MIT Libraries | DSpace@MIT

MIT Open Access Articles

Does metropolitan form affect transportation sustainability? Evidence from US metropolitan areas

Massachusetts Institute of Technology

Does metropolitan form affect transportation sustainability?
Evidence from US metropolitan areas.

Andres Sevtuk[1]
Reza Amindarbari[2]

1 Department of Urban Studies and Planning, Massachusetts Institute of Technology
2 College of Design, North Carolina State university.

Corresponding author: Andres Sevtsuk (asevtsuk@mit.edu)
MIT DUSP. Room 10-485. 77 Massachusetts Avenue, Cambridge MA 02139
Phone: +18579280928

Abstract

In this paper we examine transportation sustainability in American metropolitan areas using transportation-related CO_2 emissions, public transit accessibility and commuting times as indicators. Though variations in these indicators may stem from historic contexts, policies, institutional arrangements, social and cultural origins, the spatial structure of metropolitan areas—in particular their formal characteristics—may also be a contributing factor. To test this relationship, we identify metropolitan form metrics from prior literature that are expected to impact transportation outcomes, and choose five metrics to which we introduce significant improvements. We apply the metrics to all 166 Combined Statistical Areas (CSAs) in the US, using an open-source GIS toolbox released along with the paper. Our findings demonstrate that form-based metrics provide a better explanation to CO_2 emissions, public transit accessibility and commuting times in US metro areas than the simpler population size or density metrics typically used in practice. We also show that counter to prior literature on urban scaling laws and economies of scale, which have argued that larger cities and metro areas are more sustainable per capita, transport related CO_2 emissions and transit accessibility are actually less favorable in larger CSAs when controlling for formal characteristics of metropolitan areas. Instead of scale, compactness has the highest elasticity with respect to transportation sustainability of metro areas.

1. Introduction

Urban sprawl costs the US economy more than $1 trillion a year via infrastructure and public services (Victoria Transport Policy Institute, 2015). Because of employment decentralization, Americans travel longer distances to work, which is deteriorating the mental and physical health of the nation (Jackson, 2003). More than 15 percent of the population in nearly all US metro areas is obese, costing cities an estimated $80 billion a year only in health care (Witters 2011). A large body of planning literature has focused on how these challenges are related to urban form, land use planning and car-dependency in American cities. Built environments encourage vehicular travel and all too often discourage even moderate levels of physical activity. Less work has focused on studying how metropolitan form—that is the shape characteristics of entire metropolitan areas—affect transportation choices and related quality of life outcomes in urban regions (Angel et al 2005; Burchfield et al 2006; Wheeler 2008).

This paper evaluates the relationship between metropolitan form and transportation outcomes. To capture formal characteristics of metro areas, we deploy five different shape metrics from prior literature and introduce important improvements to their specification. We apply the proposed metrics to all 166 Combined Statistical Areas (2015 CSAs) within the continental US[1] (excluding Puerto Rico) in order to classify US metro areas based on formal similarities, and assess how metropolitan form properties relate to transportation sustainability.

The paper is structured as follows. We first present the metrics we use to capture the formal

---

[1] CSAs are delineated based on commute patterns by the US Office of Management and Budget, and are comprised of one or multiple urban cores and surrounding low-density socio-economically linked areas.

properties of metropolitan areas and describe their improvements with respect to previous applications. Next, we apply the metrics to combined statistical areas in the US and illustrate extreme outliers. The following section examines how the forms of US metropolitan areas relate to CO_2 emissions per capita, average travel time to work, and the percent of population that can walk to rail transit from home—qualities which we associate with more sustainable land-use and transportation outcomes. We demonstrate that metropolitan form metrics, such as gravity-compactness and coverage provide a more detailed understanding of these outcomes at the metropolitan level than simpler population size and density metrics that are widely used in scholarship as well as policy-making to date. In contrast to the literature on urban scaling laws, our findings suggest that it is compactness—a formal property of metropolitan areas—not simply their scale that explains lower CO_2 emissions and better transit access. Section five discusses takeaways for policy and planning and directions for future research.

2. Metropolitan form metrics

An extensive body of literature has focused on the development of urban form metrics (Angel et al. 2010; Wu et al., 2017; Clifton et al. 2008; Parent et al. 2009; Prosperi et al. 2009; Song et al. 2013; Wentz 1997). While the idea of landscape morphology and shape analysis in geography dates to at least early twentieth century (Sauer 1925), data and computational limitations till recently prohibited a wider implementation of these metrics (Boyce & Clark 1964; MacEachren 1985; Gibbs, 1961). With GIS and computing power, the field has considerably broadened. For instance, FRAGSTATS is a Spatial Pattern Analysis Program for Categorical Maps based on

remotely-sensed land cover data that offers tools for analyzing regional urbanization patterns in GIS (McGarigal & Marks 1995). Parent et al. (2009) and Angel et al.(2010) and Wentz (1997) utilized advances in GIS to develop urban form metrics for a set of global cities. Their proposed metrics describe continuous urbanized areas around metropolitan cores, examining the shape characteristics of urban edges, the relationship between a city's area and perimeter, and the ease with which the area can be traversed or circumvented. A unique focus on the overall shape of metropolitan areas, however, has so far ignored their internal structural characteristics, such as the distribution of population or jobs, or the pattern of their centers and land uses (Prosperi et al. 2009).

From a longer list of urban form metrics featured in literature, we identified five, which stood out for their expected relationship *vis a vis* metropolitan transportation outcomes. Of these, *Polycentricity* and *Gravity-compactness* capture characteristics of intra-urban spatial structure, weighed by employment and residential population respectively. There has been considerable debate about the influence of polycentric urban structure on transportation (Glaeser & Kahn 2001; Crane & Chatman 2003), which suggests that the spatial distribution of job and residents should exert an influence on commuting distances and mode-shares. Second, while density remains one of the most common metrics for describing cities and their transportation outcomes, the conventional metric has been critiqued for disregarding internal relationships between parts (Tsai 2004, US Census Bureau 2012). Our use of gravity-compactness tries to address some of these shortcomings. We also expected coverage, which describes the extent to which a metro area is blanketed with urbanized landcover, and fragmentation, which quantifies how continuous versus piecemeal the development pattern is, to affect transportation outcomes. Metropolitan

areas with less coverage and more fragmentation were expected to lead to longer commutes times and more pollution. Finally, while each of the four previous metrics describes the present conditions of metropolitan structure from a different angle, we also expected limits to future expansion to exert an influence on present density and commuting patterns. Even though expandability addresses future potential, limits to growth may be grandfathered into past and present development trends and expectations in the form of higher density, certain growth corridors or greater support for public transit (Catenaccio 2010). Where the specification of a metrics significantly diverges from existing practice, we try to distinguish it by name from analogous measures in the literature.

2.1. Gravity-compactness

Compactness measures how physically close the elements of a metropolitan region—their built-up areas, population or built form— are distributed with respect to each other. The closer they are, the more compact a metropolitan area is. In geometry, compactness is defined by how a shape corresponds to a circular disk. Bertaud and Malpezzi (2003) defined dispersion—the inverse of compactness—as the ratio between average distance from centroids of population tracts to the central business district (CBD), weighted by the population of the tracts, and the average distance of the same population from the centroid of a hypothetical circular city of the same size with uniform density. Angel et al. (2005) proposed an analogous definition of compactness, but additionally accounted for geographical constrains like water bodies or mountains. Instead of comparing the observed built-up area to an ideal circular disk, it is compared to developable land in the study area. The main limitation of both measures is that

5

they assume the existences of a center and overlook spatial relationships between different built-up areas with respect to each other. This can be problematic in cities where the CBD is less pronounced, not located at the geometric centroid of the city (e.g. coastal cities), or which have a polycentric structure. Second, Bertaud's index is not affected by the absolute changes of density—if the relative size balance in tracts is kept constant, but the population in each tract is increased at the same rate, the dispersion index remains constant.

Attempting to address these shortcomings, we have called the following metric gravity-compactness, which is computed as the weighted average gravity accessibility of all built up polygons with respect to each other (Supp. Figure 1). The metric makes use of Hansen's gravity accessibility index (Hansen, 1959), which is proportional to the population of each destination area and inversely proportional to its distance:

$$G_i = \sum_{j \epsilon M - \{i\}}^{n} \frac{W[j]}{d[i,j]}$$

Equation 1

, where Gi is the gravity index for built-up polygon [i] that is part of metro area M, W[j] is the population of destination polygon [j] and d[i,j] is the distance between their centroids. But since the measure is impacted by the total sum of weights [W], the metric has to be normalized to be comparable from city to city. Without normalization, compactness results would be higher in cities with higher population than cities with lower population, even if population is more compactly distributed in the latter. We find the normalized value G_{i_norm} by dividing each G_i with the total population in the city W_{total}.

Computing the gravity index from each built-up polygon to all other polygons in a metropolitan area, weighted by population, and taking the mean result across all polygons, can thus capture how compactly the metropolitan population is situated with respect to each other:

$$G = \frac{\sum_{i \epsilon M} W[i] * G_{i\_nrom}}{\sum_{i \epsilon G} W[i]}$$

<div align="right">Equation 2</div>

The proposed measure does not assume a central reference point, thus making it applicable to metropolitan areas of any shape.

2.2 Fragmentation

Expansion of metropolitan areas does not only occur along the edges of built up areas; leapfrog growth is common in most cities. Climatic conditions, availability of affordable land, flexible zoning regulations, ground-water distribution, and low transportation costs are often key drivers of fragmented urban growth (Burchfield et al. 2006).

Angel et al. (2005) describe *contiguity* as the ratio between the main (largest) built-up area of the city and the total built-up area of the city. The more built-up area is concentrated in the largest cluster, the more contiguous the city is. This measure is useful as long as the main built-up area constitutes a large portion of the total extent of the city. But the metric is not well suited to metro areas dominated by multiple smaller built-up clusters—it does not account for rank-size relationships between individual discontinuous areas beyond the largest cluster.

7

Similar to Angel et al. (2005), we assume that the fewer discontinuous developments there are, the less fragmented a metropolitan area is. However, our metric additionally accounts for size relationships between all built-up clusters, achieved by automating the procedure in GIS (Supp. Figure 2).

To compute fragmentation, first we compute the ratio between the area of each built-up polygon and the area of polygons smaller than it:

$$\frac{\sum_{i=1}^{n} I(A_i < A_j) A_i}{A_j} \quad j \epsilon \{1,2,\dots,n\}$$

<div align="right">Equation 3</div>

, where A_j is the area of polygon j. Fragmentation is then defined as the weighted average of these elements:

$$\frac{\sum_{j=1}^{n} \left(\frac{\sum_{i=1}^{n} I(A_i < A_j) A_i}{A_j} \right) \cdot A_j}{\sum_{j=1}^{n} A_j} = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{I(A_i < A_j) A_i}{A_{total}}$$

where I(.) is an indicator function that is 1 when its condition holds and 0 otherwise.

<div align="right">Equation 4</div>

The key improvement of the proposed index is that it accounts for the size relationships between all built-up clusters, making the metric suitable to metropolitan areas of any shape or pattern.

2.3. Area-doubling expandability

Area-doubling expandability the extent to which a metropolitan area can be doubled using available and buildable land around its current urban extent. Studies of land management have used urban growth boundaries as a limit for land supply (Hopkins and Knaap 2000). Growth boundaries, however, are insufficient for our purposes, since few metro areas have legally implemented them and their regulatory mechanisms vary widely. Administrative boundaries are also inadequate as reference land, since metropolitan area typically span several municipalities.

Saiz (2010) used a 50-kilometer radius from the centroid of the city – independent of administrative boundaries – as the search area for developable land. But as Wendell Cox (2011) pointed out, an invariant search radius makes results incomparable in cities of different size. Whereas in larger cities a 50-kilometer radius may barely cover the built-up area, in smaller towns it may contain several times their existing urban extent.

We define the reference area as twice the size of the current urban extent, which is relative to each metropolitan area. We have implemented a solution to identify the precise distance to which the boundaries of all currently built-up areas have to be offset so that the total area of each polygon exactly doubles. Unbuildable regions are subtracted from the doubled reference area. Area-doubling expandability is then defined as the ratio between the area of buildable parts of the doubled expansion area (A_e) and the existing built area (A_b). The index is one if the entire expansion area is buildable and zero if none of it is buildable:

Expandability $= A_e / A_b$

Equation 5

The key advantage is that the references area is relative to the size of each metro areas. While the choice of a doubled, as opposed to a tripled or quadrupled area is similarly arbitrary as Saiz' (2010) and Wendell Cox' (2011) measures, the exact size of this ratio is also not critical—it is rather important to use the same multiplier in comparing different metro areas.

2.4. Coverage

Coverage illustrates how much land in a metropolitan area is built out. We use this common definition here, but as reference land, deploy the convex hull —the smallest possible convex polygon circumscribing presently built up areas (Supp Figure 3). The "buildable" area of the convex hull is obtained by subtracting unbuildable areas from the complete hull:

$$Cov = \frac{A_{blt}}{(A_{cvx} - A_{unb})}$$

Equation 6

2.5. Relative polycentricity

Polycentricity typically describes the extent to which jobs in a metropolitan area are concentrated in multiple as opposed to a single center. Quantifying polycentricity is challenging since defining a center can be ambiguous and the number of centers is not necessarily a sufficient yardstick for a polycentric urban structure. Centers can have blurry boundaries and polycentricity can be

relative to the size of a city as well as the resolution of the lens with which we examine an urban landscape. An area can be a center locally, but not large enough to qualify as a center at the scale of the city or in comparison to other cities.

While a number of useful density-based methods for detecting centers have been proposed (Cleveland 1979; McDonald 1987; McDonald & McMillen 1990; McDonald & Prather 1994; McMillen & McDonald 1997; García-López & Muñiz 2010; Giuliano & Small 1991), they have either failed to account that several contiguous tracts can jointly form a single center or presumed the existence and location of a CBD. More generally, many studies have focused on identifying centers, but they have not considered the overall share of jobs or the relative size balance between centers as factors. Roca and Marmolejo (2009) proposed a functional approach for detecting sub-centers, which instead of density of zones uses the intensity of interaction with the CBD. With a similar functional approach, Limtanakool et al. (2009) utilized the Entropy Index, which describes how uniformly commuting flows are distributed among centers, as a measure of how polycentric a metropolitan region is.

Our relative polycentricity metric builds upon these methods to identify centers, but makes the employment density cutoff relative to city size. We also use Limtanakool's concept of spatial interaction between centers to include a size-balance consideration in the metric.

We use the following density-based criteria to identify centers at the census block-group level (other spatial units can also be used). First, job density in block-groups has to be two standard deviations above the mean job density in the metro area—commonly used threshold for identifying outliers. Second, if several adjacent block-groups pass the cut-

11

off threshold, they are grouped to form a joint center. Third, while many block-groups in

a metro area can have a high job density, we do not consider one a center if it does not

contain a sufficiently high share of total metro area employment. We have empirically

determined a reasonable threshold to be 10% of the square root of total jobs in the area.

This avoids identifying block-groups, whose job densities are high, but which only

contain a small fraction of metropolitan jobs in them, as centers. The minimum relative-

size threshold is adjustable. We use 10% of the square root of total jobs in the area as our

relative size cutoff below, which we found to intuitively identify subcenters in multiple

metro areas.[2]

Having defined centers, polycentricity depends simultaneously on the number of centers,

the sizes of the centers, and the relative size balance between centers found in a metro

area. We consider a region to be more polycentric if a) it has more centers, b) a greater

share of its total jobs is located in centers, and c) the more equally balanced the sizes of

its sub-centers are (the less any one center dominates). Relative polycentricity PC is thus

defined as:

PC=HI*N*Rc

Equation 7

, where N is the number of centers, and Rc is the ratio between the number of jobs found in all

centers to the total amount of jobs in the metro area as a whole. HI is Limtanakool et al.'s

Entropy Index (2009), which measures the degree to which the sizes of centers are homogenous:

$$HI = -\sum_{i=1}^{N} \frac{(z_i)ln(z_i)}{ln(N)}$$

Equation 8

, where Zi is the ratio of the number of jobs at center i to the total number of jobs in all N centers

in a region. Supplemental Figure 4 illustrates the 33 job centers identified in the Los Angeles-

Long Beach CSA—by far the most polycentric metro area in the US (PC=13.8).

3. Application to American Metropolitan Areas

We use the National Land Cover Database (MRLC 2011) and boundaries of CSAs to delineate

the extent of urbanized areas in the US. Combined statistical areas typically include a number of

core based statistical areas (CBSAs), whose individual built-up areas often form a large, regional

whole. Within each CSA, we reclassified the four land cover classes labeled as *developed* (low,

medium, high intensity, and open space) to urban, converted the urbanized extents to polygon

features and eliminated features smaller than 1 km². We also reclassified Wetlands and Open

Water land cover as unbuildable, which used for computing compactness, coverage, and

expandability. While urban growth boundaries, steep slopes and protected lands can also serve as

barriers for development and urban expansion, our application here unfortunately does not

account for these due data and time constraints. However, our open source GIS toolbox allows

13

other researchers to define unbuildable land in alternative ways using, for instance, slope, legal growth boundaries, protected areas and other restrictions in future work. We use 2010 block-group level population from the US Census and 2015 business locations from ESRI[3] to compute gravity-compactness (weighted by population) and relative polycentricity (weighted by employment) respectively. Supplementary Table 1 provides summary statistics of metropolitan form metrics across all CSAs and a full table of results including each CSA separately is also provided in supplemental Table 4.

Some descriptive statistics are worth pointing out. Data from American CSAs suggests that population density tends to be higher in larger metropolitan areas. Highest populations density is found in the two most populated CSAs: New York-Newark, NY (2,104pax/km2) and Los Angeles-Long Beach, CA (2,110 pax/km2). The lowest population density, on the other hand, is found in Edwards-Glenwood Springs, CO (567 pax/km2) and DeRidder-Fort Polk South, LA (569 pax/km2), which are both in the bottom 10%-ile in terms of CSA population (126,000 pax and 88,000 pax respectively). This is noteworthy in light of Angel's (2005) analysis of global urban expansion, where he found that as population doubled between 2000 and 2010, land area tripled—urban densities in all regions were found to be decreasing over time. Our data shows that larger metro areas in the US have higher population densities on average.

Seventy-five per cent of all metropolitan areas in the US have four or more centers, which are, on average, 28 kilometers apart from each other. Despite the continued dominance of the mono-

---

centric model in urban economic education (Alonso 1964; Muth 1969; Mills 1967), the assumptions of these models are inconsistent with the polycentric structure of employment in US metropolitan areas.

Most American metro areas are also spatially fragmented—built-up areas cover only a small portion of their convex hulls. Less than 12% of the metro area's convex hull is urbanized in seventy-five per cent of all CSAs.

Most American metropolitan areas do not face significant water barriers to growth. Seventy-five per cent of all metropolitan areas could potentially use 88% of land within an area-doubling buffer zone around them for growth—that is land that is unobstructed by water bodies.[4] Major metropolitan areas on the coasts are typically more constrained. Washington-Baltimore-Arlington CSA, Seattle-Tacoma CSA and Miami-Fort Lauderdale-Port St. Lucie CSA are the least expandable metro areas in the US, due to both oceanfront edges and wetlands that act as de facto growth boundaries around them.

3.1. Identifying unique metropolitan forms

Although metro areas of similar size generally tend to have similar spatial structure, some also stand out as unique. We use Mahalanobis distance (Appendix 1) to detect CSAs that are outliers based on their combination of five form metrics. Mahalanobis distance not only detects observations that are outliers according to one metric, but also those that have an extremely

---

[4] There may be other barriers to growth, including terrain, protected areas legal growth boundaries as discussed above.

different combination of measured outcomes.[5]

Outliers include Edwards-Glenwood Springs in Colorado, with its snake-like shape extended along a valley, extremely large linear metropolitan areas like Miami-Fort Lauderdale-Port St. Lucie and Seattle-Tacoma (Supp. Figure 5). The highly fragment and un-expandable regions of Washington-Baltimore-Arlington and Seattle-Tacoma, which are surrounded by wetlands, the enormous continuous region of Los Angeles, and Columbia-Moberly-Mexico CSA in Missouri, which contains three major built-up areas separated by agricultural lands, are also detected as outliers. Figure 1 below reports the standardized deviations from the mean in each of their metrics and supplemental Figure 5 illustrates the shapes of some of the outlier footprints (in black) and job centers (in red).

---

[5] Supplementary Table 2 shows the metropolitan form metrics for unique CSAs.
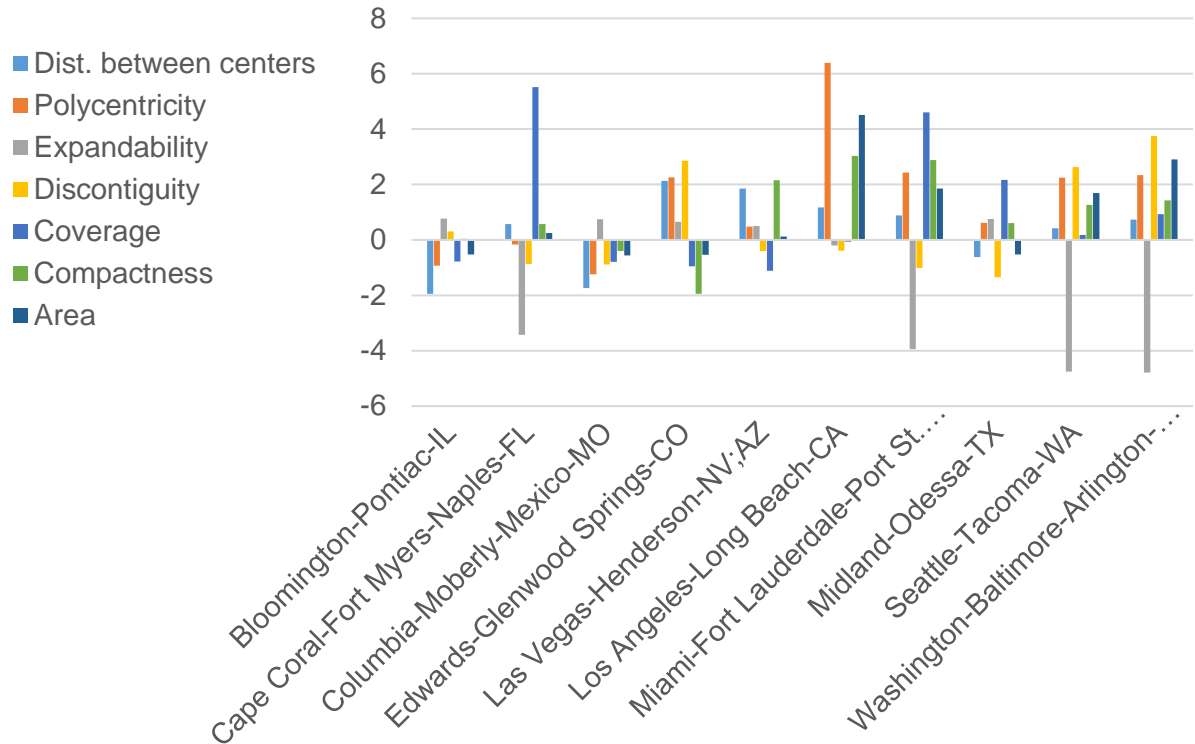
Figure 1. standardized deviation from the mean (centered at zero) for each metric for outlier metropolitan areas. While in some metropolitan areas no single metric deviates drastically from the mean, they are detected as outliers given their rare combination of measured outcomes (e.g. Columbia-Moberly-Mexico, Las Vegas-Henderson, Midland-Odessa and Bloomington-Pontiac).

We can see from Figure 1, for instance, that Los Angeles-Long Beach is an outlier due to three metrics that diverge beyond two standard deviations from the mean: 1) its exceptionally polycentric employment pattern, 2) its unusually high compactness, and 3) its unusually large area. The Seattle-Tacoma and the Washington-Baltimore-Arlington CSAs, on the other hand, are outliers due to their exceptionally low expandability (both are surrounded by water-bodies) and highly discontinuous built-up areas—a quality that is also attributable to the waterbodies that surround them.

4 Metropolitan form and transportation outcomes.

We used linear regressions to test how these metropolitan form metrics relate to transportation indicators that we use as proxies for sustainable metropolitan development—CO_2 emissions per capita, average travel time to work, and the percent of population that lives within walking distance (1km) from public transit. Though the benefits of denser urban environments remain subject to debate (Echenique et al. 2014), many planners have argued that minimizing sprawl and travel time to work, maximizing walking access to transit, and reducing CO_2 emissions per capita lead to environmentally more sustainable, less time-wasting and healthier urban environments (Frank and Engelke 2001; Glaeser 2012; Speck 2013).

Transportation-related CO_2 emissions were obtained from the Vulcan project, using 2016 data, which describes emissions in a10 by 10 km grid nationwide (Gurney et al 2009). Individual cell values were intersected with CSA built up areas to interpolate emissions to each CSA and divided by the population within the same built up polygons. Average commuting times to work were obtained from the 2016 American Community Survey, using 5-year estimates. Percent of residents living within a one kilometer (approx. 15-minute walk) from transit stations was estimated based on the US DOT Bureau of Transport Statistics National Transit Map, which includes commuter rail, heavy rail, light rail, monorail/automated guideway, streetcar rail and bus stations. This data is based on a General Transit Feed Specification (GTFS) requests made by US DOT to all transit agencies in the country, but the data is voluntarily provided and may

not represent the full picture.[6] We first computed the distance from each Census block in the US to its closest transit station, filtered out only those Census blocks that are within 1km from a station and then aggregated the population-weighted results up to CSA levels. Supplementary Table 3 presents the descriptive statistics.

The outcome variables suggested that the mean travel time to work is 21 minutes across all CSAs with a minimum of 14.4min in Clovis-Portales NM and a maximum of 32.6 in New York-Newark NY. The average percent of population with walking distance from transit across all 166 CSAs is 22%. This is an average of individual percentages, without weighting by size of CSA. The total percent of population across all CSAs that reside within 1km of stations is 46.6% and it rises to 53% if we only consider those 85 CSA that do have transit systems. The average resident in US metropolitan areas produces 5.9 tons of transportation-related CO_2 per year with, but the amount can be as high as 9-10 tons in some cases (e.g. Bowling Green-Glasgow KY and Columbia-Orangeburg-Newberry-SC) and as low as 3.4 tons per capita in the New York - Newark CSA.

In order to estimate the relationship between formal attributes and transport outcomes as elasticities, some variables had to be transformed. Transit access was included in raw form, since it was already measured as percentage. Average travel time and transportation-related CO_2 emissions were transformed with a natural log. Among the form metrics, coverage and area-doubling expandability were not transformed, since their raw form already reflects a percentage,

---

[6] The data is described at https://www.bts.gov/national-transit-map/about. Intersecting the stations with CSAs showed that roughly half (85 out of 166) of the CSAs had transit stations.

but polycentricity, fragmentation and gravity-compactness were transformed with a natural log to yield elasticities. We also include the log of CSA population as a control variable in the models, given that each of the four sustainability proxies are known to be strongly related to city size. Since 81 CSAs had no data on transit stations, only those cases with transit data were included in the corresponding model.

We first examined the relationship between each form metric to each outcome variable alone. These coefficients and R^2 statistics are reported in Table 1. We also included population and population density as sole predictors for comparison. The coefficients in these tests are uncontrolled and therefore should not be used for interpretation. Standard errors and t-statistics are reported to show how the fit of the estimated relationship varies across CSAs, but should not be interpreted as statistical significance, since the data is not a random sample, but rather reflects a full list of CSAs in the United States.

Some exploratory findings are worth noting. First, population density is often cited as a key metric for transportation planning and policy. A critical reader could also suggest that the form metrics introduced above may be equivalent or inferior to a much simple population density metric, whose computation only requires knowledge of the area and population of each CSA. When we examine the relationship between density and each of the three outcome variables in Table 1, we see that gravity-compactness consistently explains a greater share of the variation in each outcome than density. Variation in gravity-compactness alone explains 27.5% of metropolitan variation in transport-related CO2 emissions per capita, while density only explains 6.8%. Gravity-compactness explains 24.3% of variation in average commute times and 40.4% variation in transit accessibility, while the corresponding explanatory power of density is only

2.4% and 16%. Variations in coverage also provides a better explanation to CO2 emissions and average commute times than density, though density has a stronger correlation with transit accessibility. Fragmentation, polycentricity and area-doubling expandability too outperform density in their power to explain average commute times. All in all, using density as a convenient alternative to characterizing metropolitan form can be misleading.

Variable	1. Dependent = CO_2/pax — n = 163			2. Dependent = Avg travel time to work — n = 163			3. Dependent = % residents near transit — n = 85		
	Coeff.	std. error	t-value	Coeff.	std. error	t-value	Coeff.	std. error	t-value
Constant	9.455	(0.17)	(56.86)	1.906	(0.07)	(28.03)	-0.486	(0.27)	(-1.82)
Population	-0.060	(0.01)	(-4.79)	0.086	(0.01)	(16.91)	0.065	(0.02)	(3.44)
R² (adjusted)	0.117			0.633			0.114		
Constant	10.341	(0.46)	(22.28)	2.371	(0.30)	(7.86)	-2.088	(0.61)	(-3.42)
Pop. Density	-0.238	(0.07)	(-3.62)	0.097	(0.04)	(2.26)	0.355	(0.09)	(4.12)
R² (adjusted)	0.068			0.024			0.160		
Constant	9.057	(0.05)	(175.94)	2.815	(0.03)	(84.39)	-0.128	(0.08)	(-1.70)
Gravity-Compactness	-0.799	(0.10)	(-7.96)	0.478	(0.07)	(7.35)	0.997	(0.13)	(7.62)
R² (adjusted)	0.275			0.243			0.404		
Constant	8.765	(0.03)	(341.40)	2.953	(0.01)	(201.92)	0.363	(0.04)	(9.74)
Coverage	-1.017	(0.20)	(-5.00)	0.973	(0.12)	(8.40)	0.531	(0.24)	(2.17)
R² (adjusted)	0.127			0.296			0.042		
Constant	8.704	(0.05)	(173.40)	2.825	(0.03)	(109.23)	0.554	(0.07)	(7.40)
Fragmentation	-0.026	(0.03)	(-0.87)	0.142	(0.02)	(9.27)	-0.070	(0.04)	(-1.75)
R² (adjusted)	-0.001			0.340			0.024		
Constant	8.783	(0.05)	(179.27)	2.896	(0.03)	(98.18)	0.265	(0.07)	(3.81)
Relative Polycentricity	-0.100	(0.04)	(-2.57)	0.131	(0.02)	(5.57)	0.126	(0.05)	(2.49)
R² (adjusted)	0.034			0.157			0.059		
Constant	8.341	(0.10)	(85.19)	3.448	(0.06)	(61.55)	0.748	(0.11)	(6.79)
Area-doubling Expandability	0.366	(0.11)	(3.33)	-0.450	(0.06)	(-7.17)	-0.375	(0.13)	(-2.95)
R² (adjusted)	0.058			0.234			0.084		

Note: cell values show betas, with std. errors and t-values in parentheses. Significance level not reported since the dataset is not a sample, but reflects all CSAs.

Table 1. Bi-variate regressions showing the uncontrolled effects of each form metric alone on transport-related CO2 emissions per capita, average commute time and percent of population with walking access to transit. Models include population and population density as sole predictors are also included for comparison.

Table 2 reports the results of controlled models, where all form metrics as well as population of the metro were regressed together against the same three outcome variables.

Variable	1. Dependent = CO_2/pax			2. Dependent = Avg travel time to work.			3. Dependent = % residents near transit.			
	n = 163			n = 163			n = 85			
	Coeff.	Std. Error	t-value	Coeff.	Std. Error	t-value	Coeff.	Std. Error	t-value	VIF
Constant	8.482	(0.269)	(31.583)	2.032	(0.112)	(18.106)	0.222	(0.373)	(0.596)	
Population	0.083	(0.030)	(2.779)	0.087	(0.012)	(7.003)	-0.003	(0.040)	(-0.082)	7.329
Relative Polycentricity	-0.057	(0.039)	(-1.462)	-0.007	(0.016)	(-0.427)	0.028	(0.047)	(0.607)	1.397
Area-doubling Expandability	-0.108	(0.127)	(-0.849)	-0.103	(0.053)	(-1.937)	-0.177	(0.147)	(-1.201)	1.870
Fragmentation	-0.081	(0.041)	(-1.990)	0.027	(0.017)	(1.574)	-0.096	(0.050)	(-1.929)	2.739
Coverage	-0.670	(0.267)	(-2.511)	0.277	(0.111)	(2.483)	-0.413	(0.285)	(-1.452)	2.212
Gravity-Compactness	-1.136	(0.191)	(-5.932)	-0.231	(0.080)	(-2.894)	1.051	(0.250)	(4.208)	3.854
R^2 (adjusted)	0.314			0.712			0.456			

Note: cell values show betas, with std. errors and t-values in parentheses. Significance level not reported since the dataset is not a sample, but reflects all CSAs.

Table 2. OLS estimates of elasticities of metropolitan form metrics with respect to transport-related CO2 emissions, travel time to work and walking access to public transit. Highest variation inflation factor (VIF) is 8.3 for log of population.

Transportation-related CO2 emission per capita is related positively to population—larger CSAs have higher per capita emissions than smaller CSAs controlling for form characteristics—but the elasticity is very small. A one percent increase in CSAs population relates to only a 0.08% increase in per-capita emissions. This is nevertheless remarkable, given that a considerable body of urban science literature has theorized that more populated cities require less, not more, per-capita travel infrastructure and produce less per capital emissions (Glaeser 2012). Even though metro population indeed had a negative relationship to per-capita emissions when examined alone (Table 1), its sign flips when we control for metropolitan form characteristics (Table 2). Bettencourt et al. (2010), for instance, found that more populated cities produce less gasoline sales per capita, than less populated cities. Though our findings do not measure gasoline sales but CO2, the data here disagrees. It is not the size, but the form of metropolitan areas that

explains lower transport-related CO2 emissions per capita. It is natural to confound these effects, since larger metro areas also tend to be more compact.[7]

The largest form-related effect is observed with gravity-compactness—a 1% increase in compactness corresponds to a 1.136% decrease in emissions. An elasticity greater than one suggests that increasing metropolitan compactness is indeed an efficient way of reducing emissions.

Coverage has the second highest elasticity with respect to CO2 emissions—the more a CSA's convex hull is covered with development, the lower its emissions tend to be (elasticity = 0.67). Polycentricity, fragmentation and expandability are also negatively related to emissions, though their elasticities are close to zero. The six variables in the model explain 31% of variation in transport-related CO2 emissions per capita across US CSAs. Table 2 tells us that the majority of that explanatory power is attributable to compactness.

The same five form metrics plus population explain 71% of variation in average travel time to work in US metro areas. Commute times are negatively related to gravity-compactness (elasticity = -0.23)—more compact metropolitan areas have shorter commute times, controlling for covariates. Coverage is positively related to travel times, and its elasticity even exceeds that of compactness. A one percent increase in coverage is related to a 0.28% increase in travel times. Metro areas that have a larger share of their convex hull built out, have longer commute times.

---

[7] Compactness values reported in Supp. Table 4 show that the average gravity-compactness for CSAs whose population is above the 90th percentile is 1.1 (n=17). It drops to 0.6 for CSAs in the 40th-60th percentile population range (n=34), and to 0.4 for CSAs below the 10th percentile population range (n=17).

Again, the elasticities of other form variables are notably smaller. Polycentricity has a very small

negative elasticity and a low t-value. We are unable to conclude from this that more polycentric

metropolitan structure reduces travel times and helps alleviate traffic problems, which has often

been used as an argument in transport literature (Crane and Chatman 2003). For instance, Los

Angeles-Long Beach CSA—the most polycentric metro area in the US—has a mean commute

time of 27.7 minutes, which is above the 95th percentile of mean travel times. Even when LA is

excluded from the data due to its exceptionally high polycentricity, the effect does not change.

The third model suggests that access to public transit in American metropolitan areas is higher in

more compact metro areas. An elasticity greater than one is again worth highlighting here—a

one percent increase in compactness relates to a 1.05% increase in transit access. Other effects

are notably smaller and have lower t-values. Area-doubling expandability (elasticity -0.17) and

coverage (elasticity -0.41) are negatively related to transit access—metro areas whose expansion

is less constrained and those that cover more of their extents with development, have fewer

people living near transit. Other elasticities are too small to play a role. Population, which taken

alone exhibited a positive relationship to transit access in Table 1, again flipped signs to produce

a negative impact, suggesting that larger metro areas have fewer people near transit options,

when controlling for compactness and other form characteristics. A low t-value also signals

greater variation among CSAs. The model explains 45.6% of variation in transit access.

Overall, compactness stands out as the most important predictor with the highest elasticity in all

three models. How close people live to others in a metropolitan region appears to be the most

consequential variable in explaining carbon emissions and the proportion of people with walking

access to transit. Metropolitan regions that produce least pollution, have shortest commute times,

and where most residents have walking access to transit tend to have more compact forms. Coverage is the second most impactful factor, with notable elasticities as well. For average commute times, coverage has a higher elasticity than compactness.

5. Discussion

We presented five metrics to describe the forms of metropolitan areas and introduced improvements to their specification. The metrics were applied to all US CSAs, which enabled us to outline those metro areas that stand out with unique formal properties. We also analyzed relationships between the metropolitan form metrics and three different sustainable transport indicators—transport-related CO_2 emissions, commute times and pedestrian access to public transport. Gravity-compactness, which describes how close people live with respect to all other people in the metro area, was the most consequential predictor of CO_2 emissions and access to transit. Our findings suggest that the gravity-compactness metric overcomes some of the known shortcomings of the simpler population density measures often used in practice.[8] Given a number of desirable outcomes of more compact metropolitan areas—less CO_2, shorter commute times, and higher levels of population near transit—we find that the metric is well-suited to be used in metropolitan planning and policy making.

---

[8] The US Census Bureau has recently shifted to including weighted population density measures in addition to raw population density measures in their data (US Census Bureau, 2012). While a weighted density measure accounts for variations within population distribution, a key advantage of the proposed compactness measure lies in the fact that it captures internal variations in population distributions as well as distances between built-out regions within the same metropolitan area, thereby providing a more accurate and context-sensitive depiction of metropolitan configuration.

Coverage and fragmentation metrics can be used to describe spatial fragmentation of metropolitan areas. While coverage describes the percent of buildable land in a convex hull that is actually built out, irrespective of shape, fragmentation specifically captures how disjointed the development pattern is. Both metrics are positively related to average commute times to work. Controlling for covariates, metro areas with a higher land cover, as well as those with more fragmented development, tended to have longer commute times to work. At the same time, more compact and more polycentric metropolitan areas tended to reduce commute times and enable more people to gain access to transit within walking distance.

Contrary to some existing literature, our analysis showed that per-capita CO_2 emissions are not necessarily smaller in bigger CSAs. Controlling for form characteristics, metropolitan population is actually positively related to CO_2, though the effect is very small. Rather than the size of a metro area, more compact form is are key to explaining lower emissions.

An interesting metropolitan form typology that is highly discontinuous, but at the same time compact is presented by Washington-Baltimore-Arlington, Seattle-Tacoma and Miami-Fort Lauderdale-Port Lucia CSAs (Supplemental Figure 5). Each of these three metropolitan areas is highly constrained by water bodies around their edges (low expandability). Each of them also contains a rail transit system that has become a magnet for densification and growth. Their forms are compact and polycentric and at least partially concentrated around public transport, similar to the well-known patterns of Copenhagen and Stockholm in Europe. Despite their low coverage and highly fragmented development patterns, these CSAs demonstrate that it is possible to achieve lower than average transport-related CO_2 emissions with highly discontinuous, but compact growth around polycentric rail corridors.

The open-source GIS toolbox accompanying this paper enables researchers to automate and customize the calculation of shape metrics on metropolitan areas elsewhere. Future work could examine how effective different policy and planning mechanisms have been in guiding the formal pattern of metropolitan areas. The effects of policies, such as urban growth boundaries, gasoline taxes, large-scale transit investments or form-based development codes can be examined when metropolitan shape metrics are compared across the same metropolitan areas over multiple decades. Longitudinal data for such comparison has been recently made available by the World Bank, covering at least three decades (1990, 2000 and 2010) world-wide. Having scalable metrics to describe metropolitan form offers an important step towards quantifying the effects of past, present and future policy changes in different urban regions.

References

Alonso, W. (1964). Location and land use. *Toward a general theory of land rent*. Harvard University Press.

Angel, S., Sheppard, S., Civco, D. L., Buckley, R., Chabaeva, A., Gitlin, L. & Perlin, M. (2005). *The dynamics of global urban expansion*. Washington, DC: World Bank, Transport and Urban Development Department.

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

Bertaud, A., & Malpezzi, S. (2003). *The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition*. Madison, WI.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kuhnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences April 24*, *104*(17), 7301–7306.

Boyce, R. B., & Clark, W. A. V. (1964). The concept of Shape in Geography. *Geographical Review*.54(4), 561– 572.

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

Catenaccio, G. (2010). Urban Growth Boundaries: Two American Examples. *Projections, the MIT Journal of Planning.*, *issue 10*, 31–51.
https://dusp.mit.edu/sites/dusp.mit.edu/files/attachments/project/projections_10_a.pdf

Cleveland, W. S. (1979) Robust locally-weighted regression and smoothing scatterplots, Journal of the American Statistical Association, 74, pp. 829–836.

Clifton, K., Ewing, R., Knaap, G. J., & Song, Y. (2008). Quantitative analysis of urban form: a multidisciplinary review. *Journal of Urbanism*, 1(1), 17-45.

Cox, W. (2011). Constraints on housing supply: Natural and regulatory. *Econ Journal Watch*, 8(1), 13-27.

Crane, R., & Chatman, D. G. (2003). As Jobs Sprawl, Whither the Commute? *Access Magazine*, (Fall), 14–19.

https://cloudfront.escholarship.org/dist/prd/content/qt4863x8g5/qt4863x8g5.pdf

Echenique, M. H., Hargreaves, A. J., Mitchell, G., & Namdeo, A. (2014). Growing Cities Sustainably. *Journal of the American Planning Association*, *78*(2), 121–137.

Frank, L. D., & Engelke, P. O. (2001). The built environment and human activity patterns: exploring the impacts of urban form on public health. *Journal of Planning Literature*, *16*(2), 202-218.

Garcia-López, M. À., & Muñiz, I. (2010). Employment decentralisation: Polycentricity or scatteration? The case of Barcelona. *Urban Studies*, 47(14), 3035-3056.

Gibbs, J. (1961). A Method for Comparing the Spatial Shapes of Urban Units. In J. Gibbs (Ed.), *Urban Research Methods*. Princeton: Van Nostrand Co.

Giuliano, G., & Small, K. A. (1991). Subcenters in the Los Angeles region. *Regional science and urban economics*, 21(2), 163-182.

Glaeser, E. (2012). *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. The Penguin Press HC.

Glaeser, E. L., & Kahn, M. E. (2001). *Decentralized Employment and the Transformation of the American City* (NBER Working Paper No. 8117). https://www.nber.org/papers/w8117

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S., & de la Rue du Can, S. (2009). The Vulcan Project: High Resolution Fossil Fuel Combustion CO2 Emission Fluxes for the United States. *Environmental Science & Technology*, *43*(14), 5535–5541.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, 25(2), 73-76.

Hopkins, L. D., & Knaap, G. J. (2000). Portland, Oregon: An inventory approach and its implications for database design. In Vernez-Moudon A. & Hubner M. (Eds) *Monitoring Land Supply with Geographic Information Systems; Theory, Practice, and Parcel-Based Approaches*. Wiley, 67-88.

Jackson, L. E. (2003). The relationship of urban design to human health and condition. *Landscape and urban planning*, 64(4), 191-200.

Limtanakool, N., Schwanen, T., & Dijst M. (2009). Developments in the Dutch urban systems on the basis of flows. *Regional Studies*, 43(2), 179–196.

MacEachren, A. M. (1985). Compactness of geographic shape: Comparison and evaluation of measures. *Geografiska Annaler. Series B. Human Geography*, 53-67.

McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2), 242-258.

McDonald, J. F. and McMillen, D. P. (1990) Employment subcenters and land values in a polycentric urban area: the case of Chicago, Environment and Planning A, 22, pp. 1561–1574. https://doi.org/10.1068/a221561

McDonald, J. F., & Prather, P. J. (1994). Suburban employment centres: The case of Chicago. *Urban studies*, 31(2), 201-218.

McGarigal, K., & Marks, B. J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure.

McMillen, D. P., & McDonald, J. F. (1997). A nonparametric analysis of employment density in a polycentric city. *Journal of Regional Science*, 37(4), 591-612.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 197-210.

MRLC (2011). *National Land Cover Database 2011 (NLCD 2011).* https://www.mrlc.gov/tools

Muth, R. F. (1969*). Cities and Housing; the Spatial Pattern of Urban Residential Land Use*.

Parent, J., Civco, D., & Angel, S. (2009). Shape Merics, Center for Land Use Education and Research. http://clear.uconn.edu/tools/Shape_Metrics

Prosperi, D., Moudon, A.V., & Claessens, F. (2009) The question of metropolitan form: An introduction. *Footprint*, Delft Architecture Theory Journal, 5, 1-4.

Roca Cladera, J., Marmolejo Duarte, C. R., & Moix, M. (2009). Urban structure and polycentrism: Towards a redefinition of the sub-centre concept. *Urban Studies*, *46*(13), 2841-2868.

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

Sauer, C. O. (1925). *The morphology of landscape*. University of California Publications in Geography 2, 19-54

Speck, J. (2013). Walkable City. *How downtown can save America one step at a time*. North Point Press.

Song, Y., Merlin, L., & Rodriguez, D. (2013). Comparing measures of urban land use mix. *Computers, Environment and Urban Systems*, *42*, 1-13.

Tsai, Y. (2004). Quantifying Urban Form: Compactness versus "Sprawl." *Urban Studies*, *vol 42*(issue 1).

US Census Bureau. (2012). *Patterns of Metropolitan and Micropolitan Population Change: 2000 to 2010. 2010 Census Special Reports.* Washington, D.C.

Victoria Transport Policy Institute (2015). *Analysis of Public Policies That Unintentionally Encourage and Subsidize Urban Sprawl.* http://static.newclimateeconomy.report/wp-content/uploads/2015/03/public-policies-encourage-sprawl-nce-report.pdf

Wentz, E. A. (1997). Shape Analysis in GIS. *Auto Carto* 13,7-10.

Wheeler, S. M. (2008). The Evolution of Built Landscapes in Metropolitan Regions. *Journal of Planning Education and Research*, 27(4), 400–416.

Witters, D. (2011). *More Than 15% Obese in Nearly* All U.S. Metro Areas. Gallup.

Wu, B., Yu, B., Wu, Q., Chen, Z., Yao, S., Huang, Y., & Wu, J. (2018). An Extended Minimum Spanning Tree method for characterizing local urban patterns. *International Journal of Geographical Information Science*, *32*(3), 450–475.