which an observation in the universe of cities must belong. After distributing all 4231 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 Fig. 2 below.

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

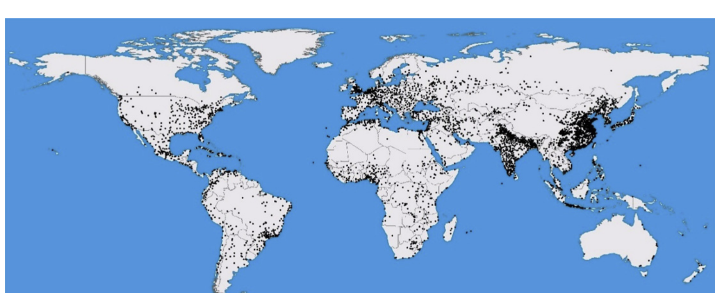

Fig. 1. The universe of all 4231 cities that had 100,000 or more in 2010.

Fig. 2. The global sample of 200 cities.

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 States 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 km 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 Fig. 3 below.

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 Fig. 4 below.

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

1 *Fringe* open space pixels, all open space pixels within 100 m 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 ha 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 Fig. 5 below.

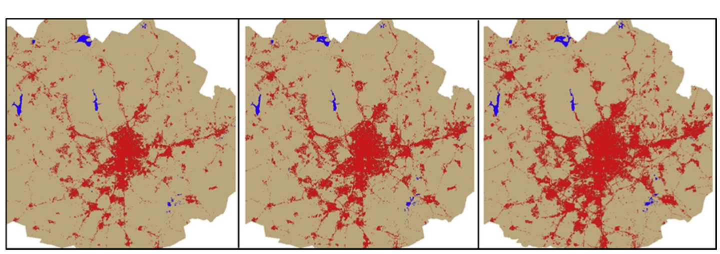

Fig. 3. The three-way classification of Madrid into water (blue), built-up (red), and open space (brown), at the T1 (1991), T2 (2002) and T3 (2010). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

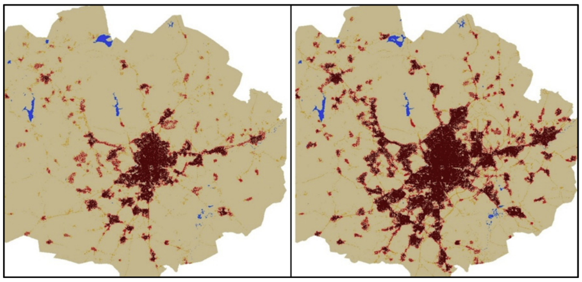

Fig. 4. The sub-classification of built up area into urban pixels (dark red) suburban pixels (red) and rural pixels (ochre) in Madrid, Spain in May 1991 (left) and May 2010 (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

The differentiation of the study area into the three classes of built-up, three classes of open space, and water facilitated the creation of rules that were used to identify *urban clusters* across the study area. Urban clusters are discrete patches of urbanized open space and their interior built-up areas, surrounded by rural open space. 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, Fig. 5 suggests that there were dozens of urban clusters in 1991 and 2010. We can see the clusters more clearly in Fig. 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), was 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.

We employed a rule based on the size and geographic proximity of clusters to each other to determine whether they would 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, 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 was always equidistant from edge of the cluster. The buffer distance for a given cluster was a function of the built-up area of the cluster, resulting in a buffered area equal to one-quarter the area of the cluster. The inclusion rule united all clusters whose buffers intersect one another into a singular set. The new set of clusters defined the city's *urban extent*. Fig. 7 shows urban extent of Madrid in 1991 and 2010.

The formulation of the inclusion rule was the result of attempts by the research team to group urban clusters in ways that corresponded to accepted notions of what constituted the spatial extent of cities. In certain respects, the task was a form of pattern recognition. Sometimes the pattern was easy to discern, when there was a single large cluster surrounded by open countryside, and other times it was more difficult, when there were many clusters of varying sizes in different proximities to each other, similar to, but typically more complex than the Madrid example.

It is important to note here that the choice of a universal inclusion 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. Correlations among compactness attributes in the global sample of cities

Given our brief survey of twelve compactness attributes of cities in Section 1.2 above, in this essay we have chosen to focus on four compactness attributes of urban footprints— Cohesion, Proximity, Exchange, and Fullness—in the global sample of cities. We do have data for five other compactness attributes of cities—Density, Saturation, Contiguity, Walkability, and Connectivity—for the global sample of cities, and we do plan to investigate those attributes further in the global sample of cities and publish additional essays on the results of our investigations in the future.

Altogether, we have data for nine of the twelve attributes described in Section 1.2, all except Mass and Mix, for all cities in the global sample for three time periods: 1990, 2000, and 2014. The descriptive statistics for these nine compactness attributes of cities in the global sample for the year 2014 are given in Table 1 below.

Table 2 below presents the correlations between these nine compactness attributes. Many of these correlations are significant, namely the probability that they are not correlated (their *p*-value) is smaller than 5 percent. Many of these correlations are expected, as we noted earlier in Section 1.3.

First, Saturation is correlated with Density because it is a factor of Density. Second, Density is correlated with Contiguity, as reported by Baruah et al. (2017): Cities with less leapfrogging and less fragmented urban extents are denser. Third, Density is correlated with Walkability, as reported by Baruah et al. (2017): Cities with denser street grids are denser. We have no explanation of why density would be correlated with the density of arterial roads, but it is. It is important to note here that, as we suspected, Density is not correlated with any of the shape compactness attributes of cities: Proximity, Cohesion, Extent, and Fullness. It strengthens our contention that density and shape compactness are two independent attributes of cities that require separate analytical and policy approaches.

We do find, however, that Saturation, as well as Contiguity are correlated with the shape compactness attributes of cities. Cities with

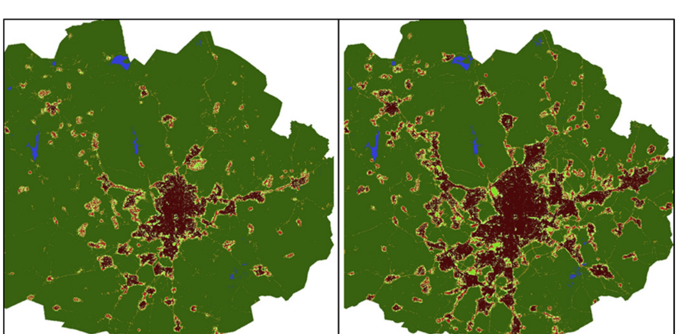

Fig. 5. The sub-classification of open space into fringe open space (light green), captured open space (bright green), and rural open space (dark green) in Madrid, Spain in May 1991 (left) and May 2010 (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

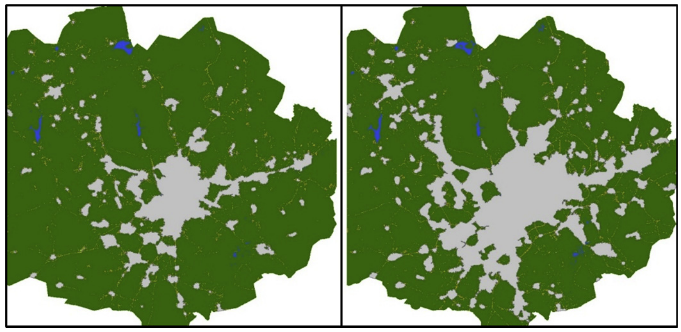

Fig. 6. Urban clusters (grey) in Madrid in 1991 (left) and in 2010 (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

more compact footprints—in terms of Proximity, Cohesion, Exchange, and Fullness—are also more contiguous and more saturated, as expected. We should expect Fullness, in particular, to be correlated with Saturation, and it is: Cities with physical barriers to expansion would be expected to be more saturated. More saturated cities are also found to have significantly more 4-way street intersections and significantly denser arterial grids, but we have no adequate explanation for these correlations.

Finally, we find very high correlations, exceeding 0.90, between the three attributes of the shape compactness of cities: Proximity, Cohesion, and Exchange. The three measures indeed focus on related properties of city shapes: the closeness of all locations to the CBD, the closeness of all locations to each other, and the closeness of all locations to their centroids. Given these high correlations, we shall report on the results pertaining to only one of them—Cohesion—in the remaining sections of this essay. The results for the two other attributes are no different. Not surprisingly, we find high correlations, of the order of 0.6+, between Fullness and these three shape compactness indicators. This is to be expected because Fullness is not different than Exchange in cities on flat terrains and, more generally it measures Exchange in the presence of physical barriers to expansion. These physical barriers render the shapes of city footprints less compact, hence the lower correlations with Proximity, Cohesion, and Exchange.

To conclude this section, the most relevant finding for the purposes of acting on compactness attributes of cities with the purpose of climate change mitigation—and, more specifically, the reduction of greenhouse gas emissions from road transport by reducing the total Vehicle Kilometers Traveled (VKT)—is that action on densification and action on the shape compactness of cities are independent interventions, where progress on one is unlikely to cause progress on the other. This lends support to our contention that action on the shape compactness of cities is a new and important component of any strategy seeking to effect climate change through changes in urban form.

Table 1
Descriptive statistics for nine compactness attributes circa year 2014 for the universe of cities, weighted by city weights.

Compactness Attribute	No. of Cities	Mean	95% Conf. Intervals		Minimum	Maximum
Density	200	52.4	47.0	57.8	7.4	372.4
Saturation	200	0.63	0.62	0.64	0.48	0.84
Contiguity	197	0.64	0.61	0.67	0.19	1.00
Proximity	200	0.77	0.75	0.78	0.36	0.96
Cohesion	200	0.76	0.74	0.77	0.38	0.96
Exchange	198	0.64	0.62	0.66	0.19	0.86
Fullness	198	0.71	0.70	0.73	0.22	1.00
Walkability	198	20.5	18.7	22.4	1.4	77.6
Connectivity	198	1.4	1.3	1.5	0.3	3.7

2.4. Measuring the shape compactness of urban footprints

In this section, we define the precise indices that were used to measure Proximity, Cohesion, Exchange, and Fullness in the global sample of cities. Following Angel et al. (2010, 444), we construct these

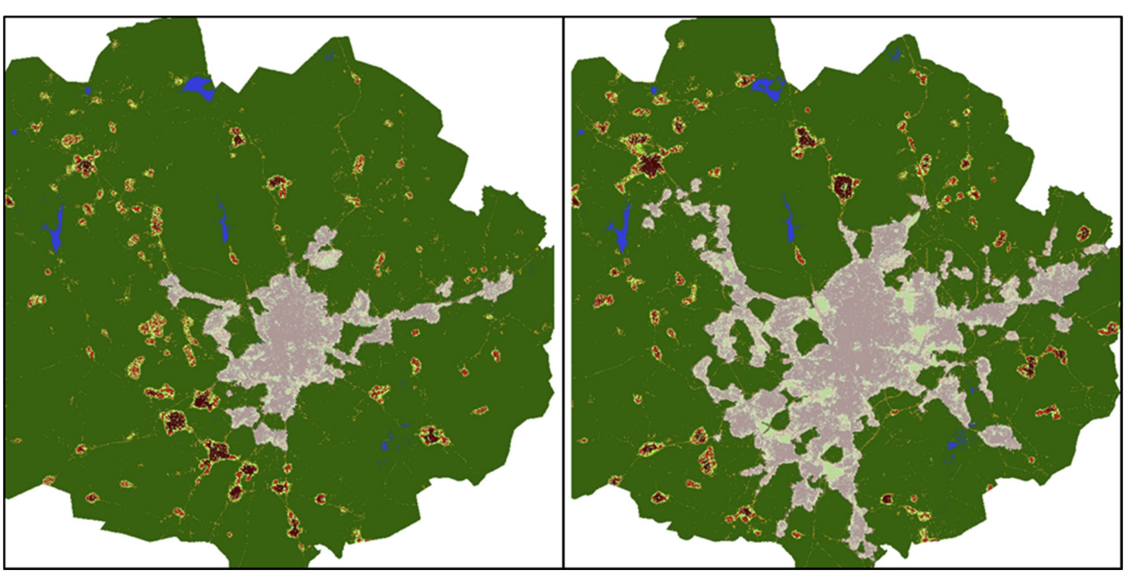

Fig. 7. The Urban Extent of Madrid (grey) in May 1991 (left) and May 2010 (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Table 2
The Pearson Correlation Matrix for nine compactness attributes for the year 2014 (correlations with *p*-values smaller than 5% are in bold).

Compactness Attribute	Density	Saturation	Contiguity	Proximity	Cohesion	Exchange	Fullness	Walkability	Connectivity
Density	1								
Saturation	**0.243**	1							
Contiguity	**0.330**	−0.027	1						
Proximity	0.040	**0.194**	**0.178**	1					
Cohesion	0.052	**0.193**	**0.201**	**0.996**	1				
Exchange	0.036	**0.216**	**0.184**	**0.965**	**0.950**	1			
Fullness	0.082	**0.226**	**0.197**	**0.630**	**0.627**	**0.635**	1		
Walkability	0.139	**0.238**	0.085	**0.228**	**0.236**	**0.242**	0.111	1	
Connectivity	**0.220**	**0.307**	−0.104	−0.032	−0.043	0.002	0.131	**0.206**	1

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 were used to construct, measure, and analyze the four compactness indices studied 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 has a 30-meter 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% 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 (PRX)** of a city is the ratio of the average beeline distance from all points in the Equal Area Circle to its center and the average beeline distance from all points of the city's urban extent to its Central Business District (CBD) identified by its City Hall.
- The **Cohesion Index (COH)** of a city is the ratio of the average beeline distance from all points to all other points in the Equal Area Circle and the average beeline distance from all points to all other points in the city's urban extent.
- The **Exchange Index (EXC)** of a city is the share of its urban extent within an Equal Area Circle located at the centroid of its urban extent.
- The **Fullness Index (FUL)** 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 are urban footprints when we take physical barriers to urban expansion into account.

- The **Compactness Correction Factor (CCF)** 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 the four compactness indices for all cities in the global sample. 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.

Fig. 8 shows the 2014 urban footprints of 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 extents. 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 footprints of cities are shown in declining order of their Cohesion Index, from the highest in the sample, Shanghai, China, with a Cohesion Index of 0.96, to the lowest in the sample, Cabimas, Venezuela, with a Cohesion Index of 0.36. This should be interpreted to mean that the average distance between two locations in Cabimas is three times longer that it would have been if its urban footprint were a circle. In Shanghai, on the other hand, there is no appreciable difference between the average beeline distance between all locations within its urban footprint and the average distance between locations in its Equal Area Circle.

The distribution of values for the Cohesion Index in 2014, shown in Fig. 9 below, resembles a normal distribution, with an average value of 0.77. The values for the Index are constrained between 0 and 1 but there are very few cities that do not display a tendency towards being circular. Hence, there are only very few low values for the Proximity Index, and none below 0.35.

Fig. 10 shows cities with the highest 16 and 16 lowest Fullness Index values in the global sample of cities. The orange circles are the Buildable Land Circles centered at the centroids of their 2014 urban extent. In the top left corner of each image, values are given for the Fullness Index (FUL). The maps of the urban footprints of cities are shown in declining order of their Fullness 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 Fullness Index of 0.22. This should

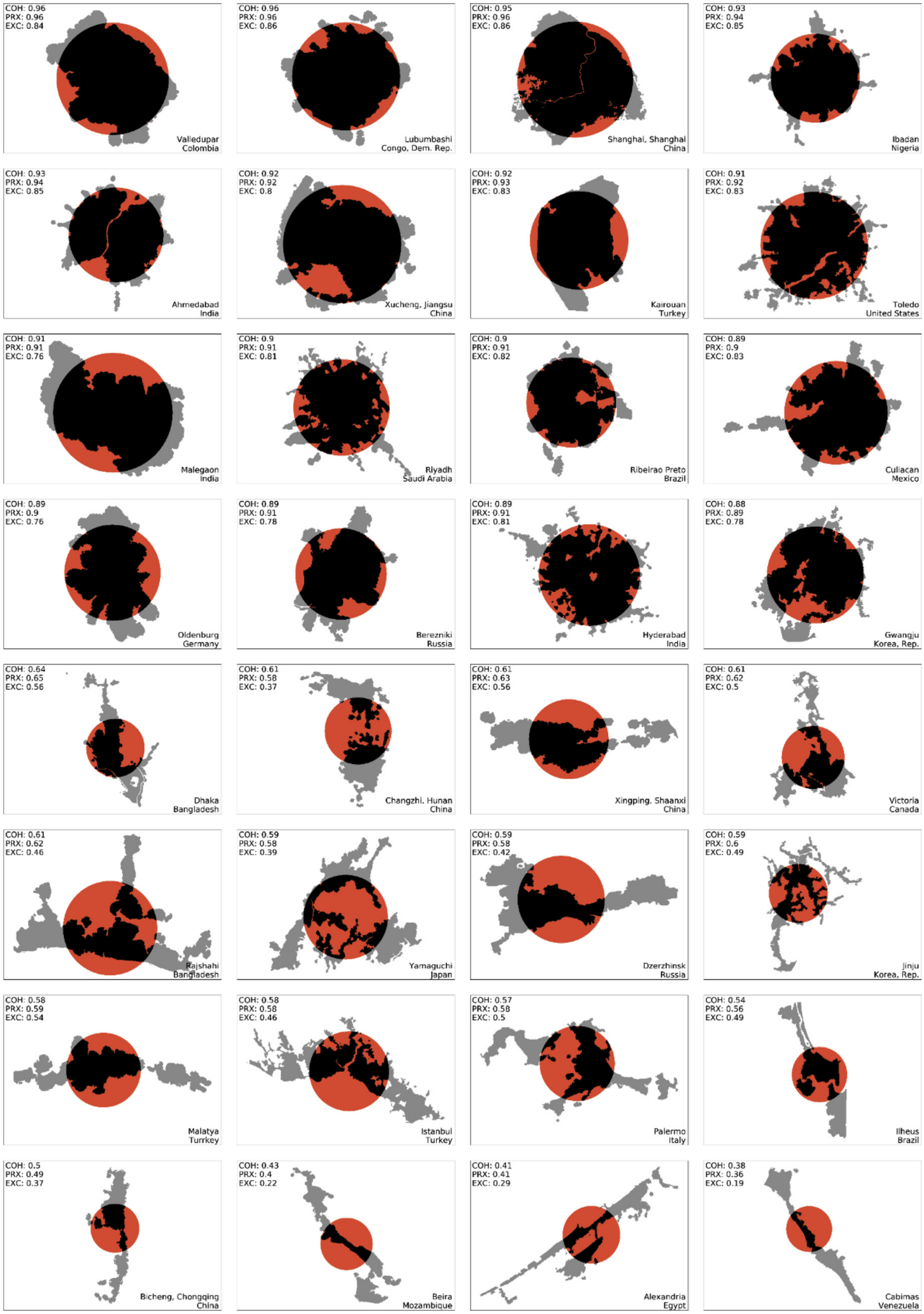

Fig. 8. Cities with the highest 16 and lowest 16 values for the Cohesion Index in the global sample. The orange circles are the Equal Area Circles centered at the centroids of their urban extents in 2014. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

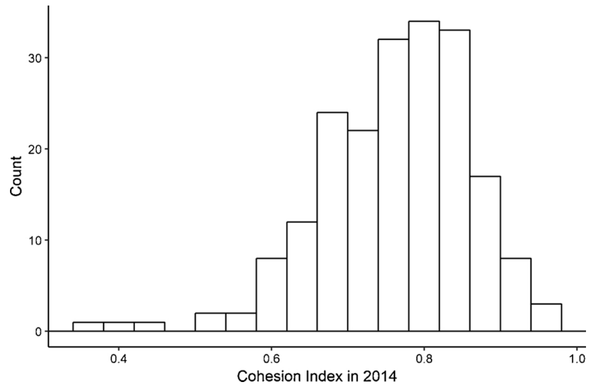

Fig. 9. The distribution of compactness index values, in this case for the Cohesion Index for 2014, in the global sample of cities resembles a normal distribution with an average value of 0.77.

be interpreted to mean that the urban footprint of Caracas—even though it is far from resembling a perfect circle—is as compact as can be given the physical landscape features in and around it. In the base of Beira, its expansion—and the expansion of its neighboring cities—has resulted in unifying their urban clusters into one meandering urban extent with a very low Fullness value, because Beira is surrounded by a lot of buildable land where development could have resulted in a more compact urban footprint.

Fig. 11 shows cities with the highest 28 and lowest 4 ratios between the Fullness Index and the Exchange Index—i.e. the Compactness Correction Factor—in the global sample of cities. The dark orange circles are the Equal Area Circles and the light orange circles are the Buildable Land Circles, both centered at the centroids of their 2014 urban extent. In the top left corner of each image, values are given for the Compactness Correction Factor (CCF), for the Fullness Index (FUL), and for the Exchange Index (EXC). The maps of the urban footprints of cities are shown in declining order of their Compactness Correction Factors (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 should be interpreted to mean 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, Fig. 12 shows 16 cities with the highest increase and 16 cities with the highest decline in their Cohesion Indices between 1990 and 2014 in the global sample of cities. 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 footprints 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.

It should be noted in passing here that Angel, Parent, Civco, and Blei (2012), 257–319) and Angel (2012), 223–247) mapped the urban footprints and analyzed the change in their compactness values in a representative group of 30 cities over a two hundred year period, 1800-2000. Graphic animations of the expansion of these cities over time are available online in *The Atlas of Urban Expansion—2016 Edition* (Angel et al., 2016). The results of this analysis are not duplicated here and the reader is referred to these sources for further study of the change in the compactness of urban footprints over a much longer time period.

Given the maps of the urban footprints 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.

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 footprints 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 footprints of cities in 2014, for example—(5.9 ± 1.4%)—was on slopes exceeding 15 percent. And like steep slopes,

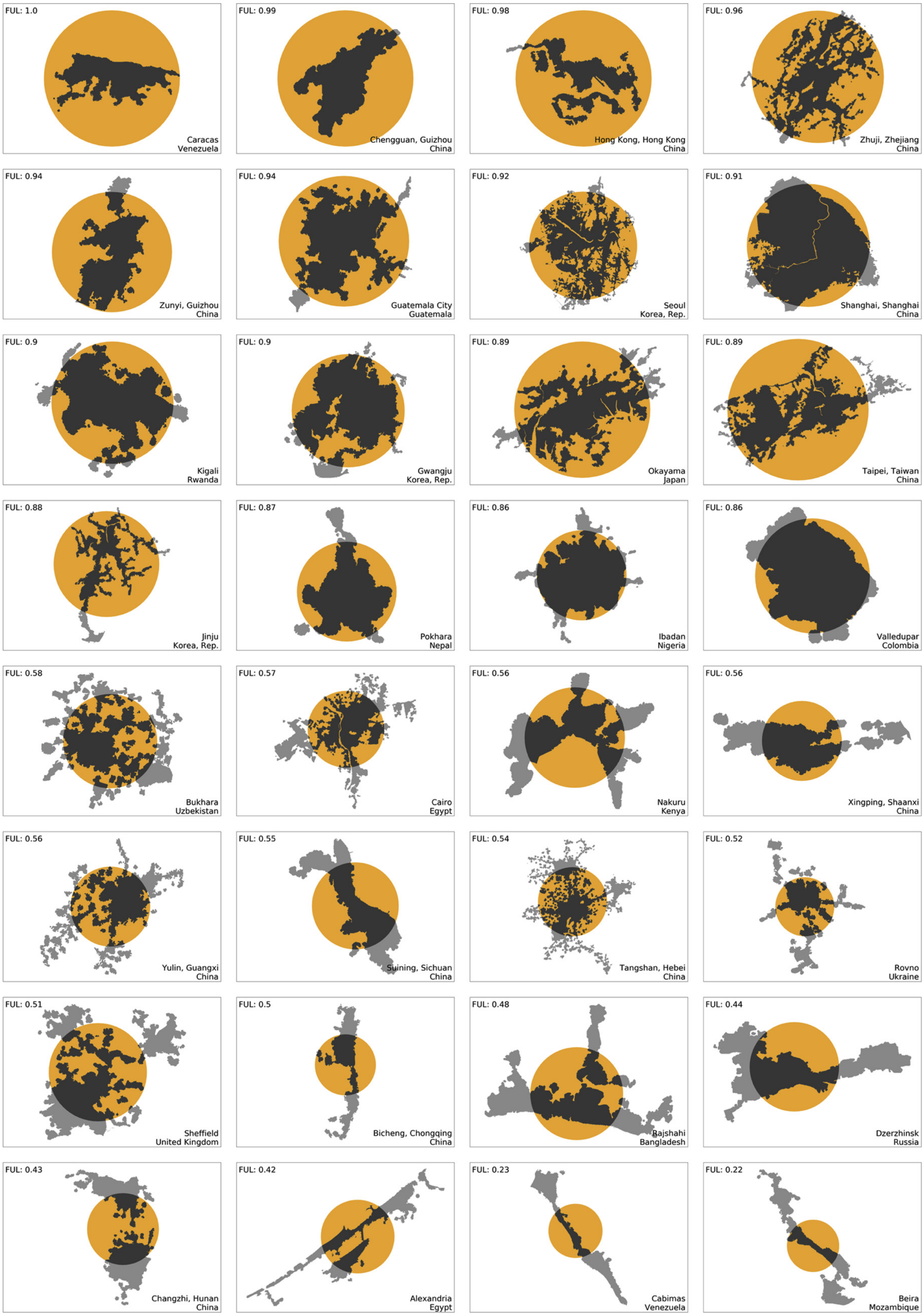

Fig. 10. Cities with the highest 16 and lowest 16 Fullness Index values in the global sample. The orange circles are the Buildable Land Circles centered at the centroids of their 2014 urban extents. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

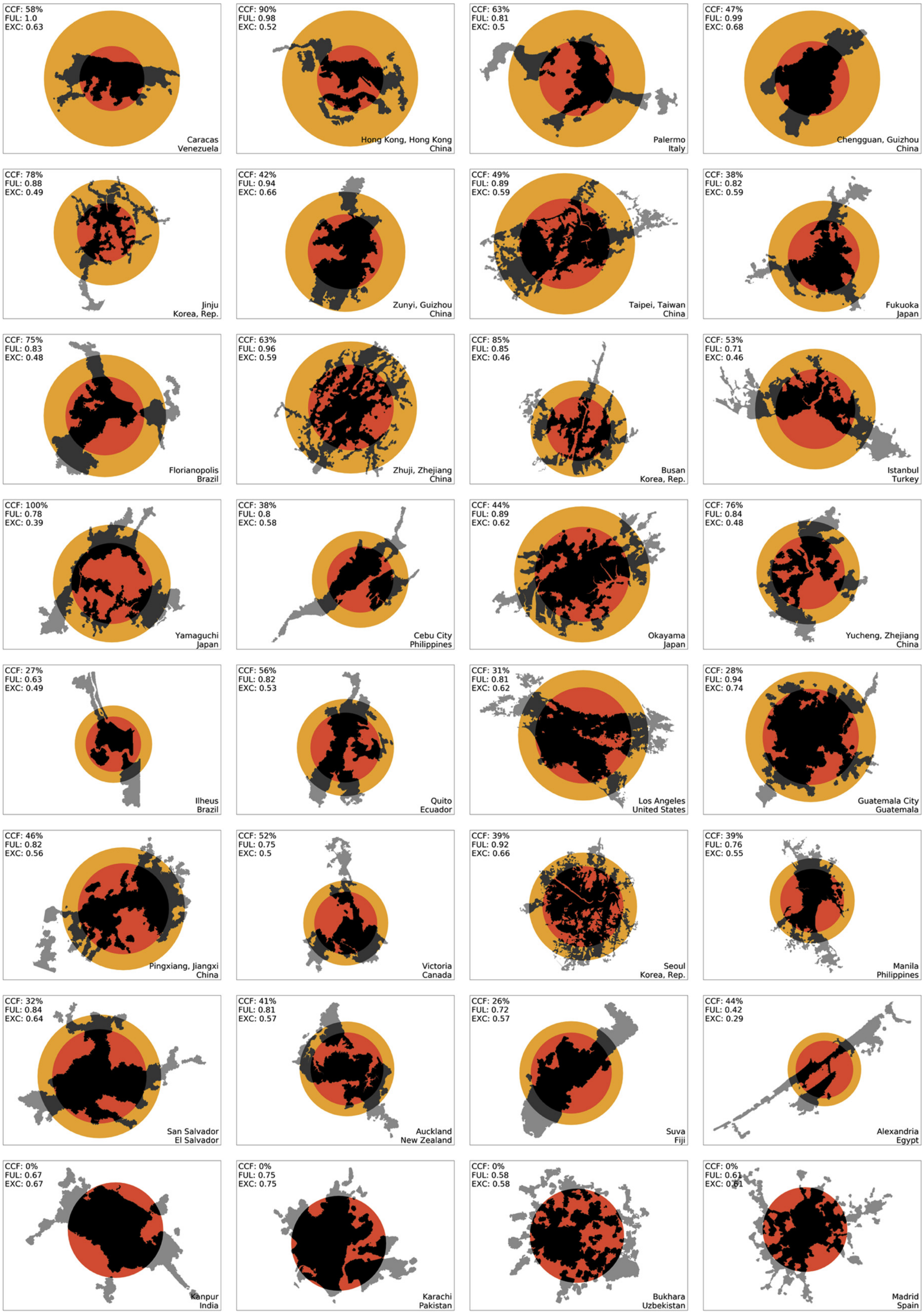

Fig. 11. Cities with the highest 28 and lowest 4 Compactness Correction Factors in the global sample. The dark orange circles are the Equal Area Circle and the light orange circles are the Buildable Land Circle, all centered at the centroids of their 2014 urban extents. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

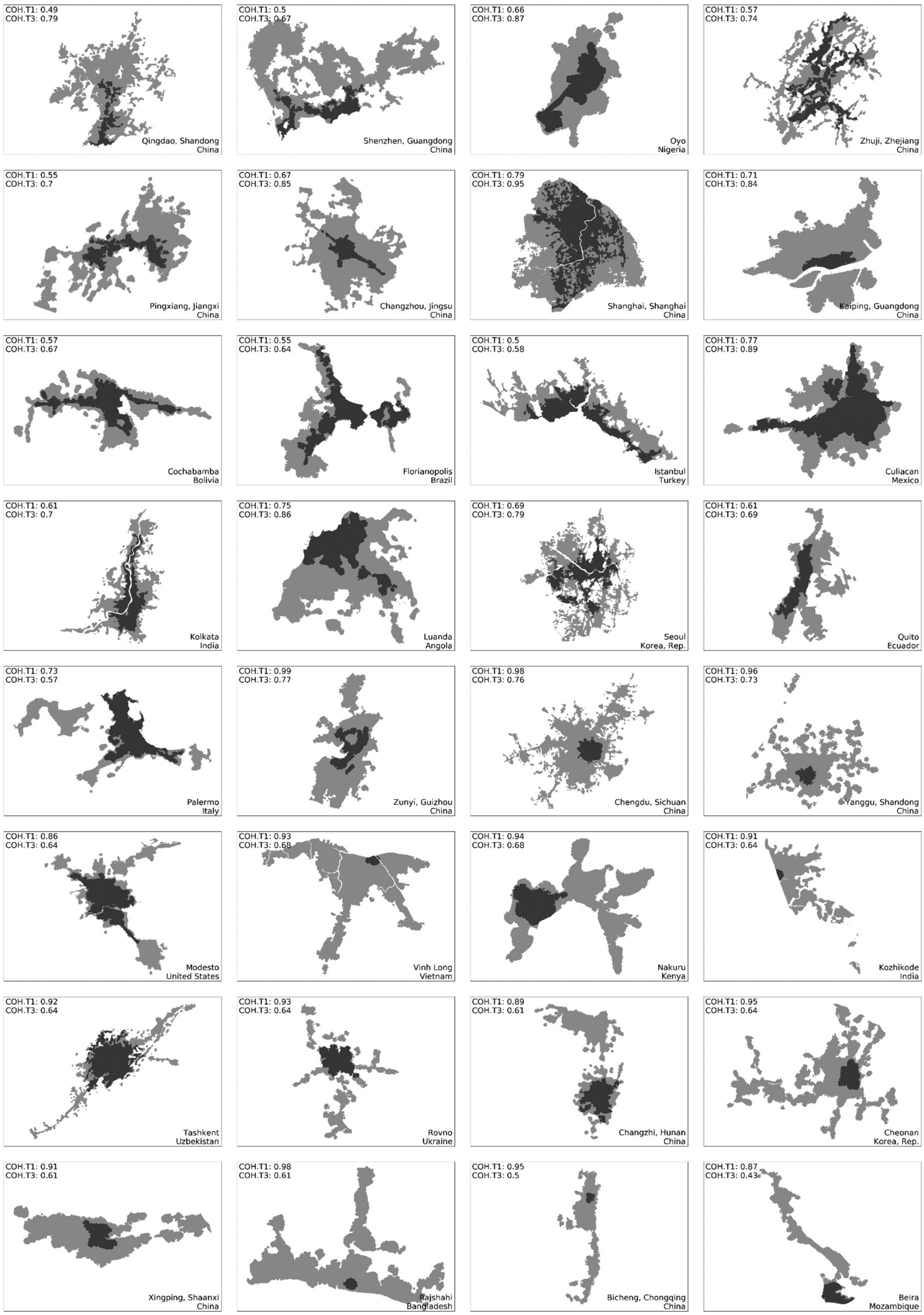

Fig. 12. 16 cities with the highest increase and 16 cities with the highest decline in Cohesion Index values between 1990 and 2014 in the global sample of cities.

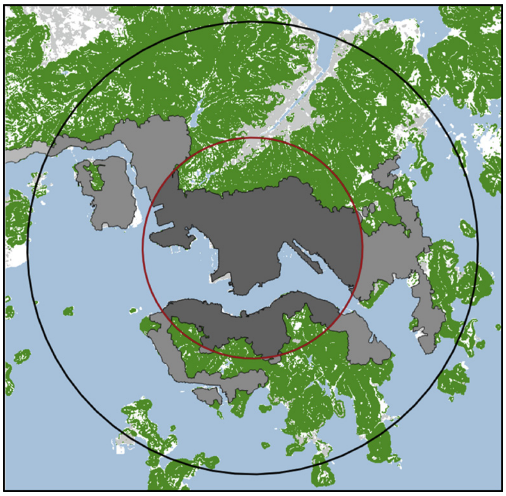

Fig. 13. The 2014 Urban Extent of Hong Kong (grey) occupies only 52 percent of its Equal Area Circle (red) and is, therefore, among the least compact cities in the global sample. Yet it occupied 98 percent of its Buildable Land Circle (black), confirming that—given its physical environment—it is as compact as can be (Buildable land shown in white, non-buildable land in green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Fig. 14. 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.

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

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 quintile 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 Fullness 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 Fig. 13).

More generally, the footprints of cities surrounded by natural barriers—be they high slopes or bodies of water—tend to be less compact. The Buildable Land Ratio—defined earlier as the ratio of the areas of Equal Area Circle and the Buildable Land Circle—is a measure of the degree to which a given city is exposed to natural barriers. The smaller that 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 and Cohesion Index in 2014 are significantly correlated to each other. For a 10 percent increase in Buildable Land Ratio we can expect the Cohesion Index to increase, on average, by 1.8 percent (Adjusted $R^2 = 0.14$). 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 located on open, flat plains.

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°. We find that in the global sample of cities the share of the area of urban footprints on slopes higher than 15° highly correlates with the Compactness Correction Factor, with R-squared of 0.61. For a 10 percent increase in the Compactness Correction Factor, we can expect a 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% of its urban extent is in areas with slopes that are steeper than 15 percent (see Fig. 14 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 footprints. 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, often at a higher cost and at higher levels of risk.

Steep slopes are only one kind of physical 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 ground 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.

3.1.2. Merging of adjacent settlements

As cities expand outwards, their urban footprints come to include settlements—cities, towns 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 of, so called, planned non-compactness. 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 footprints 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 urban extent of the previous period; (2) extension: building at the edge of and away from the urban extent of the previous period; (3) leapfrog: building in areas surrounded by rural open space away from the urban extent of the previous period; and (4) inclusion: incorporating distinct settlements already built in the previous period into the city's 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 two or more settlements into a common, single urban footprint. 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 below.

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 by 2014, less than two-thirds its 1991 value, largely due to the inclusion of existing settlements on its periphery into 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 Fig. 15 below).

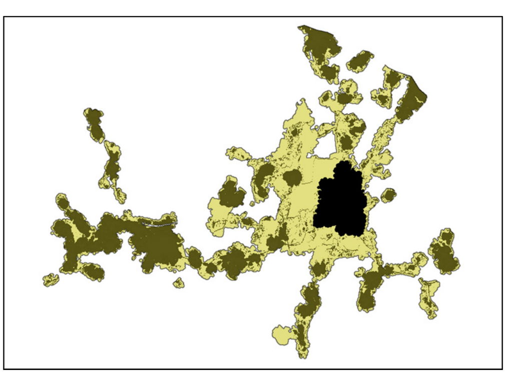

Fig. 15. As Cheonan, South Korea, expanded between 1991 (black) and 2014 (light green), it merged with a large number of settlements on its periphery (light and dark red). 53 percent of the area added to the city's urban extent during this period was attributed to 'inclusion', reducing its Cohesion Index from 0.95 in 1991 to 0.64 in 2014 (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

3.1.3. Inter-city roads and rail lines

The compactness of contemporary cities—where commuter trip destinations are dispersed everywhere rather than located predominantly at city centers—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 commuter 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 that of 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 Fig. 16 below). It has experienced a

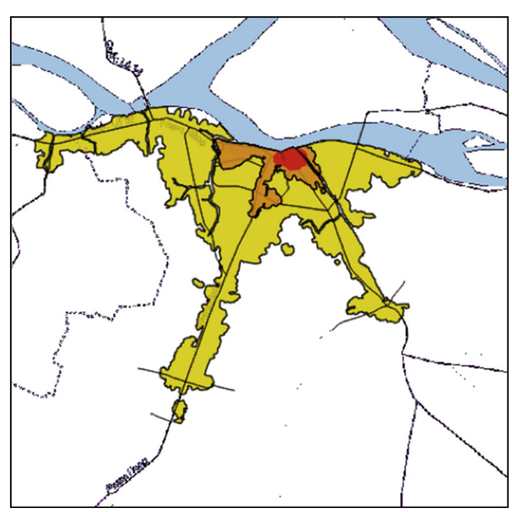

Fig. 16. By 2014, Vinh Long, Vietnam, has expanded along four intra-city roads leading away from its center becoming less compact than it was in 1990 in the process.

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.

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 by road or rail 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 footprints 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 a matter of 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 footprints 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 cite 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 converted to urban use (Angel, Valdivia, & Lutzy, 2007). This often results in highly fragmented urban footprints—i.e. a smaller level of saturation of urban footprints by built-up areas—and hence in larger urban footprints, but not necessarily in less compact ones. Indeed, the urban footprints 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 ha, around 13 percent of the land area of England. Overall there was a decrease of 790 ha (less than 0.05 percent) in the area of Green Belt between 31 March 2016 and 31 March 2017″ (U.K. Department of Communities & Local Government, 2017).

Sheffield, England, is in the global sample of cities. Its urban extent in 2014 encompassed not only the city of Sheffield, but also a number of other nearby cities in South Yorkshire County, including Rotherham, Barnsley, and Doncaster. As a result, its centroid has moved towards the Northeast, further away from Sheffield's Central Business District (see Fig. 17). The Cohesion Index of Sheffield's urban extent in 2014 was quite low—it ranked the 32nd lowest in the global sample of 200 cities—and we suspect that it was largely because of the South Yorkshire Country Green Belt (shown in green in Fig. 17). Sheffield and its neighboring cities and towns are situated in relatively flat land—much of it under cultivation—and its urban extent could, in principle, be quite compact. Yet, as Fig. 17 shows, 49 percent of its Buildable Land Circle is occupied by its greenbelt. As a result, its 2014 Fullness Index was 0.51, the 9th lowest value in the global sample of cities. Policy decisions—in this case land use restrictions—appear to have a major impact on the shape compactness of cities. Sheffield was only half as compact as it could have been if its urban expansion were not constrained by its green belt. In truth, Sheffield has no room to expand, except beyond its green belt towards the East, the Southeast, and the South as it is bound on the West by the Peak District National Park and on the North by the West Yorkshire Green Belt. Future urban expansion will thus make it even less compact. Again, while the green belt provides a high level of amenity value in Sheffield, it already increases average trip length by some 50 percent, leading to substantial increase in 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 footprints could not be investigated in sufficient detail at the present time.

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 river, 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. Preferences 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 along the Cebu Strait in the Philippines. It is also evident in cities not in the global sample like Miami, Florida, along the Atlantic Ocean, or Montevideo, Uruguay, along the La Plata River.

Alexandria, to take one example, could have expanded in a southeasterly direction into the flatlands of 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 2014 Fullness Index, 0.42, was the third lowest in the global sample of 200 cities (see Fig. 18 below).

3.1.6. Land market distortions

When land markets function properly, there is very little leapfrogging as cities expand outwards. In the United States, for example, most leapfrogging was found to be one kilometer or less away from the built-up areas of cities (Burchfield et al., 2006). 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—the National Fund for Workers' Housing (Instituto del Fondo Nacional de la Vivienda para los Trabajadores) 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.

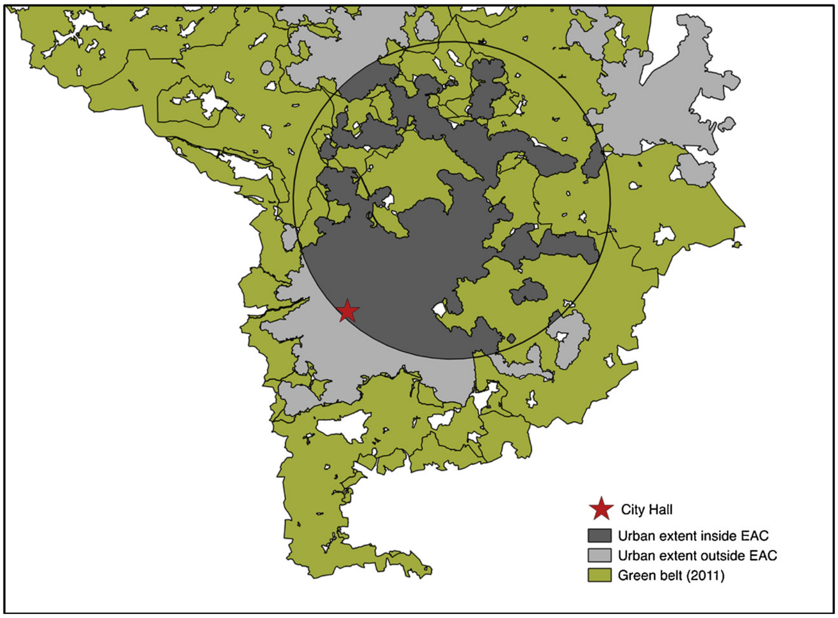

Fig. 17. The greenbelt in metropolitan Sheffield, England, (shown in green) occupies half of its Buildable Land Circle, making Sheffield one of the least compact cities in the global sample of cities. The urban extent of the city is shown in dark grey inside the circle and in light grey outside it. City Hall is shown as a red star. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

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 Fig. 19). Locating housing estates in distant locations also resulted in increased levels of abandonment. The overall rates of abandonment

Fig. 19. INFONAVIT Housing on the urban periphery in Mexico (Photo Credit: Habitat D.F.).

Fig. 18. The city of Alexandria in Egypt, expanded further along the Mediterranean shore between 1990 (dark grey) and 2014 (light gray), becoming less compact in the process. The land in the area shown in white is buildable. (For the interpretation of references to colour in this figure legend, the reader is referred to the web version of this article).

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 were 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 footprints must await the collection of better and more extensive global data.

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 footprints does begin to emerge. We can begin to distinguish some of the key forces—to be sure, 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 Fig. 20 below shows the average values of the four compactness indices defined earlier for the universe of cities as a whole, comprising all 4231 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 represented by each city in the sample. The table and the figure also show the 95 percent confidence intervals for these indices. They show that the average values of all four indices declined. The confidence intervals for the year 2000 overlap with those of 1990 and 2014, suggesting that the decline in the average value in the 1990–2000 and the 2000–2014 periods was not statistically significant. The confidence intervals for the year 1990, however, do not overlap with those of 2014, suggesting that the decline in the average values 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.

We can obtain a stronger result by looking at the weighted average of the *change in compactness* in individual cities between two periods, rather than comparing the weighted average of compactness values in two time periods. We applied a weighted paired *t*-test to compare the compactness values in the two periods for individual cities. It tests whether the differences in compactness indices between two periods

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

	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]
Fullness Index	0.752 [0.734, 0.769]	0.737 [0.721, 0.753]	0.713 [0.695, 0.731]

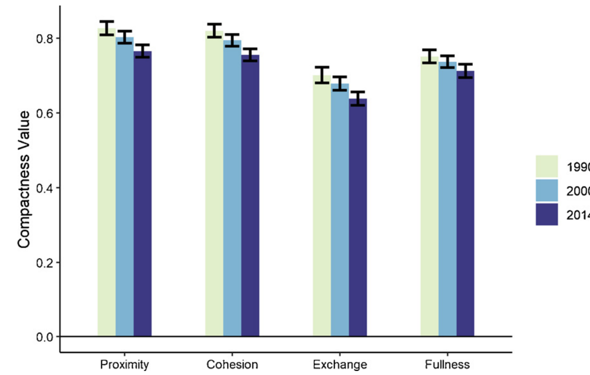

Fig. 20. All four types of compactness indices decreased over the three periods. The decline in compactness between 1990 and 2014 was significant at the 95 percent confidence level.

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.

Given this stronger result, we can conclude that the shape compactness of cities the world over has been in significant decline during both 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 or less pronounced 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, but it does 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 rather weak, as only the Proximity Index and Cohesion Index show a significant difference (at the 95 percent confidence level) in the magnitude of decline between cities in more developed countries and cities in less developed countries.

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 between two time

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.

	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]
Fullness Index	−0.014 [−0.027, −0.001]	−0.024 [−0.039, −0.010]	−0.038 [−0.058, −0.019]

Table 5
The weighted mean and confidence interval 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
Fullness Index	−0.031 [−0.059, −0.003]	−0.036 [−0.071, −0.001]	−0.017 [−0.056, 0.022]	0.47

periods 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 99 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. 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^2 of 0.236. The results also pertain to the other two compactness indices, and to both the 1990–2000 and the 2000–2014 period.

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 annual rate of decline of urban extent densities in less developed countries between 1990 and 2014, 2.0 ± 0.4%, was significantly faster than the annual rate of decline in density in more developed countries during that period, 1.3 ± 0.3%. The implications of these findings are highlighted in the concluding section of this essay.

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 ⅔*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 $0.9054R$ (see, e.g., García-Pelayo, 2005,). 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.

Table 6
A multiple regression model with the percentage change of Proximity Compactness during the 1990–2014 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

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 cities in the sample for all three 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.

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 also have data on the average beeline commute distance traveled for 40 U.S. cities for the year 2000 (see Angel & Blei, 2016). The U.S. Census provides data on the origins and destinations of commuter trips from one census tract to another, but does not provide data on network distances of those trips. We used beeline distances between centroids of census tracts as proxies for network distances. Using the number of trips between pairs of census tracts as weights, we then calculated the average beeline commute distance for each of the 40 cities. We constructed a multiple regression model where the logarithm of the average beeline commute distance is the dependent variable and the logarithms of population, density, and cohesion are the independent variables. The model (see Table 8) explains 85% of the variation in observed average distance of journey to work in a stratified sample of 39 U.S. cities.[1]

The model is weaker than the model shown in Fig. 7. The coefficient for the Exchange Index is only significant at the 10% confidence level (p-value – 0.0629). Still, the model suggests that the observed average observed beeline commuting distance in U.S. cities—a number that should, in principle, be associated with Vehicle Miles Traveled (VMT) and hence with energy expended in travel and with greenhouse gas emissions from travel—varies with population, density, and the cohesion of the urban footprint of cities in a similar fashion to the model shown in Table 7 above: (1) Other things being equal, the larger the population, the greater the distance traveled on the journey to work. A 10% increase in city population results in a 2.7% increase in the average distance of the journey to work; (2) Other things being equal, the higher the population density in the city, the shorter the distance of the journey to work. A 10% increase in population density results in a 3% decrease in the average distance of the journey to work; and (3) Other things being equal, the more compact the shape of the urban footprint of the city, the shorter the distance of the journey to work. A 10% increase in shape compactness results in a 2.1% decrease in the average distance of the journey to work.

Table 7
Linear regression with the log of the average distance between points within the urban extent as the dependent variable in the global sample of cities in 2014. Adjusted R-squared = 0.996, N = 198.

Variable	Coefficient B	P-value
Log Population Size	0.502	< .0001
Log Population Density	−0.504	< .0001
Log Exchange Index	−0.655	< .0001

Table 8
Linear regression with the log of the average beeline commute distance between points within the urban extent as the dependent variable in 39 U.S. cities in 2000. Adjusted R-squared = 0.853, N = 39.

Variable	Coefficient B	P-value
Log Population Size	0.269	< .0001
Log Population Density	−0.320	< .0001
Log Cohesion Index	−0.252	0.045

In short, the shape compactness of cities has a similar effect on the average distance of the journey to work that density has: Increasing either of them shortens it. The order of magnitude of the effects is similar too, but the observed effects in this model are weaker in comparison with the earlier model, e.g. a 10% increase in shape compactness results in a 6.5% decrease in the journey to work in the first model, but only in 2.1% decrease in the distance of the journey to work in the second one.

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 footprints as well as those that increase the average densities of urban footprints. 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 footprints 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—which can be 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). All of these factors, as we saw in Section 1 of this essay, have been identified in the compact city literature as potentially contributing to climate change, but there is no mention in the panels' assessment of the possible impact of the shape

[1] We eliminated New York from the model because its urban extent (9,511 km^2) is larger than a single metropolitan labor market, where all workplaces are accessible to all residences within a tolerable commute time. If New York is not eliminated, the p-value for the Log Cohesion Index is 0.160 and Cohesion is not significant at the 5% p-value.

compactness of cities on climate change. If indeed population density matters to climate change—at least insofar as higher densities result in shortening distances among locations within cities—then the shape compactness of cities should matter as well.

Güneralp 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 greenhouse gas 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 greenhouse gas emissions in cities. Higher-density cities make for shorter trips and therefore have lower totals of vehicle kilometers traveled (VKT). They also make public transit more feasible, both leading to lower energy use and lower GHG emissions. In Günerlap *et al's* assessments as well, the shape compactness of urban footprints has not yet entered this discussion.

It follows, 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 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 & Renaud, 1995, 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 (Fig. 21), a decline associated with increasing urban incomes and the increased availability of affordable urban transportation. In this essay we have also reported on the decline in the compactness of urban footprints over time. Both trends do not bode well for actions aimed at combating greenhouse gas emissions through urban form.

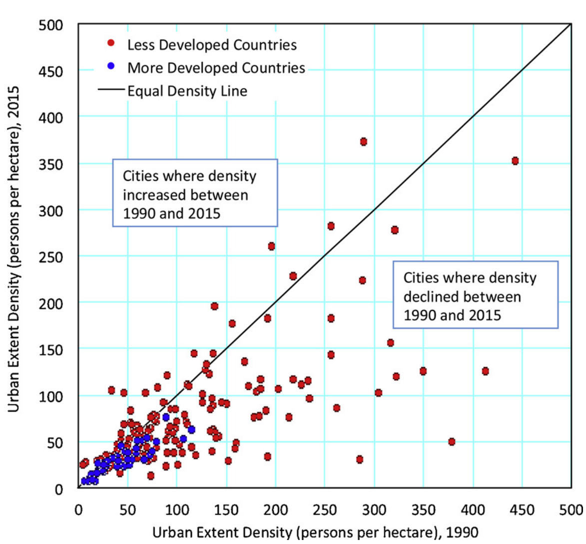

Fig. 21. In the global sample of 200 cities, average urban extent densities declined in 72 percent of cities in less developed countries and in 75 percent of the cities in more developed countries between 1990 and 2015.
Source: *Atlas of Urban Expansion—2016 Edition*, online at www.atlasofurbanexpansion.org, 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 than for 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 greenhouse gas 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 is typically the preferred course of action for those concerned with energy conservation, with the mitigation of global greenhouse gas 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 by restricting housing supply and by citizens who want to protect green spaces on the urban periphery or reduce greenhouse gas emissions associated with longer trips. 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, calling both for acceptable densification and for orderly urban expansion, 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 footprints 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 footprints, 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 the shape of 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 for public open spaces, large and small, 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 footprints 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 footprints. 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 at a higher cost.

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 ambitious plans to guide urban expansion—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 the shape of urban footprints more compact, namely the desires of households and businesses for greater proximity, being 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—be they commuter rail lines, freeways, or expressways that allow for travel at higher speeds in some directions but not in others—on urban expansion should also be taken into account when seeking to guide it, as these tend to make the shape of urban footprints 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 that can help equalize travel times along alternate routes so as to reduce 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 (Vásconez, Galarza, Angel, Montezuma, & Fonseca, 2015), for example—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 the shape of urban footprints more compact or less compact, with serious implications for our plans to mitigate climate change in the years to come.

Funding source

The research work leading to this paper was supported by grants from the United Nations Human Settlemnents Programme (UN-Habitat) in Nairobi, Kenya; the Lincoln Institute of Land Policy in Cambridge, MA, USA, and the Marron Institute for Urban Management at New York University in New York, NY, USA.

Declarations of interest

None.

References

Ahlfeldt, G., & Pietrostefani, E. (2018). *Demystifying compact urban growth: Evidence from 300 studies from across the world, OECD Regional Development Working Papers, 2018/03.* Paris: OECD Publishinghttps://doi.org/10.1787/bbea8b78-en.

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

Angel, S. (2012). *Planet of cities.* Cambridge MA: Lincoln Institute of Land Policy170–185.

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

Angel, S., & Parent, J. (2011). Non-compactness and voter exchange: Towards a constitutional cure for gerrymandering. *Northwestern Interdisciplinary Law Review, 4*(1), 89–130.

Angel, et al. (2016). *Atlas of urban expansion*online at:(2016 edition). www.atlasofurbanexpansion.org.

Angel, S., Valdivia, I. M., & Lutzy, R. (2007). *Urban expansion, land conversion, affordable housing, and energy use in China: The case of Zhengzhou, Henan*The China Policy Workshop Paper. Online at. Princeton Universityhttp://wws.princeton.edu/research/pwreports_f07/wws591a.pdf.

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

Angel, S., Parent, J., Civco, D. L., & Blei, A. M. (2012). *Atlas of Urban expansion.* Cambridge MA: Lincoln Institute of Land Policy.

Baruah, N. G., Henderson, J. V., & Peng, C. (2017). *Colonial legacies: Shaping African cities. SERC discussion papers, SERCDP226.* London, UK: Spatial Economics Research Centre, London School of Economics and Political Science.

Bechle, M. J., Millet, D. B., & Marshall, J. D. (2011). Effects of income and urban form on urban NO2: Global evidence from satellites. *Environmental Science and Technology, 45*(11), 4914–4919.

Beer, A. (1994). Spatial inequality and locational disadvantage. *Urban Policy and Research, 12*(3), 181–199.

Bento, A. M., Cropper, M. L., Mobarak, A. M., & Vinha, K. (2003). The impact of urban spatial structure on travel demand in the U.S. *The Review of Economics and Statistics, 87*(August (3)), 466–478.

Bertaud, A., & Renaud, B. (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.

Bonfantini, B. (2013). *Centri Storici: infrastrutture per l'urbanità contemporanea, 64*, 153–161 Territorio.

Boyko, C. T., & Cooper, R. (2011). Clarifying and Re-conceptualizing density. *Progress in Planning, 76*, 1–61.

Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics, 65*(1), 91–98.

Bunting, T., Filion, P., & Priston, H. (2002). Density gradients in Canadian metropolitan regions, 1971–96: Differential patterns of central area and suburban growth and change. *Urban Studies, 39*, 2531–2552.

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

Burton, E. (2000). The compact city: Just or just compact? A preliminary analysis. *Urban Studies, 37*(11), 1969–2006.

Burton, E. (2001). The compact city and social justice. *Housing, Environment and Sustainability,* (April), 1–16.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity and design. *Transportation Research D, 2*(3), 199–219 1997.

Cheshire, P. C., & Hilber, C. A. L. (2008). Office space supply restrictions in Britain: The political economy of market revenge. *The Economic Journal, 118*(529), F185–F221.

Churchman, A. (1999). Disentangling the concept of density. *Journal of Planning Literature, 13*(4)), 389–411.

Duranton, G., & Turner, M. A. (2011). The fundamental law of Road congestion: Evidence from US cities. *American Economic Review, 101*, 2616–2652.

Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis. *Transportation Research Record, 1780*, 87–114.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association, 76*, 265–294.

Ezeh, A., Oyebode, O., Satterthwaite, D., Chen, Y., Ndugwa, R., Sartori, J., Mberu, B., Melendez-Torres, G. J., Haregu, T., Watson, S. I., Caiaffa, W., Capon, A., & Lilford, R. J. (2016). The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet* Published online October 16.

Forsyth, A., Oakes, J. M., Schmitz, K. H., & Hearst, M. (2007). Does residential density increase walking and other physical activity? *Urban Studies, 44*(4), 679–697.

Frank, L. D., & Pivo, G. (1994). Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking. *Transportation Research Record, 1466*, 44–52.

Gaigné, C., Riou, S., & Thisse, J.-F. (2011). *Are compact cities environmentally Friendly? Discussion paper DE2011-15.* Research Institute for Economics and Business Administration (RIEB), Kobe University March.

Galarza Sánchez, N., Liu, Y., Angel, S., & 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*, 3475–3482. online at: http://iopscience.iop.org/article/10.1088/0305-4470/38/16/001/pdf.

Gehl, J. (2010). *Cities for people.* Washington D.C: Island Press 288 pp.

Grammemos, F. (2011). European urbanism: Lessons from a city without suburbs. *Planetizen, 10*(February), online at: https://www.planetizen.com/node/48065.

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., Fragkias, M., Li, X., & Seto, K. C. (2017). Global scenarios of urban density and its impacts on building energy use through 2050. In: *Proc. Natl. Acad. Sci. 114*(August (34)), 8945–8950. https://doi.org/10.1073/pnas.1606035114 online at:.

Harari, M. (2017). *Cities in bad shape: Urban geometry in India, working paper.* The Wharton School, University of Pennsylvania October.

Hidalgo, D., & Gutiérrez, L. (2013). BRT and BHLS around the world: Explosive growth, large positive impacts and many issues outstanding. *Research in Transportation Economics, 39*(1), 8–13.

Holtzclaw, J., Clear, R., Dittmar, H., Goldstein, D., & Haas, P. (2002). Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and Use—Studies in Chicago, Los Angeles and San Francisco. *Transportation Planning and Technology, 25*, 1–27.

Horvath, A. (2004). Construction materials and the environment. *Annual Review of Environment and Resources, 29*, 181–204.

Jones, C., Leishman, C., MacDonald, A., & Watkins, D. (2010). In M. Jenks, & C. Jones (Eds.). *Economic viability, dimensions of the sustainable City*. Oxford: Springer.

Kockelman, K. (1997). Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Francisco Bay area. *Transportation Research Record: Journal of the Transportation Research Board, 1607*, 116–125. https://doi.org/10.3141/1607-16.

Matsumoto, T. (2011). *Compact City policies: Comparative assessment. Compact City Policies: Comparative Assessment, 47th ISOCARP Congress 2011*1–12.

McGarigal, K., & Marks, B. J. (1994). *FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure, version 2.0.* Oregon State University.

Melia, S., Parkhurst, G., & Barton, H. (2011). The paradox of intensification. *Transport Policy, 18*(1), 46–52.

Mitlin, D., & Satterthwaite, D. (2013). *Urban poverty in the global South: Scale and nature.* London and New York: Routledge.

Mogridge, M. J. H. (1985). Transport, land use and energy interaction. *Urban Studies, 22*, 481–492.

Müller, D. B., Liu, G., Løvik, A. N., Modaresi, R., Pauliuk, S., Steinhoff, F. S., et al. (2013). Carbon emissions of infrastructure development. *Environmental Science & Technology, 47*, 11739–11746.

Næss, P. (2005). Residential location affects travel behavior—but how and why? The case of Copenhagen metropolitan area. *Progress in Planning, 63*, 167–257.

National Research Council (2009). *Driving and the built environment: The effects compact development on motorized travel, energy use and Cemissions.* 257 pp. Available at:Washington, D.C: The National Academies Press. http://www.nap.edu/catalog.php?record_id=12747.

Neuman, M. (2005). The compact city fallacy. *Journal of Planning Education and Research, 25*(1), 11–26.

Noland, R. B. (2001). Relationships between highway capacity and induced vehicle travel. *Transportation Research Part A: Policy and Practice, 35*, 47–72. Available at: http://www.sciencedirect.com/science/article/pii/S0965856499000476.

OECD (2012). *Compact City policies: A comparative assessment. OECD green growth studies*. OECD Publishing.

OECD (2015). *OECD urban policy reviews: Mexico 2015 transforming urban policy and housing finance.* OECD Publishinghttps://doi.org/10.1787/9789264227293-en.

Permana, A. S., Perera, R., & Kumar, S. (2008). Understanding energy consumption pattern of households in different urban development forms: A comparative study in Bandung City, Indonesia. *Energy Policy, 36*, 4287–4297.

Ritter, C. (1822). Die Erdkunde im Verha¨ltniss 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.

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine, 25*, 80–91.

Saiz, A. (2010). The geographic determinants of housing supply. *Quarterly Journal of Economics, 125*(3), 1253–1296.

Salon, D., Boarnet, M. G., Handy, S., Spears, S., & Tal, G. (2012). How do local actions affect VMT? A critical review of the empirical evidence. *Transportation Research Part D: Transport and Environment, 17*, 495–508. https://doi.org/10.1016/j. trd.2012.05.006 ISSN: 1361-9209.

Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G. C., Dewar, D., Huang, L., Inaba, A., Kansal, A., Lwasa, S., McMahon, J. E., Müller, D. B., Murakami, J., Nagendra, H., & Ramaswami, A. (2014). *Chapter 12: Human settlements, infrastructure and spatial planning. 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, Average radial distance of points within a circle, n.d., online at: http://mathforum.org/library/drmath/view/62529.html.

Thomas, L., & Cousins, W. (1996). A new compact city form: Concepts in practice. In M. Jenks, E. Burton, & K. Williams (Eds.). *The compact City: A sustainable urban form?*. UK: Oxford.

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.

United Nations (1993). *Agenda 21: Earth summit—The United Nations programme of action from Rio*.

Vásconez, J., Galarza, N., Angel, S., Montezuma, R., & Fonseca, S. (2015). *Preparando ciudades de rápido crecimiento para su expansión: Informe sobre Valledupar y Montería, 83*, Colombia: Medio Ambiente y Urbanización129–154.

Vorontsova, A. V., Vorontsova, V. L., & Salimgareev, D. V. (2016). The development of urban areas and spaces with the mixed functional use. *Procedia Engineering, 150*, 1996–2000.

Wilson, G., & Baldassare, M. (1996). Overall "sense of community" in a suburban region: The effects of localism, privacy and urbanization. *Environment and Behavior, 28*(1), 27–43.

World Health Organization (WHO) (2011). *Health co-benefits of climate change mitigation-transport sector: Health in the green economy.* Geneva: World Health Organization.

Zhou, B. B., & Kockelman, K. M. (2008). Self-selection in home choice: Use of treatment effects in evaluating relationship between built environment and travel behavior. *Transportation Research Record: Journal of the Transportation Research Board, 2077*, 54–61.

Shlomo Angel (corresponding author) 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.

Sara Arango Franco is a data scientist at the Marron Institute of Urban Management at New York University.

Yang Liu is a research scholar and statistician at the Marron Institute of Urban Management at New York University.

Alejandro M. Blei is a research scholar at the Marron Institute of Urban Management at New York University.