WORKING PAPER SERIES

Title

Fragmented by Nature: Metropolitan Geography, Urban Connectivity, and Environmental Outcomes

Authors

Luyao Wang

Albert Saiz

Weipeng Li

Date

May 2024

Center for Real Estate

MIT Center for Real Estate Research Paper No.

24/03

MIT Center for Real Estate

Fragmented by Nature: Metropolitan Geography, Urban Connectivity, and Environmental Outcomes
Luyao Wang, Albert Saiz, Weipeng Li
MIT Center for Real Estate Research Paper No. 24/03
May 2024

Abstract

Physical geography has long been identified as critical for urban development, land use, and environmental outcomes in cities worldwide. However, the literature has yet to provide comprehensive, quantitative analyses of the global extent and impact of urban geographic barriers. Our study introduces three novel indexes: the *share of natural barriers, non-convexity* (a measure of natural fragmentation), and the *average road detour*, to measure and study the practical reach and effects of natural barriers around global cities. We calculate these indexes for areas in and around four separate global city-boundary definitions, augmenting the original data with relevant additional variables. We find that natural barriers lead to more complex transportation environments and are associated with higher urban densities, smaller urbanized footprints, taller buildings, and less pollution, but also with lower incomes and smaller populations. To draw meaningful policy conclusions, comparative research about environmental, economic, and social outcomes across global cities should always account for their surrounding geographies.

Luyao Wang
Massachusetts Institute of Technology & Wuhan University

Albert Saiz
Massachusetts Institute of Technology

Weipeng Li
College of Information and Communication

1 Introduction

Physical geography has long been acknowledged as a critical factor shaping urban development, land use, and socio-economic outcomes worldwide [1, 2]. Cities have always been influenced by their geographic context, often developing around natural resources, such as water bodies, fertile land, or mineral deposits [3, 4]. Topographical features, such as mountains, coastlines, and rivers, partially determine the expansion and land use patterns of cities [5, 6]. With concerns about urban sustainability and the need for efficient land use in the face of rapid urbanization, understanding the impact of geographic barriers in urban centers is more critical than ever [7–9].

A scientific literature studying comparative global city expansion has emerged, due to the increased availability and consistency of data [7, 10]. For instance, [11] measure the density and fragmentation of 35 urban areas worldwide and [12] similarly analyze 1,570 global metropolitan areas. Focusing on the land area of a sample of 90 international major urban centers, [13] find that cities with larger populations, high income, more underground water reserves, and low agricultural soil prices expanded further. Other work has addressed the global interaction between transportation systems and urban density [14–16], or between zoning and tall buildings [17].

Extant research has also made some progress shedding light on the interaction between the physical environment and human behavior. It is well-known that bedrock soils [18] and seismic activity [19] affect construction costs and the practicability of tall buildings. These geographic factors have important bearing on local economic activity and employment [20, 21]. [22] show that the availability of aquifers and terrain ruggedness are determinants of suburban sprawl in 275 US metropolitan areas.

[23] demonstrates that geographic barriers affect the supply of housing in the US. Constrained areas display higher housing prices and slower population growth. Using satellite data and historical maps, [24] shows that complex urban shapes reduce urban productivity and population density in India, also increasing commuting costs and lowering land values. These findings are replicated for Latin America by [25]. The study of urban morphology—which arises from a combination of geographic, historical, economic, transportation, and urban planning factors—also delves into the measurement of the shape of cities and their growth [26, 27]. In that vein, [28] show that the *ex post* shapes of the urban footprints within 200 large urban settlements are conditioned by topographic characteristics.

[29] analyze the micro-effects of natural barriers using data on real-time car trips within the Boston metro area. They find that natural barriers generate traffic delays and worsen downtown congestion: removing such barriers would shrink average travel times by 17 percent and city-center traffic by 13 percent. Therefore, policies that eliminate bottlenecks—such as investments on bridges and tunnels—improve urban efficiency above and beyond the trips that are directly benefited. [30] show that the presence of hills and water bodies reduce average commuting speeds across cities in India, and [31] reproduce these findings across global cities.

Despite the progress, there is still relatively scarce research providing holistic, global, and comparative understanding of the role of natural barriers on urban outcomes. This may be attributable to the challenges associated with obtaining and

processing large-scale spatial data, or to the inherent complexities in comparing diverse geographic contexts [32, 33].

For instance consider Figure 1, mapping Barcelona (Spain, panel A) and Atlanta (USA, panel B). Both metro areas had similar population in 1990 but have been often contrasted as polar opposites in terms of sustainability: Atlanta's built-up area was 28 time larger than Barcelona's [34, 35]. Urban planning, car-dependence, and history account for most of the differences. However, their surrounding geographies may also have played a role, with Barcelona sandwiched between mountains and the Mediterranean. Interestingly, the city of Sant Cugat—in the northeast of Barcelona, but separated by the Collserola mountain range—displays much lower densities: 2,023 people per square kilometer, as opposed to 16,339 in Barcelona proper. Another interesting comparison is with the city of Madrid (Spain, not pictured)—with no substantial physical constraints—at 5,514 people/km^2.

Oslo (Norway, panel C) and Helsinki (Finland, panel D) provide another suggesting contrast. While both areas suffer from natural fragmentation, Oslo displays a *non-convexity* index (explained below) that is about 80% larger. Whereas the population of Finland (5.541 million in 2023) is only 2.5% larger than Norway's (5.408 million), Helsinki's urban center's population (911,950) is 13% larger than Oslo's (807,340). Nonetheless, Oslo is much denser at 3,336 people/km^2, compared to 2,651 in Helsinki.

Are these examples illustrative of the impact of natural geography on urbanization? In this paper we conclusively show that more fragmented geographic landscapes are indeed associated with denser, smaller, greener, and less land-hungry cities around the globe. Nonetheless, they also tend to spawn poorer urban areas and to display road connectivity challenges. Hence, we argue that any comparison of social, economic, and environmental outcomes between cities needs to control for the impact of natural geography. Otherwise, researchers and policymakers may fully attribute differences across cities—for instance in energy use—to policy or human behavior, rather than to natural constraints.

In this study, we introduce and calculate three measures of geographic barriers surrounding urban centers—represented by bodies of water, steep slopes, and national boundaries. We do so for 37,675 urban units in four global datasets: our update of the Global Human Settlement Layer Urban Centres Database (GHSL-2023-UCDB)[36, 37], the Global Hierarchical Urban Boundaries (GHUB)[38], the Global Urban Boundaries (GUB)[39], and the Atlas of Urban Expansion (AUE)[40]. These datasets cover the two most commonly used definitions of cities: population-based (UCDB) or building-based (GHUB, AUE, and GUB). We also compute socio-economic attributes within each urban unit from a variety of external data sources and update several attributes (population, build-up area, income per capita, etc.) for the UCDB using the latest releases of global gridded data [41]. A detailed description of the four datasets is presented in Supplementary Table S1, S2 and Figure S1.

While hundreds of social, economic, and policy variables affect urban growth, topographical features capture truly exogenous or primordial drivers. The current size, density, shape, and most other attributes of cities are outcomes, or endogenous variables. A way to frame the questions we are interested in is by thinking about the full ex-ante distribution of potentially urbanizable environments, prior to the explosion of

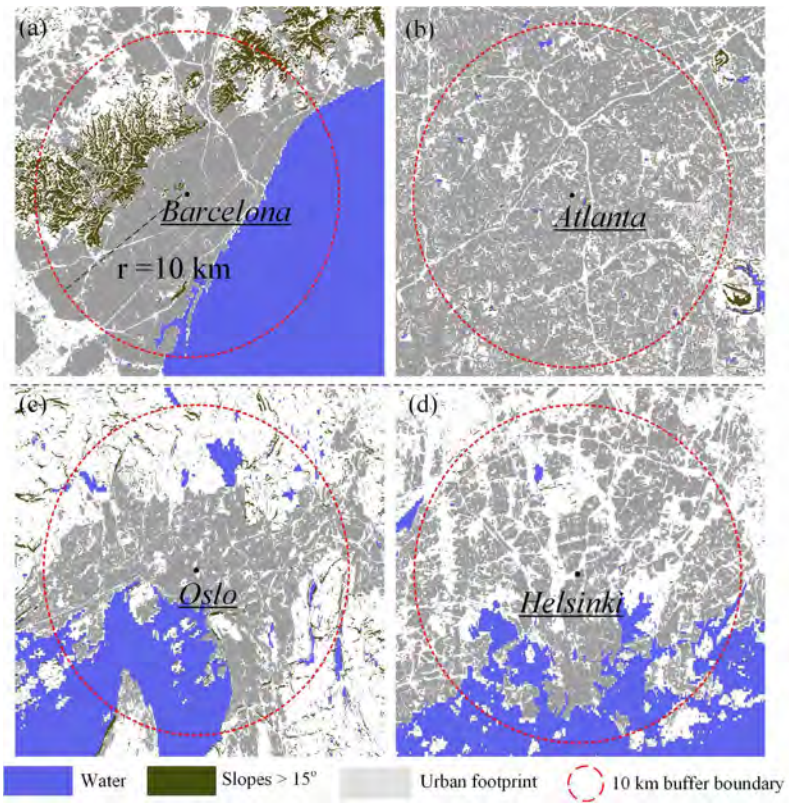

Fig. 1 Urban built-up area and physical barriers in Barcelona (Spain) **(A)**, Atlanta (USA)**(B)**, Oslo (Norway) **(C)** and Helsinki (Finland) **(D)**

human population in the past two centuries. Conditional on gathering the minimum population to qualify as urban: which ex-ante environments have tended to spawn larger, denser, richer, taller, and more impervious-surfaced cities today? In order to answer this question, we require an exogenous and comparable definition of the environment around each potential city. In our research, we endeavor to provide such measurement and answers to the questions above.

2 Measurement

A strand of research engages in the measurement of city shapes (e.g. [42]). [43] summarize these metrics, which are critical to understanding urban geometry, patterns of sprawl, and city-specific paths for further expansion. However, our approach must differ, as current city shapes and sizes are endogenous outcomes arising *ex post* from factors such as those under study. Statistical inference would not be valid if we adjusted our measurements to their cities' current population or density/area, as these are precisely the outcomes that we want to explain (see Supplementary Section 1).

Consequently, we compute measurements around fixed radiuses from the urban centroid. These provide homogeneous, comparable, ex-ante metrics for the environment surrounding *potentially-substantial* urban entities, independently of their ex-post population. We focus on 5 and 10 km radiuses, which encompass areas larger than most global cities (Supplementary Table S3). However, results are not sensitive to the choice of radius, as the correlation in the measurements across them is always very high (Supplementary Table S7, S8).

Our first metric—the *share of barriers*—reframes and expands previous measures [23, 44] to most significant global cities. It captures the overall share of the area within 10 km of the city center impeding urban expansion due to internal bodies of water, oceans, international borders, or steep-sloped terrain—at or above a 15-degree inclination (see *Methods* and Supplementary Figure S4).

Formal housing can hardly be built cost-effectively at slopes at or above 15 degrees. Terracing the landscape is possible, but very costly and associated with risks of landslides. [23] finds land with slopes above 15 degrees encompassed 47.62 percent of the land in Los Angeles, but only 3.65 percent of its housing. [28] also chose a 15-degree threshold arguing that *"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 solution."* According to these authors 94.1 percent of the development in 200 large global cities was in areas with slopes below 15 degrees.

A few formal homes or—likelier—slums are built on very steep areas in Caracas, Medellín, Rio, Lima, and other cities. Conditions in these settlements are poor: slums are precisely an endogenous outcome of formal developers shunning these areas. [45] study slum formation in extremely steep-sloped terrains, which they situate at above 14 degrees. They find that—even considering informal housing—building at slopes above 14 percent is rare. Adding water areas—with an almost zero percent of people living there—makes the *share of barriers*, as defined in this paper, a good index for the area that is very unlikely to be urbanized.

We also provide the reader with this variable's counterpart at 5-km radii, but the two measurements are very highly correlated. Both variables capture the physical availability of land for potential urban development in and around global city centers. Note that access to oceans and bodies of water also confers economic advantages in pisciculture and navigation. To account for these, we also control for coastal proximity in the regressions below.

Our second measure—the cities' *average dyadic nonconvexity*—focuses on barriers between locations in and around urban areas. For a geographically representative sample of points—uniformly scattered around 10 km of the city center—we calculate the average share of the straight-line segment between them that falls within a natural barrier, thereby impeding their connectivity. These ratios capture local *non-convexity*: in topology, a convex set is such that all segments between any two points are fully contained within it (see *Methods* and Supplementary Figure S5.). We also provide this metric at 5 km radii from global city centers.

This measure is calculated for each origin- and destination-point dyad—from the bottom up. It can therefore be disaggregated over any subset of the landscape continuum, still carrying a consistent interpretation [29]. Other measures of shape become meaningless as one segments the original polygon into smaller components.

Regarding the relationship between the first two measurements, note that natural barriers are necessary but not sufficient to generate local nonconvexities. For instance, a hypothetical circular, flat island of radius 5km would display a *share of barriers*—as measured around 10 km of its center—of 0.75 $\left(1 - \frac{\pi 5^2}{\pi 10^2}\right)$. However, all potential trips between inner points would be unimpeded, resulting in an average local *nonconvexity* of zero. Therefore, barriers must be present *and* effectively fragment flat land in order to compute into the second metric.

Finally, a third metric relates to the road trips across potential origin-destination dyads within a radius around the city center. This variable is proportional to the average *detour* factor (see *Methods* and Supplementary Figure S6): the ratio between the *minimum* road trip distance permitted by the current street network and the straight-line distance between points. The minimum mileage between destinations captures the impact of local topography, but also conscious policy decisions about road system design and budgeting—such as the density of the street network and investments in tunnels and bridges. The final measure averages road *detour* factors between all origin-destination dyads across a uniformly distributed grid of points around the city center—within 5 and 10 km radii.

We adopt these three new global urban indexes to examine the role of geographic barriers on connectivity and city-level outcomes. Regressions below are shown at 10-kilometer radii from urban centers, but available and unchanged at other radii (see Supplementary Section 4—Robustness checks).

3 Results

3.1 Urban Natural Barriers and Nonconvexity Worldwide

Figure 2A displays the spatial distribution of natural constraints in and around cities in the UCDB dataset (other datasets shown in Supplementary Figure S7-S10). Cities with high *share of barriers* are predominantly situated in Central, East, and Southeast Asia, and Southern Europe. Populous Asian countries, such as India and China, contain the largest number of cities with surrounding high *shares of barriers* (157 cities in India and 137 cities in China), followed by Indonesia with 128. In contrast, only 28 cities in the United States and 26 cities in the United Kingdom are highly constrained.

Most cities impacted by geographic barriers lay near the ocean (Figure 2B), such as New York City, Tokyo [46], or the coastal zones in China, India, and Indonesia. However, many cities in the east coast of Latin America are less impacted, due to their location slightly inland or in inner harbors. There are 1,181 cities—such as Kabul, Bishkek, and Dushanbe—that are corseted in by large mountain systems, more frequently in Central Asia, Eastern South America, and East Africa.

The expansion of some cities is also impeded by international borders—as in Tijuana, Haparanda, or Geneva [47]. Border cities are mainly in the Middle East, the

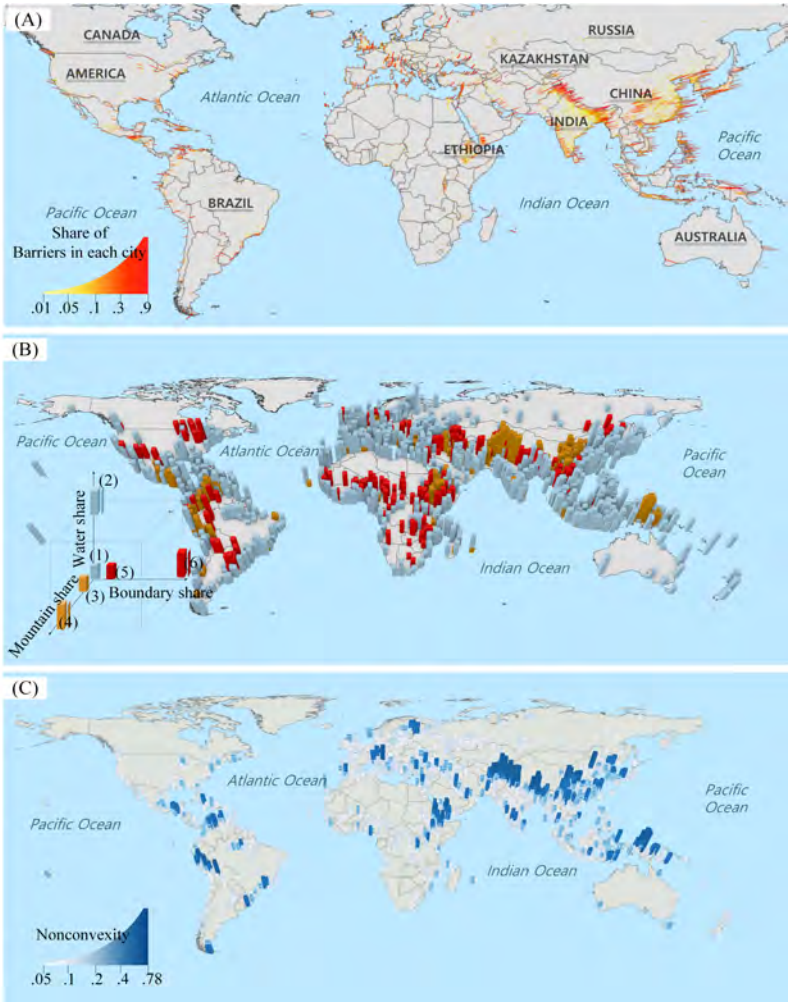

Fig. 2 **(A)** The global *share of barriers* around cities in the GHS-UCDB boundaries. Color and height of the bar represent the value of the index. **(B)** Global cities by dominant component of barriers: (1) water dominant at low (<0.3) shares, (2) water dominated at high (≥0.3) shares, (3) mountains dominant at low shares, (4) mountain dominants at high shares, (5) international borders dominant at low shares, (6) international border dominant at high shares. **(C)** Global cities with major *nonconvexities* (cities with *nonconvexity* < 0.05 are undisplayed).

Northeastern United States and—disproportionally—in the landlocked countries of Central and Southern Africa. A few cities suffer from the impacts of the three kinds of barriers, like Lugano, in Switzerland (Supplementary Figure S3).

The *share of barriers* index, however, masks variation in the morphology of natural geographic features—as evidenced in Supplementary Figure S4. Consequently, we next address the *average dyadic nonconvexity* index (Supplementary Figure S5).

Figure 3C reveals that cities with high geographic *non-convexity* are predominantly situated in South-Central and Southeast Asia—primarily in the Central and

South-East coastal areas of China, the Eastern periphery of India, and the Southern Himalayan Range. While most African cities enjoy relatively benign geographic conditions, Ethiopia is an exception, with 79 cities displaying high *nonconvexity* indexes. A similar pattern is observable around a handful of cities in the western region of South America and the southern sector of North America. However, an anomaly lies in India's coastal cities, which present a high *share of barriers* yet low *nonconvexity* indexes, suggesting that their landmasses remain relatively unfragmented.

3.2 Global North versus South

The data shows that the average *share of barriers* around Global North cities (0.12) is higher than around Global South cities (0.091). However, the average *nonconvexity* around cities in the South (0.039) is marginally higher than in the North (0.036), as depicted in Figure 3A, 3C. This is also evident in the GHUB and GUB datasets (see Supplementary Figure S11, S12).

While the *nonconvexity* index allows us to gauge land connectivity in and around cities, it does not capture how the transportation system amplifies or mutes natural fragmentation. The *detour* index quantifies road connectivity, inclusive of bridges and tunnels built to counteract physical impediments. Using UCDB delineations, Global South urban environments display 24% longer detours (Global South: 0.78, Global North: 0.63), as seen in Figure 3B. A similar trend is also observed in the GHUB and GUB datasets (see Supplementary Figure S13, S14). Countries with very high *detours* (>0.8) are predominantly located in Africa, East and Southeast Asia, and South America (Figure 3E).

3.3 Natural barriers and road connectivity in and around cities

We now examine the associations between geographic barriers and city outcomes. We use data from the UCDB [36] or from overlays of the 2023 GHSL [41], and other gridded global datasets computed to UCDB boundaries (details in Supplementary Table S2, S4). Moreover, we incorporate country-level data sourced from the World Development Indicators (WDI) database [48] as well as information from the Freedom House index [49], where less democratic countries obtain higher "autocracy" scores.

Our first set of regressions (Table 1) set the average street *detour* between points in and around the cities as the dependent variable. In Column 1, the only explanatory variable is the *share of barriers* within a 10 km radius of the city center (for results on 5 km radiuses see Supplementary Table S7). Geographic constraints have an impact on road distances between points. However, barriers that generate nonconvexities effect minimum driving distances much more strongly (column 2)—the regression's adjusted R-squared nearly triples, from 0.059 to 0.167. Of course, mountains and water bodies have *additional* impacts on trip speed and congestion beyond the effects on the design of the road system that we find here [29].

Columns 3 and 4 saturate the model with controls. To capture social, economic, cultural, and policy differences across nations we include 156 country fixed effects—dropping singleton city-and-country data points. Therefore, we now focus on differences in *detour* factors across cities *within* each country. We account exhaustively

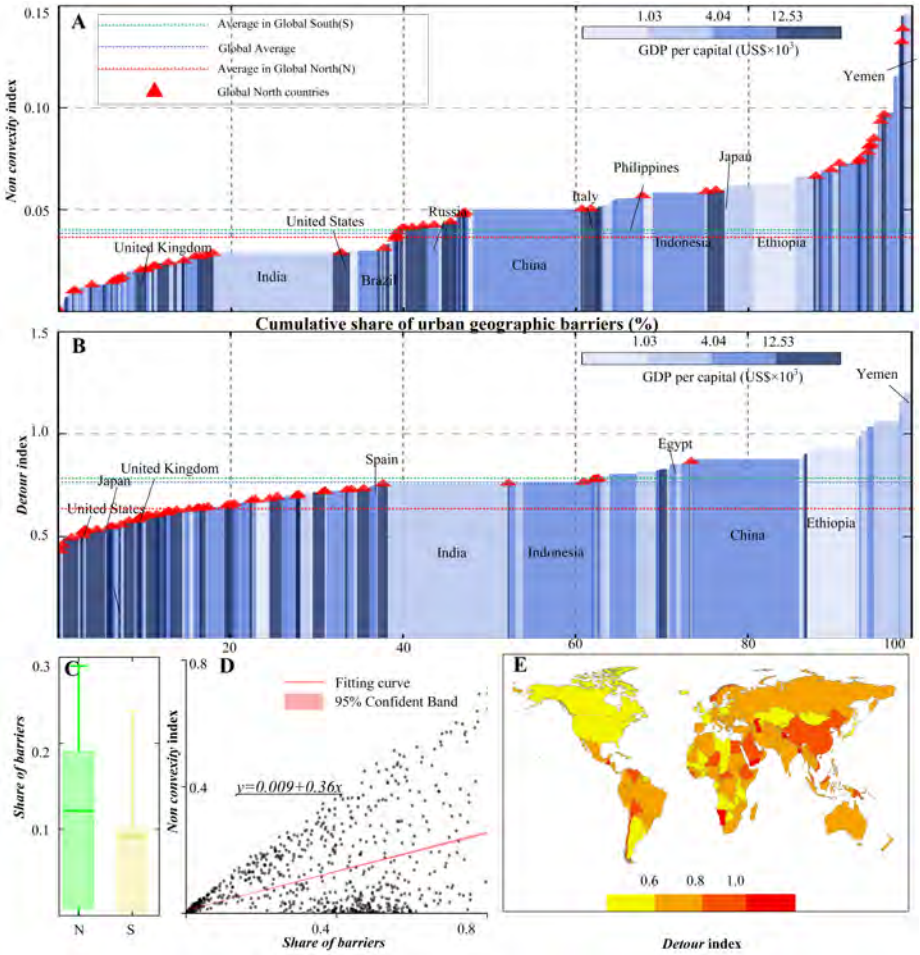

Fig. 3 **(A)** Average *nonconvexity* index by country. The dashed lines in red, green, and blue colors denote the average *nonconvexity* index in Global North countries (0.036), Global South countries (0.039) and all countries (0.038), respectively. The bar width represents the portion of urban geographic barriers of each country to the global total. Global North countries are marked with red triangles. **(B)** The average *detour* index of each country. The dashed lines with red, green, and blue colors denote the average *detour* index of Global North countries (0.63), Global South countries (0.78) and all countries (0.76), respectively. **(C)** Comparison of the average urban *share of barriers* in Global North and South countries (N: Global North, S: Global South). **(D)** Relationship between the average urban *nonconvexity* index by country and the corresponding *share of barriers*. **(E)** Distribution of the *detour* index across global countries.

for other locational fundamentals including: 34 dummy variables for soil characteristics, which may affect paving costs; 14 dummies for climate type; 16 dummies for biome taxonomies, capturing among others crop fertility and, hence, relative agricultural land costs; a dummy for coastal cities; a dummy for cities that are not in riverine watershed areas; a dummy for capital cities; and the logs of temperature, precipitation, and average elevation; finally, the logs of GDP per capita, population, and built-up

Table 1 Impacts of geographic barriers on street/road patterns

	Detour Index within 10 km of City Center					
	(1)	(2)	(3)	(4)	(5)	(6)
Average Local Nonconvexity within 10km of City Center		2.540***		2.005***	1.968***	2.004***
		(0.0495)		(0.0585)	(0.0722)	(0.0618)
Share of barriers within 10km of City Center	0.682***		0.600***		0.0309	
	(0.0238)		(0.0297)		(0.0356)	
Dummy for Ocean or Lake Intersection			0.0385***	0.0317***	0.0280***	0.0235**
			(0.0110)	(0.00985)	(0.0107)	(0.0106)
Dummy of Capital City			0.0994***	0.0783**	0.0789***	0.0619*
			(0.0315)	(0.0306)	(0.0306)	(0.0323)
City not in Major River Watershed			-0.0647***	-0.0490***	-0.0499***	-0.0589***
			(0.0093)	(0.0090)	(0.0091)	(0.0091)
Log of Average Precipitations in 2015			0.0287***	0.0116	0.0119	0.0074
			(0.0083)	(0.008)	(0.008)	(0.0076)
Log of Average Temperature in 2015			-0.0423	-0.0132	-0.0109	-0.0435
			(0.0345)	(0.0334)	(0.0335)	(0.0327)
Log Average Elevation			0.0177***	0.0051	0.0054	0.0004
			(0.0034)	(0.0033)	(0.0034)	(0.0034)
Log of GDP per Capita			-0.0026	-0.0016	-0.0017	0.0027
			(0.0029)	(0.0028)	(0.0028)	(0.0029)
Log of Total Resident Population in 2015			0.0023	-0.0215***	-0.0213***	-0.0042
			(0.0074)	(0.0072)	(0.0072)	(0.0071)
Log of Build-up Area within Urban Unit			-0.0844***	-0.0588***	-0.0588***	-0.0787***
			(0.0068)	(0.0067)	(0.0067)	(0.0066)
COUNTRY: Urban population (% of total population)						0.0004
						(0.0004)
COUNTRY: Average Freedom House index from 1973-2015						0.0277***
						(0.0031)
COUNTRY: Population ages 0-14 (% of total population)						-0.0034***
						(0.001)
COUNTRY: Literacy rate (% of adult people)						-0.0022***
						(0.0004)
COUNTRY: Government expenditure (% of GDP)						0.0003
						(0.0011)
COUNTRY: Log of Country's GDP per Capita in 2015						0.0457***
						(0.0106)
Country Fixed Effects	NO	NO	YES	YES	YES	NO
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES	YES
Observations	13,093	13,093	12,449	12,449	12,449	11,116
Adjusted R-squared	0.059	0.167	0.237	0.281	0.281	0.253

Note: In column 6, variables prefaced by "COUNTRY: " are country-level indicators derived from WDI and Freedom House.
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

land area in each city are also included. The results are in line with the earlier ones. Constraints to the geographic expansion of cities reduce road connectivity much more wherever they generate nonconvexities.

Remember that, while related, the two right-hand-side geographic metrics are different. For instance, Tripoli (Libya) is highly land-constrained—with 50.7% of the area around its downtown "lost" to the Mediterranean. However, the city is sited at the peak of a triangle of land bulging into the sea thereby displaying a low average inner

nonconvexity (0.04). When we include both variables (column 5), mostly the latter matters.

We can draw other ancillary conclusions from our baseline specification in Table 1, column 4. Dry places, with no rivers and low precipitation tend to feature shorter-distance road trips: whereas rivers may not generate large nonconvexities they necessitate of bridges, thereby inducing detours from straight-line itineraries. Cities at high elevations also display higher average route detours. Naturally, detours are shorter as cities become more populated and occupy larger areas, because denser street networks are bound to follow from urban expansion.

Column 6 explores the data variation between nations. We now eliminate the country fixed effects, but account for the covariance within nations by clustering standard errors at that level. This allows us to control for several national characteristics. Road distances tend to be lower in high income countries. These results are consistent with those in [31], who use a sample of 1,190 major cities and a different outcome: driving speed as ascertained by Google Maps' algorithms. However, they may also partially capture poorer coverage of OSM in developing countries.

Two intriguing new results arise: more literate and, especially, more democratic countries—characterized by low autocracy "Freedom House" indexes—tend to have denser city and peri-urban street networks. This is consistent with the idea that democratic governments cater better to the needs of their citizens. In contrast, previous evidence has shown that dictatorships tend to invest more in national rapid highways connecting different parts of the country, with much clearer military value [50].

3.4 Natural Barriers: Global Urban Imprint

We not turn to other associations between geographic barriers and major socioeconomic and environmental outcomes. These associations might be mediated through poorer road connectivity [29] or other channels [23]. Most such channels could be endogenous to the dependent variables in the study. In contrast, spatial nonconvexities are not reverse-caused by the left-hand-side variables and impervious to contemporaneous change.

In the regressions of Table 2, each city represents one observation and the main explanatory variable is its *nonconvexity*. As seen in Table 1, the *share of barriers* is a highly co-linear and weaker variable. We also control for the rich battery of variables in Table 1, column 4, including country, biome, climate, and soil fixed effects (ancillary coefficients in Supplementary, Table S6). Nonetheless, the results are robust to the omission of these additional controls. We show results for five datasets, from top to bottom panel: our computed GHSL-2023-UCDB; the less-precise 2019 vintage of the UCBD (for reproducibility); our computations based on GHUB and GUB delineations; and the AUE.

Column 1 reveals that naturally nonconvex environments ultimately begot lower urban populations. Qualitative results are ironclad-consistent across datasets, but point estimates differ. They are stronger for the GHSL-2023-UCDB and the AUE, with city definitions that select for relatively larger populations. This suggests that natural constraints are more problematic for expansion once the city has reached a minimum size.

Table 2 Associations between urban nonconvexity and socioeconomic outcomes in four datasets

	UCDB (2023 vintage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log of Total Resident Population in 2015	Log of resident population density in build-up area	Share of build-up area within urban unit	Log of GDP Per Capita	Log of resident population density in build-up area	Share of build-up area within urban unit	Log of average build-up area height
Average Local Nonconvexity, 10km	-1.686***	1.575***	-0.133***	-0.904***	1.439***	-0.0970***	0.184***
	(0.174)	(0.0775)	(0.0076)	(0.187)	(0.0772)	(0.0072)	(0.0375)
Log of GDP Per Capita in 2015					-0.088***	0.0046***	0.0186***
					(0.0037)	(0.0003)	(0.0018)
Log of Total Resident Population in 2015				0.365***	0.0373***	0.0206***	0.138***
				(0.0098)	(0.0042)	(0.0004)	(0.002)
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
Adjusted R-squared	0.261	0.659	0.542	0.604	0.678	0.644	0.687
	UCDB (2019 vintage)						
Average Local Nonconvexity, 10km	-0.607***	2.997***	-0.223***	-1.070***	2.194***	-0.177***	0.0756*
	(0.123)	(0.173)	(0.0164)	(0.143)	(0.148)	(0.0158)	(0.0409)
Observations	13,109	13,102	13,109	12,796	12,789	12,796	12,761
	GHUB						
Average Local Nonconvexity, 10km	-0.481	0.511***	-0.0529***	-0.684***	0.717***	-0.0480***	0.425***
	(0.337)	(0.128)	(0.0129)	(0.149)	(0.0967)	(0.0112)	(0.0459)
Observations	10,135	10,135	10,142	10,082	10,082	10,082	10,082
	GUB						
Average Local Nonconvexity, 10km	-0.936***	0.354***	-0.049***	-0.821***	0.602***	-0.0298***	0.373***
	(0.276)	(0.113)	(0.0124)	(0.13)	(0.082)	(0.011)	(0.0469)
Observations	14,055	14,055	14,073	13,963	13,963	13,963	13,963
	AUE						
Average Local Nonconvexity, 10km	-4.634**	2.672**	-0.108	-2.094	2.603***	-0.0211	0.773
	(2.318)	(1.037)	(0.123)	(1.522)	(0.926)	(0.118)	(0.547)
Observations	200	200	200	200	200	200	200

Note: Standardized beta coefficients; t statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each coefficient captures the estimate of the impact of urban geographic nonconvexity in a separate regression with the variable in the column name as dependent variable. All columns control for city-level variables, including "Dummy for Ocean or Lake Intersection", "Dummy of Capital City", "City not in Major River Watershed", "Log of Average Precipitation", "Log of Average Temperature", "Log of Average Elevation", "Total Resident Population". Fixed effects (Country, Soil , Climate and Biome Types) are also included, except in the regressions using the AUE data. Column 4,5,6,7 also control for the "Log of Population" and columns 5,6, and 7 additionally for the "Log of GDP per Capita." Details of the regressions are in Supplementary Section 3 and Table S6.

In our discussion, we sometimes discuss quantitative magnitudes, for which we use the median parameter estimates across datasets. Note that the distribution of urban *nonconvexity* is approximately log-normal, with most cities displaying measurements close to zero. Hence, to focus on the cities where these effects matter, we discuss the effects of moving from the median (*nonconvexity*=.013) to the 90^{th} percentile (*nonconvexity*=0.105). Doing so would be associated with a population that is approximately 8.5 percent smaller.

Column 2 shows results on the log of urban density: city population divided by the total urban built-up area. The relationship between topographical complexity and urban land use is not obvious a priori. On the one hand, natural land fragmentation may encourage leapfrogging. [51] report on such phenomenon around the mountains surrounding Phoenix (Arizona, USA). On the other hand, barriers may corset real estate development into relatively small flat land areas, as in Hong Kong [52]. [22] report negative associations between mountains and sprawl in 275 US metro areas, but simultaneously positive ones regarding terrain ruggedness. These two variables are highly correlated, so that it is difficult to ascertain their combined effect.

Results, in column 2 of Table 2, strongly suggest that the constraining effect of geography is dominant worldwide: a move across the 50-90 percentile in barriers is associated with 14.5 percent higher urban densities. Moreover, it is also associated with a reduced built-up imprint as a percentage of the urban area (column 3).

In Table 2, column 4 we also find that geographic fragmentation in and around cities is associated with lower GDPs per capita. This is not driven by simple agglomeration economies [53, 54]—implying a mechanical impact from lower populations to lower income—because we control for city size in this specification. The 50-90 effect implies 11 percent lower incomes per capita. Highly constrained cities may therefore experience lower productivity, likely because of increased transportation costs [29] and more expensive housing [23].

In columns 5 and 6 we repeat the exercises in columns 2 and 3, controlling now for the logs of GDP per capita and population. Results do not change much. Even conditional on their lower incomes and sizes, geographic barriers beget denser cities with smaller built-up shares.

Finally, we examine average building height, resulting from satellite estimates on overall building volumes in the urban areas, divided by their total built-up surface. Consider POP_i as the total population of city i and express it as $POP_i = A_i^{B,R} \times \varphi_i \times \delta_i$, where $A_i^{B,R}$ is the land area in the city occupied by residential buildings, φ_i is the average floor-to-area ratio (FAR)—the square meters of housing per square meter of residential land—, and δ_i the number of people per housing square meter [55]. $A_i^{B,R}$ is not observed, but we do have measurements of A_i^B, the overall built-up area—including non-residential buildings, parks, streets, and sidewalks. Defining σ_i as the share of built-up area that is devoted to residential uses ($A_i^{B,R} = \sigma_i \times A_i^B$) yields: $\ln(\frac{POP_i}{A_i^B}) = \ln(\sigma_i) + \ln(\varphi_i) + \ln(\delta_i)$. Accommodating more people on a given built-up area can be achieved through increasing building heights/FAR (φ_i). Indeed, this is what we find in Table 2, column 5. The estimates are imprecise, but they are mostly below those on density. Therefore, higher densities must also be achieved through a more intense residential use of land—higher σ_i—or via smaller home floor plates—higher δ_i. These three mechanisms may require of relatively higher land prices [56], consistent with [23]. However, there is no data on home values at the extent that we cover here.

3.5 Urban Environmental Outcomes

Finally, we move our attention to environmental outcomes. These are only provided at a global scale by the UCDB, and therefore we pivot into using that dataset exclusively.

Table 3 Associations between urban nonconvexity and socioeconomic outcomes

	(1)	(2)	(3)	(4)
	Log Average Night Time Light Emission in 2015	Average greeness	Log Residential CO_2 Emissions non-organic 2015	Log Transportation CO_2 Emissions non-organic 2015
Average Local Nonconvexity, 10km	-0.772***	0.0174*	-0.896***	-1.598***
	(0.154)	(0.0106)	(0.134)	(0.271)
Log of Total Resident Population in 2015	0.367***	-0.0177***	1.338***	1.372***
	(0.0104)	(0.0007)	(0.0093)	(0.0186)
Log GDP per Capita in 2015		-0.003***	0.226***	0.405***
		(0.0007)	(0.0083)	(0.017)
Other Control Variables	YES	YES	YES	YES
Observations	12524	12374	12795	12541
Adjusted R-squared	0.691	0.684	0.898	0.672

Note: Standardized beta coefficients; t statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each coefficient captures the estimate of the impact of urban geographic *nonconvexity* in a separate regression with the variable in the column name as dependent variable. All columns control for city-level variables, including "Dummy for Ocean or Lake Intersection", "Dummy for Capital City", "City not in Major River Watershed", "Log of Average Precipitation", "Log of Average Temperature" and "Log of Average Elevation". All columns also control for country, soil type, climate and biome fixed effects.

Table 3 starts with the log of night light intensity as dependent variable (column 1). This metric is oftentimes used as a proxy for local income [57]. Consistent with our earlier results, we find topographical land fragmentation to be associated with reduced urban night lights.

Column 2 now turns to the average greenness of the urban area, as calculated by the UCDB. These data resulted from processing Landsat annual Top-of-Atmosphere (TOA) reflectance composites. Their algorithm processes surface color data to capture land with semi-natural vegetation cover: e.g. street trees, lawns, parks, gardens, forests, or green roofs. We find natural *nonconvexity* to be associated with slightly greener urban areas, which is consistent with their smaller share of built-up area found earlier.

Finally, columns 3 and 4 in Table 3 focus on pollutant emissions—regarding the residential and transportation sectors respectively. The literature typically finds reduced energy use in buildings [58] and transportation [59] in denser cities, as well as less CO_2 emissions [60, 61]. However, this conclusion depends on the scale, measurement, or definitions of density and proximity between locales [62].

In our data, there is a very strong negative association between geographic fragmentation and both types of emissions. Moving across the 50-90 percentile range of urban *nonconvexity* yields residential and transport CO_2 emissions that are, respectively, 11% and 20% smaller—likely because of the reduced land footprint of topographically complex cities. We caveat the results with the fact that the local CO_2 emissions estimates by [62] were produced with statistical models that may take into account urban density.

Of course, the accumulation of pollutants—as opposed to their emission—may be higher in areas adjacent to mountains depending on the specific diffusion processes allowed by the local orography (e.g.,[63]). Note also that, in table 3, we are controlling

for income and population: because both variables are associated with more emissions—and due to their negative relationship with barriers—the unadjusted impact of urban geographic nonconvexities would be even more strongly negative.

4 Discussion and Conclusions

We have provided three new measurements to account for the impact of natural barriers on urban environmental and socioeconomic outcomes using 4 different definitions of global urban areas, spanning 37,675 urban units in 182 countries around the globe.

We first calculated the share of 5 and 10 km disks around cities that is lost to oceans, internal bodies of water, highly sloped terrain, and international borders. This variable captures land availability for urbanization around city centers—or its lack thereof—extending the previous coverage of this kind of measure to most cities worldwide.

However, the impact of natural barriers on urban functionality is determined not solely by their size or proportion but also by their morphology, as measured by *nonconvexities*. Hence, we also generated a measure of set convexity by calculating the average share of the straight segment between any two points around city's centroid that is intersected by such features. This index captures natural land fragmentation. Additionally, we have provided a metric for road connectivity between points in and around the city by calculating their average *detour* factor.

After characterizing the global distribution of barriers, we have provided—to the best of our knowledge—the first worldwide comparative analysis of the impact of geographic fragmentation on urban socioeconomic and environmental outcomes. We documented a very robust set of results. Natural nonconvexities are critical determinants of average road distances, thereby impeding connectivity in and around urban entities. Cities that sprouted in areas with topographic land fragmentation tend to be smaller, denser, taller, greener, and less polluting, but also poorer an darker at night.

A growing comparative corpus of research on global cities studies the major "triad" of urbanization outcomes: population, built-up area, and density [64]. A branch of this literature associates higher urban density with lower energy use [58, 59, 65]. The results in this paper suggest that future comparative work on socioeconomic and environmental outcomes across cities should adjust for natural barriers as baseline controls. Otherwise, researchers and policymakers may attribute human agency to outcomes that are partially driven by geography. Future work should investigate the channels that drive the relationships identified here, as well as their interactions with policy and other local city features. To that effect, we provide the datasets in open-source format.

Further research can also use these or similar geographic data to study other dimensions of urban morphology [66–68], thereby extending studies such as [28] to a global sample of cities. An outstanding question for further research is the role of housing markets in land constrained cities at a global level. Future work should also study the comparative impact of difficult geographic conditions on housing and infrastructure construction costs.

Data measurements such as the ones in this paper may also assist policymakers in the creation of formulae for the assignation of budgetary resources. Physically fragmented cities will require of more resources to ensure social and economic connectivity. In practice, cities with non-convexity indexes at above the 75th percentile in our data may require specific budgetary assistance and studies about the impact of topography on infrastructure adequacy and urban connectivity. By better comparing the extent of barriers across cities, policymakers can make more informed decisions regarding land use, transportation infrastructure, and public facilities, ultimately promoting more equitable and sustainable urban growth.

In low-laying areas where land fragmentation is by due to ocean intrusion, inlets, wetlands, or riverine estuaries, future sea level rise may not only endanger life and destroy structures, but also fragment the urban continuum thereby increasing intra-city transportation costs. Beyond mitigation efforts and relocation assistance to those directly affected, policymakers in these cities may consider reinforcing local transportation systems and plan for economic policies that foster metropolitan connectivity. Regional strategies to link the areas fragmented by sea level rise to new and alternative hinterlands may become necessary.

Other direct policy applications arise in the new cities and large-scale satellite suburbs that are increasingly being built ex novo around the world. Detailed calculations and projections about future fragmentation driven by barriers as these developments grow is critical. More broadly, planners should consider geographic hurdles—even 20-50 years in the future—to ensure that the directionality of urban growth is not impeded or fragmented by barriers.

Methods

Data sources

Global databases on cities

We utilize four global databases on cities. (1) The primary dataset utilized in our analysis is the Global Human Settlement Layer Urban Centres Database (GHS-UCDB) [36], which currently stands as one of the most comprehensive databases on cities. The latest version(UCDB R2019A 1.2), released in April 2020, identifies 13,135 global cities as of 2015. The GHS-UCDB offers a multidimensional perspective, encompassing five categories: Geography, Socio-economic factors, Environmental aspects, Disaster Risk Reduction, and Sustainable Development Goals. (2) The second dataset is the Atlas of Urban Expansion [40], providing detailed and precise information regarding the spatial expansion of 200 representative cities worldwide. (3) The third dataset is The Global Hierarchical Urban Boundaries (GHUB) [38], which is a vector dataset that delineates the spatial distribution of urban areas, built-up areas, and various landscapes, including urban open spaces and water bodies. GHUB combines elements used in the definition of the UCDB and AUE datasets and has identified 10,242 urban settlements across the world. (4) The fourth dataset is the Global Urban Boundaries (GUB) dataset [39], which is the first of its kind to offer multi-temporal global urban boundary data.

The above four datasets cover two different and most commonly discussed definition of cities: *people-based* (UCDB) and *building-based* (GHUB, GUB and AUE). Despite the inherent differences in these definitions, the delineation of urban extents for most cities appears to be consistent across the four datasets through the comparison in [37]. We also select several cities (see Supplementary Figure S1) to demonstrate that their urban extents generally align with each other. A detailed discussion and comparative overview of the four datasets is presented in Supplementary Table S1.

Geographic barriers and socioeconomic attributes

For geographic barriers, Water bodies (oceans, reservoirs, and rivers) are derived from Global 3 arc-second Water Body Map (G3WBM) at 90 m spatial resolution [69]. Slopes are calculated mainly from NASA Shuttle Radar Topographic Mission 90m-resolution Digital Elevation Database v4.1 (SRTM version 4.1)[70]. National boundaries are extracted from Version 3.6 of the Database of Global Administrative Area (GADM) [71].

The socioeconomic attributes used in the regressions are derived through external data sources if the attributes were not provided by their original urban centre datasets. For the GHS-UCDB dataset, we mainly adopt its city-level variables from its original attributes and update several variables (population, build-up area, income per capita, etc.) using the latest estimates from gridded global data sources—such as the GHSL 2023 data package[41]. We also incorporated country-level indicators to the UCDB, which are derived from World Development Indicators (WDI) database[72] and Freedom House database [49]. For the GHUB and GUB, we overlaid their delineations with several gridded data sources to derive their attributes. The main attributes that we computed include population, build-up area, coastal city, capital city, major river watershed, precipitations, temperatures, elevation, GDP, built-up volume, built-up height, soil type, climate, and biome. The specifics on the calculations of these attributes, along with the data sources and how they correlate with each dataset, are further elaborated on in Supplementary Table S2. For the AUE, we adopt the same attributes as GHUB and GUB, except for two of its original variables: population and built-up area.

Computation of geographic indexes

Share of barriers

The *share of barriers* index quantifies the presence of geographic barriers in a typical urban region, as measured in studies examining the feasibility of urban expansion and housing supply elasticity in US cities[23]. As circular urban development patterns often emerge [73, 74], and in order to avoid endogeneity in the definition, geographic barriers are commonly assessed using a circular buffer around the city center [23, 24]. Consequently, we define the urban area as a 10 km or 5km radii around the center, as derived from the four datasets, to ascertain whether geographic barriers within this region impact potential urban functions. The *share of barriers* index thus represents the ratio of the three types of geographic barriers to the total urban area within a 10 km radius of the city center, regardless of its shape. In our definition, the *share of*

barriers within an i km distance from each urban center is calculated by:

$$S_i = \frac{\sum P_i + \sum W_i + \sum \widehat{B}_i}{\pi \times i^2} \tag{1}$$

where W_i and P_i denote the areas of water coverage and steep-slope areas located within the i(5 or 10) km circle around the city center, respectively. $\widehat{B}_i$ represents the area outside the national boundary excluding its overlaps with mountains and water. Finally, S_i denotes the *share of barriers* index of the urban unit. While water and slopes do not overlap, we avoid duplications with regard to borders. We choose a representative city, Lugano, Switzerland, to illustrate the distribution of barriers and the extraction of its buildable urban area (Supplementary Figure S3).

Average Dyadic Nonconvexity

The *nonconvexity* index measures the effects of geographic barriers on the straight-line connectivity between different locations, distinguishing between cities with similar *share of barriers* but different connectivity (Supplementary Figure S4). To derive our *nonconvexity* index, we generate evenly distributed sampling points in a grid at horizontal and vertical intervals of 0.5 km at radii 10 km (0.25 km at radii 5 km). We then intersect this point grid to the relevant urban disks (at radii 10 and 5 km from each urban center). For each dyadic combination of origin and destination inner points—this is, excluding points that fall into barriers themselves we calculate their straight-line distance, and also the share intersected by geographic barriers (including water, slopes, and national boundaries).

The *nonconvexity* index is derived in three steps, as shown in Supplementary Figure S5: (1) We calculate the length of the straight-line connection between each pair of inner points—those in developable land—and the total length of parts intersected by geographic barriers, through which: (2) the share of intersected parts for are calculated. (3) The average share of lines intersected by barriers are set as the *nonconvexity* index of the urban area. The process can be formulated as:

$$NC = \frac{1}{n^2} \sum \sum \frac{SLDI_{p_i p_j}}{SLD_{ij}} \tag{2}$$

where SLD_{ij} denotes the length of straight-line distance between point i and j, $SLDI_{ij}$ denotes the length of lines intersected by geographic barriers between point $SLDI_{ij}$, n denotes the number of points located within the urban center at an interval of 5 or 10 km, and NC denotes the *nonconvexity* index of the urban center.

Detour index

The goal of the *detour* index is to measure a minimum bound for the extra driving distances that are caused by geographic barriers and by the design of the street/road network. To achieve this goal, we follow the procedures below, represented in Supplementary Figure S6:(1) Creating sample nodes: As before, we use the sample points created during the calculation of *nonconvexity* (0.5 km spatial intervals at radii 10 km

and 0.25 km spatial intervals at radii 5 km). We use these points to represent the potential commuting origins or destinations within the city. The interaction and connectivity between these points can approximate the daily commutes within the city under the impacts of geographic barriers. The coordinates of points in each urban center are recorded.(2) Calculating the two types of distance: We calculate two distance metrics: the Euclidean distance and the minimum road distance for each origin-destination pair of sample nodes in each city. The Euclidean distance is defined as the straight-line distance (see Fig. S4) connecting two sample nodes. The minimum road distance is defined as the distance of the shortest roads (e.g., streets, bridges, and tunnels) connecting origin and destination nodes and is derived through *OSMnx* [75]. (3) Generating the *detour* index: Based on the definitions above, we define the *detour* between point p_i and p_j as,

$$detour_{ij} = \frac{d_{ij}}{D_{ij}} - 1 \tag{3}$$

where d_{ij} is the *minimum road distance* between p_i and p_j, D_{ij} is the *Euclidean distance* between p_i and p_j. We further calculate the *detour* index of the whole city, as the average of all point dyads:

$$Detour = \frac{\sum detour_{ij}}{n^2} \tag{4}$$

where n represents the number of points within 10 km (or 5 km for $detour_{5km}$) of the city center. The denominator is the number of point pairs. The detours defined by Eq. (3) and Eq. (4) indicate the lowest bound of the extra commuting distance caused by the combination of geographical barriers and the design/investments in the existing road system, including bridges and tunnels.

Robustness checks

Test with 5 km radius To check on the robustness of the results to our definitions of radius, we reproduce the results in Table 1 and Table 2 for the all global databases using a 5 km radius. The results in Supplementary Table S7, S8 indicate our findings remain unchanged.

Data trimming. We present the results using a trimming technique to the GHSL23-UCDB (our data processing of UCDB's boundaries), commonly used to address concerns about extreme observations. Concretely, conservatively, we remove observations that are at the top and bottom decile of the density distribution. The results based on this "clean" sample— now available in Supplementary Table S.9 —are almost identical to the ones using the full sample.

Excluding national boundaries. Some borders are less permeable than others. Generally, land beyond the border is outside of the control of local planners and land registries. Therefore, we expect that expansion beyond borders be more difficult on average. Nonetheless, to ascertain if this is an issue, Table S10 shows the main results with a new measure of *nonconvexity* that includes areas beyond borders within the compact set (this is, we ignore the existence of borders). The results ignoring international borders are identical to the ones found in Table 2. We show results using

our new GHSL23 database with UCDB boundaries for brevity, but the results are also the same for the other databases.

Endogeneity of the Urban footprint. We have discussed why the employment of exogenous 10 km (or 5km) buffers is better than using the cities' urban footprint. Utilizing city boundaries or other topographic measures *within* the current footprint would introduce endogeneity issues. A detailed discussion of this endogeneity issue is available in the Supplementary Section 3-Robustness checks and the related results are in Supplementary Figure S21, Table S11 and Table S12. Details of case studies mentioned in main text are shown in Figure S22, and robustness test for large cities is discussed in Table S13.

Data, Materials, and Software Availability

The original datasets analyzed in this study are publicly available, as referenced in the article. The indicators derived from the datasets, the three urban indexes measuring geographic barriers, and the regressions we use in this study can be found in https://github.com/WilliamLiPro/Fragmented_by_Nature.

Acknowledgements

We are grateful for the research assistance from Dongxiao Niu, Zehua Ma and Xu Wang for their helpful comments during the experiment and earlier draft of this paper.

Author contributions

L.W. and A.S. designed the project, performed the analysis, wrote and edited the manuscript. W.L. designed the algorithm and reviewed the paper.

Additional information

- Supplementary Information accompanies this paper at https://github.com/WilliamLiPro/Fragmented_by_Nature
- Competing interests: The authors declare no competing interest.

References

[1] Batty, M. The size, scale, and shape of cities. *Science* **319**, 769–771 (2008).

[2] Spaces of globalisation: Reasserting the power of the local. *Capital & Class* **24**, 150–151 (2000). URL https://doi.org/10.1177/030981680007000110.

[3] Glaeser, E. L. & Gottlieb, J. D. The wealth of cities: Agglomeration economies and spatial equilibrium in the united states. *Journal of Economic Literature* **47**, 983–1028 (2009).

[4] Nagendra, H., Bai, X., Brondizio, E. S. & Lwasa, S. The urban south and the predicament of global sustainability. *Nature Sustainability* **1**, 341–349 (2018).

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

[6] Barredo, J. I., Caudullo, G. & Dosio, A. Mediterranean habitat loss under future climate conditions: Assessing impacts on the natura 2000 protected area network. *Applied Geography* **75**, 83–92 (2016).

[7] Seto, K. C., Fragkias, M., Güneralp, B. & Reilly, M. K. A meta-analysis of global urban land expansion. *PloS one* **6**, e23777 (2011).

[8] Güneralp, B., Güneralp, İ. & Liu, Y. Changing global patterns of urban exposure to flood and drought hazards. *Global Environmental Change* **31**, 217–225 (2015).

[9] Acuto, M., Parnell, S. & Seto, K. C. Building a global urban science. *Nature Sustainability* **1**, 2–4 (2018).

[10] Angel, S., Parent, J., Civco, D. L., Blei, A. & Potere, D. The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050. *Progress in Planning* **75**, 53–107 (2011).

[11] Schneider, A. & Woodcock, C. E. Compact, dispersed, fragmented, extensive? a comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies* **45**, 659–692 (2008).

[12] Liu, L. & Meng, L. Patterns of urban sprawl from a global perspective. *Journal of Urban Planning and Development* **146**, 04020004 (2020).

[13] Angel, S. *et al. The dynamics of global urban expansion* (World Bank, Transport and Urban Development Department Washington, DC, 2005).

[14] Glaeser, E. L. & Kahn, M. E. in *Sprawl and urban growth* (ed.Jacques-François, T.) *Handbook of regional and urban economics*, Vol. 4 2481–2527 (Elsevier, New York, 2004).

[15] Cervero, R. Linking urban transport and land use in developing countries. *Journal of Transport and Land Use* **6**, 7–24 (2013).

[16] McIntosh, J., Trubka, R., Kenworthy, J. & Newman, P. The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000. *Transportation Research Part D: Transport and Environment* **33**, 95–110 (2014).

[17] Jedwab, R., Barr, J. & Brueckner, J. K. Cities without skylines: Worldwide building-height gaps and their possible determinants and implications. *Journal of Urban Economics* **132**, 103507 (2022).

[18] Wells, M. *Skyscrapers: Structure and design* (Laurence King Publishing, London, 2005).

[19] Moehle, J. P. Seismic design of reinforced concrete buildings (2015).

[20] Rosenthal, S. S. & Strange, W. C. The attenuation of human capital spillovers. *Journal of Urban Economics* **64**, 373–389 (2008).

[21] Barr, J., Tassier, T. & Trendafilov, R. Depth to bedrock and the formation of the manhattan skyline, 1890–1915. *The Journal of Economic History* **71**, 1060–1077 (2011).

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

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

[24] Harari, M. Cities in bad shape: Urban geometry in india. *American Economic Review* **110**, 2377–2421 (2020).

[25] Duque, J. C., Lozano-Gracia, N., Patino, J. E. & Restrepo, P. Urban form and productivity: What shapes are latin-american cities? *Environment and Planning B: Urban Analytics and City Science* **49**, 131–150 (2022).

[26] Angel, S., Parent, J. & Civco, D. L. Ten compactness properties of circles: measuring shape in geography. *The Canadian Geographer/Le Géographe canadien* **54**, 441–461 (2010).

[27] Sevtuk, A. & Amindarbari, R. Does metropolitan form affect transportation sustainability? evidence from us metropolitan areas. *Environment and Planning B: Urban Analytics and City Science* **48**, 2385–2401 (2021).

[28] Angel, S., Franco, S. A., Liu, Y. & Blei, A. M. The shape compactness of urban footprints. *Progress in Planning* **139**, 100429 (2020).

[29] Saiz, A. & Wang, L. Physical geography and traffic delays: Evidence from a major coastal city. *Environment and Planning B: Urban Analytics and City Science* **50**, 218–243 (2023).

[30] Akbar, P., Couture, V., Duranton, G. & Storeygard, A. Mobility and congestion in urban india. *American Economic Review* **113**, 1083–1111 (2023).

[31] Akbar, P. A., Couture, V., Duranton, G. & Storeygard, A. The fast, the slow, and the congested: Urban transportation in rich and poor countries (2022). Preprint on webpage at https://congress-files.s3.amazonaws.com/2023-07/Global_mobility_paper.pdf.

[32] Cutter, S. L. The landscape of disaster resilience indicators in the usa. *Natural Hazards* **80**, 741–758 (2016).

[33] Hinkel, J. & Klein, R. J. Integrating knowledge to assess coastal vulnerability to sea-level rise: The development of the diva tool. *Global Environmental Change* **19**, 384–395 (2009).

[34] Bertaud, A. Clearing the air in atlanta: transit and smart growth or conventional economics? *Journal of Urban Economics* **54**, 379–400 (2003).

[35] Suzuki, H., Cervero, R. & Iuchi, K. *Transforming cities with transit: Transit and land-use integration for sustainable urban development* (World Bank Publications, 2013).

[36] Florczyk, A. *et al.* Description of the ghs urban centre database 2015. *Public Release* **1**, 1–75 (2019).

[37] Dijkstra, L. *et al.* Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics* **125**, 103312 (2021).

[38] Xu, Z. *et al.* Mapping hierarchical urban boundaries for global urban settlements. *International Journal of Applied Earth Observation and Geoinformation* **103**, 102480 (2021).

[39] Li, X. *et al.* Mapping global urban boundaries from the global artificial impervious area (gaia) data. *Environmental Research Letters* **15**, 094044 (2020).

[40] Angel, S., Blei, A. M., Civco, D. L. & Parent, J. *Atlas of Urban Expansion* (Lincoln Institute of Land Policy Cambridge, MA, 2016).

[41] Pesaresi, M. & Politis, P. Ghs-built-s r2023a - ghs built-up surface grid, derived from sentinel2 composite and landsat, multitemporal (1975-2030) (2023). URL http://data.europa.eu/89h/9f06f36f-4b11-47ec-abb0-4f8b7b1d72ea.

[42] Irwin, E. G. & Bockstael, N. E. The evolution of urban sprawl: Evidence of spatial heterogeneity and increasing land fragmentation. *Proceedings of the National Academy of Sciences* **104**, 20672–20677 (2007).

[43] Reis, J. P., Silva, E. A. & Pinho, P. Spatial metrics to study urban patterns in growing and shrinking cities. *Urban Geography* **37**, 246–271 (2016).

[44] Lutz, C. & Sand, B. Highly disaggregated land unavailability. *Available at SSRN 3478900* (2023).

[45] Müller, I., Taubenböck, H., Kuffer, M. & Wurm, M. Misperceptions of predominant slum locations? spatial analysis of slum locations in terms of topography based on earth observation data. *Remote sensing* **12**, 2474 (2020).

[46] Van Ham, M., Uesugi, M., Tammaru, T., Manley, D. & Janssen, H. Changing occupational structures and residential segregation in new york, london and tokyo. *Nature Human Behaviour* **4**, 1124–1134 (2020).

[47] Casey, E. S. & Watkins, M. Up against the wall: Re-imagining the us-mexico border (2014).

[48] Bank, W. World bank country and lending groups (2017).

[49] House, F. Freedom in the world 2022 (2022). URL https://freedomhouse.org/report/freedom-world. Https://freedomhouse.org/report/freedom-world. Accessed August 15 2023.

[50] Saiz, A. Dictatorships and highways. *Regional Science and Urban Economics* **36**, 187–206 (2006).

[51] York, A. M. *et al.* Land fragmentation under rapid urbanization: A cross-site analysis of southwestern cities. *Urban Ecosystems* **14**, 429–455 (2011).

[52] Mahtab-uz Zaman, Q., Lau, S. S. & Mei, S. H. The compact city of hong kong: a sustainable model for asia? *The compact cities: Sustainable urban form for developing countries* 255–268 (2000).

[53] Duranton, G. & Puga, D. in *Micro-foundations of urban agglomeration economies* (eds Henderson, V. & Thisse, J. F.) *Handbook of regional and urban economics*, Vol. 4 2063–2117 (Elsevier, 2004).

[54] Rosenthal, S. S. & Strange, W. C. in *Evidence on the nature and sources of agglomeration economies* (eds Henderson, V. & Thisse, J. F.) *Handbook of regional and urban economics*, Vol. 4 2119–2171 (Elsevier, 2004).

[55] Jedwab, R., Loungani, P. & Yezer, A. Comparing cities in developed and developing countries: Population, land area, building height and crowding. *Regional Science and Urban Economics* **86**, 103609 (2021).

[56] Ahlfeldt, G. M. & McMillen, D. P. Tall buildings and land values: Height and construction cost elasticities in chicago, 1870–2010. *Review of Economics and Statistics* **100**, 861–875 (2018).

[57] Gibson, J., Olivia, S. & Boe-Gibson, G. Night lights in economics: Sources and uses 1. *Journal of Economic Surveys* **34**, 955–980 (2020).

[58] Güneralp, B. *et al.* Global scenarios of urban density and its impacts on building energy use through 2050. *Proceedings of the National Academy of Sciences* **114**, 8945–8950 (2017).

[59] Kaza, N. Urban form and transportation energy consumption. *Energy Policy* **136**, 111049 (2020).

[60] Glaeser, E. L. & Kahn, M. E. The greenness of cities: Carbon dioxide emissions and urban development. *Journal of urban economics* **67**, 404–418 (2010).

[61] Lee, S. & Lee, B. The influence of urban form on ghg emissions in the us household sector. *Energy policy* **68**, 534–549 (2014).

[62] Crippa, M. *et al.* Gridded emissions of air pollutants for the period 1970–2012 within edgar v4. 3.2. *Earth Syst. Sci. Data* **10**, 1987–2013 (2018).

[63] Schmitz, R. Modelling of air pollution dispersion in santiago de chile. *Atmospheric Environment* **39**, 2035–2047 (2005).

[64] Güneralp, B., Reba, M., Hales, B. U., Wentz, E. A. & Seto, K. C. Trends in urban land expansion, density, and land transitions from 1970 to 2010: A global synthesis. *Environmental Research Letters* **15**, 044015 (2020).

[65] Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P. & Seto, K. C. Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Sciences* **112**, 6283–6288 (2015).

[66] Longley, P. A. & Mesev, V. On the measurement and generalisation of urban form. *Environment and Planning A: Economy and Space* **32**, 473–488 (2000).

[67] Herold, M., Scepan, J. & Clarke, K. C. The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and planning A* **34**, 1443–1458 (2002).

[68] Herold, M., Goldstein, N. C. & Clarke, K. C. The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote sensing of Environment* **86**, 286–302 (2003).

[69] Yamazaki, D., Trigg, M. A. & Ikeshima, D. Development of a global˜ 90 m water body map using multi-temporal landsat images. *Remote Sensing of Environment* **171**, 337–351 (2015).

[70] Jarvis, A., Reuter, H. I., Nelson, A., Guevara, E. *et al.* Hole-filled srtm for the globe version 4. *available from the CGIAR-CSI SRTM 90m Database (http://srtm. csi. cgiar. org)* **15**, 5 (2008).

[71] Hijmans, R., Garcia, N. & Wieczorek, J. Gadm: database of global administrative areas, version 3.6 (2018). Https://gadm.org/. Accessed August 20 2023.

[72] Bank, W. *World development indicators 2016* (The World Bank, 2016).

[73] Hamilton, B. W. Wasteful commuting again. *Journal of Political Economy* **97**, 1497–1504 (1989).

[74] Wheaton, W. C. Commuting, congestion, and employment dispersal in cities with mixed land use. *Journal of Urban Economics* **55**, 417–438 (2004).

[75] Boeing, G. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems* **65**, 126–139 (2017).

Supplementary Information for

Fragmented by Nature: Metropolitan Geography, Urban Connectivity, and Environmental Outcomes

Luyao Wang[1, 2], Albert Saiz [1, *], Weipeng Li [3, *]

[1] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge MA, USA

[2] State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

[3] Research Center of Big Data, College of Information and Communication, Wuhan, China

*Email: saiz@mit.edu or liweipeng@nudt.edu.cn

This PDF file includes:

Section 1 to Section 5
Figures S1 to S21
Tables S1 to S13

Contents

Section 1: Descriptions of the four city-boundary datasets

1.1 Basic information on city-boundary datasets

1. GHS-UCDB-2019

The first dataset utilized in our analysis is the Global Human Settlement Layer Urban Centres Database (GHS-UCDB) (Florczyk et al., 2019), which stands as one of the most comprehensive databases on cities. The UCDB defines a city—urban centre—as a contiguous area with a density exceeding 1,500 people per square kilometer and a minimum population of 50,000 (Dijkstra and Poelman, 2014). This database compiles population and built-up area data from a variety of satellite images and external data sources. Consequently, the latest revision, GHS STAT UCDB2015MT GLOBE R2019A 1.2, released in April 2020, identifies 13,135 global cities as of 2015. The database offers extensive information on global cities, encompassing five dimensions of urban units: geography, socioeconomics, environment, disaster risk reduction, and sustainable development goals. Moreover, many attributes in the database span a broad time frame, covering for 25 or even 40 years (from 1975 to 2015).

2. GHSL-UCDB-2023

Some data errors and discrepancies were observed in the GHS-UCDB-2019, especially affecting smaller cities. The Global Human Settlement Layer (GHSL) team writes: *"After the publication of the GHS P2022 new independent data showed an anomaly in the performance of the multi-temporal model that was not visible during the model development5. According to the JRC internal tests, the anomaly was expected to introduce a positive bias in predicted change rates of built-up surfaces and buildup volumes after the year 2000. The positive bias is especially remarkable in the rural domain as set by the GHS-SMOD R2022A"*(Pesaresi et al., 2023). Because built-up areas are used to estimate local population, these errors introduced bias in the calculation of population and density levels in the smaller cities. We have therefore recalculated the critical variables in our study using the urban center boundary files from the original UCDB but applying the most recent data updates to the GHSL of 2023 and other global gridded datasets, including new estimates for: population, built-up area, density, and income per capita.

3. AUE

The Atlas of Urban Expansion (AUE) is an open-source online resource with maps, satellite images, and data on spatial changes in cities around the world, created in partnership by New York University, the Lincoln Institute of Land Policy and UN-Habitat (Angel et al., 2016). The AUE datasets define a city based on the proportion of built-up area in $30\,m \times 30\,m$ cells and their proximity to other built-up cells. In 2010, this method identified 4,245 cities worldwide. A population-weighted sample of 200 of these cities were selected to provide detailed information. The Atlas provides extensive geographic and quantitative data on urban expansion and its key attributes across these 200 global cities. This data is primarily derived from medium-resolution Landsat satellite imagery and census data. The attributes include the overall urban extent, average population density, fragmentation of built-up areas by open spaces, compactness of geographic shape, and the proportions of infill, extension, leapfrog development, and newly added built-up areas within the urban extents for three time periods: circa 1990, 2000, and 2014. Additionally, the Atlas of Urban Expansion offers arterial road distribution data for 200 cities in vector (*.shp*) format. These roads were identified using three data sources: OpenStreetMap, Google Maps, and Bing Maps, where the roads are available as map layers.

4. GHUB

The Global Hierarchical Urban Boundaries (GHUB), developed by Wuhan University, is a vector dataset that delineates the spatial distribution of urban extents, built-up areas, and various landscapes, including urban open spaces and water bodies (Xu et al., 2021). GHUB defines the minimum size for an urban settlement as 5 square kilometers, aligning with Florczyk et al. (2019), who considers a population of 50,000 as the threshold for an independent urban settlement. Additionally, Angel et al. (2016) notes that the average population density in globally stratified sampling cities is 101 persons per hectare. Using these criteria, GHUB has identified 10,242 urban settlements exceeding five square kilometers. The primary data sources for GHUB include the Global Artificial Impervious Area (GAIA) dataset (Gong et al., 2020) and the EU Joint Research Center's (JRC) water body dataset (Pekel et al., 2016), both of which offer a high spatial resolution of 30 meters. The 2018 editions of these datasets were utilized to develop the GHUB products. GHUB' s accuracy and consistency have been validated through comparisons with the Atlas of Urban Expansion (AUE) and the Urban Centres Database (UCDB) (Xu et al., 2021).

5. GUB

The Global Urban Boundaries (GUB) dataset (Li et al., 2020), the first of its kind to offer multi-temporal global urban boundary data, was developed by Iowa State University and Tsinghua University. GUB identifies urban boundaries based on the spatial distribution of artificial impervious areas derived from Landsat data. This delineation process was carried out on the Google Earth Engine platform, resulting in the creation of 30-meter resolution global urban boundary datasets for seven representative years: 1990, 1995, 2000, 2005, 2010, 2015, and 2018. The urban boundaries delineated by GUB demonstrate strong concordance with results obtained from nighttime light data and human interpretation. These boundaries effectively outline the urban extents of cities, as validated against high-resolution Google Earth imagery. GUB includes urban units larger than 1 square kilometer and filters out smaller clusters below this size threshold. In 2015, the GUB dataset identified 65,582 urban clusters, of which 14,098, with areas larger than 5 square kilometers, were chosen as the study areas for our research.

1.2 Dataset Choice and quality assessment

Accurately measuring urban areas is critical for addressing challenges related to environmental sustainability and urban planning. Our study aims to assess the impacts of geographic barriers on urban connectivity and socioeconomic outcomes. To achieve this, we require two types of information: first, the boundaries of cities, which enable us to pinpoint their location, characterize city ranges, and subsequently match to their surrounding geographic environments; second, the socioeconomic attributes within these urban extents, such as population density, building heights, GDP, and other environmental indicators.

Accurately measuring the extent of a city is challenging, due to the complex nature of urban systems. Current global urban mapping efforts typically use one of two approaches: building-based or people-based. The building-based approach classifies spaces as urban or non-urban by standard pixel unit, primarily using remote sensing data at appropriate resolution levels. This method, although straightforward, often fails to capture nuanced differences in boundaries that are due gradual urban-to-rural transitions. In contrast, people-based approaches assess the quantitative characteristics of each pixel or patch, such as density and population size, as well as their spatial relationships, including continuity and proximity. Urban boundaries are then delineated based on the characteristics of the population distribution.

An important reason for our dataset selection is it includes the two types of city definition: **people-based (**UCDB) and **building-based** (AUE, GHUB and GUB). Therefore, we can investigate whether the impacts of geographic barriers are consistent and significant across different datasets and city definitions. Despite the inherent differences in these definitions, the basic delineations of urban extents for most cities appear to be largely consistent across the four datasets (Dijkstra et al., 2021). People-based and building-based definitions have respective advantages and disadvantages to address different scientific questions. These pros and cons have been extensively compared and discussed in (Dijkstra et al., 2021). Our study is ultimately skeptic about specific city delineations. A comparative overview of the four datasets is presented in Table S1-1.

Table S1-1. Comparison of the four global urban datasets

Name	Definition of urban center (People-based or Building-based)	Number of urban units	Time coverage
GHS-UCDB	People-based (Density >= 1500 /km2 and Population >= 50 000)	13,135	Geography: 2015 Attributes: 1975, 1990, 2000, 2015
AUE	Building-based	200	Geography: 1990*, 2000*, 2014* Attributes: 1990*, 2000*, 2014*
GHUB	Building-based (Area >= 5 km2)	10,242	Geography: 2018 No attributes.
GUB	Building-based	65,582 (14,098 >= 5 km^2)	Geography: 1990,1995, 2000, 2005, 2010, 2015, and 2018 No attributes.

*Time coverage of some cities in AUE may be from proximate years.

Another consideration is the availability of urban boundary datasets. Currently, the number of global-scale urban boundary products, especially those that include socioeconomic attributes, is very limited. This limitation is largely due to the challenge of obtaining reliable global data sources with both temporal and spatial consistency. Consequently, we selected the GHSL-UCDB-2019 as one of our primary datasets due to its comprehensive attribute data. We have rigorously examined all sources of these attributes and updated some of them (as well as extended) using the latest version of GHSL products (e.g., population, built-up area, and built-up height) to create our own GHSL-UCDB-2023 update, which we use in our study (details provided in the following section). The AUE products are well-recognized for their high accuracy and reliable attributes; however, they cover only 200 large cities, and their latest data is from around 2014. This limitation leads us to use the AUE primarily as a comparative dataset. In contrast, the GHUB and GUB datasets, produced from higher resolution (30m) and more recent remote sensing products, offering finer granularity compared to UCDB products (1000m). Besides these four datasets, we also considered using two others—the global urban extent product (MGUP) produced from MODIS datasets from 2001 to 2018 based on an automated mapping approach (Huang et al., 2021), and the percolation-based urban area product (PCCA) that quantifies urban areas using multi-source data based on percolation theory (Cao et al., 2020). The MGUP products, with a spatial resolution of 250m, are lower in resolution than both GHUB and GUB, and they lack the attribute data available in GHSL-UCDB, leading us to discount them as candidate products. The PCCA datasets, which include additional GIS data such as OSM and social sensing data, offer high accuracy (0.85). However, they only partially cover cities

in 28 countries and exhibit lower accuracy outside of China. Therefore, while they cannot be used directly for our global-scale analysis, we primarily utilize them as validation datasets.

After selecting the four urban boundary datasets, we conducted quantitative assessments of their data quality through two distinct comparisons. The first involved comparing each urban product to the AUE, which is widely seen as more accurate than most binary urban/non-urban products. We selected a range of cities from various countries (refer to Figure S1-1) to demonstrate that their urban delimitations generally align with one another. To assess accuracy in greater detail, we employed a random points approach similar to (Huang et al., 2021). This method involves generating random test points within the urban boundaries of each product and verifying whether these points also fall within the "urban extent" of the AUE. The number of random points used ranged from 26,351 to 30,491, with accuracy results displayed in Table S1-2. UCDB showed a high accuracy of 78.5% compared to the AUE, while GHUB and GUB also demonstrated good alignment, with accuracies of 80.3% and 84.1%, respectively. These results are consistent with those from the MGUP study, which assessed the overall accuracy of GHUB and AUE at 82.18% using 11,273 test points. It is important to note that the data sources used to delineate urban boundaries in the AUE dataset date back to around 2014, which is earlier than those used for the other three datasets. Because we would expect some misalignment to arise due to the timing differences, these tests validate the robustness of the global urban datasets used in the study.

In addition, we have also conducted a further validation exercise using another external data source—the PCCA (Cao et al., 2020). By integrating data on population, nighttime lights, and road distribution, the PCCA products show high accuracy compared to Landsat 8 images or high-resolution Google Earth imagery. The overall accuracy of their three types of products is documented at 0.93, 0.89, and 0.89 respectively. Moreover, PCCA products utilize population estimates based on mobile phone data, together with an exhaustive coverage of the road network in China. This results in even higher accuracy within that country. To validate the accuracy of our three global urban boundary datasets in China, we selected 43-57 cities from each dataset that align well with the PCCA delimitations. We then applied the same random point approach to quantify the accuracy for each city, with the number of points ranging from 300 to 1000 per city, depending on city size. The accuracy of the three products compared to the PCCA vary slightly, as illustrated in Figure S1-2. Consequently, we repeated the assessment three times and calculated the average accuracy for each dataset. The results, presented in Table S1-3, show that all three datasets maintain an accuracy greater than 70%, confirming a satisfactory data quality.

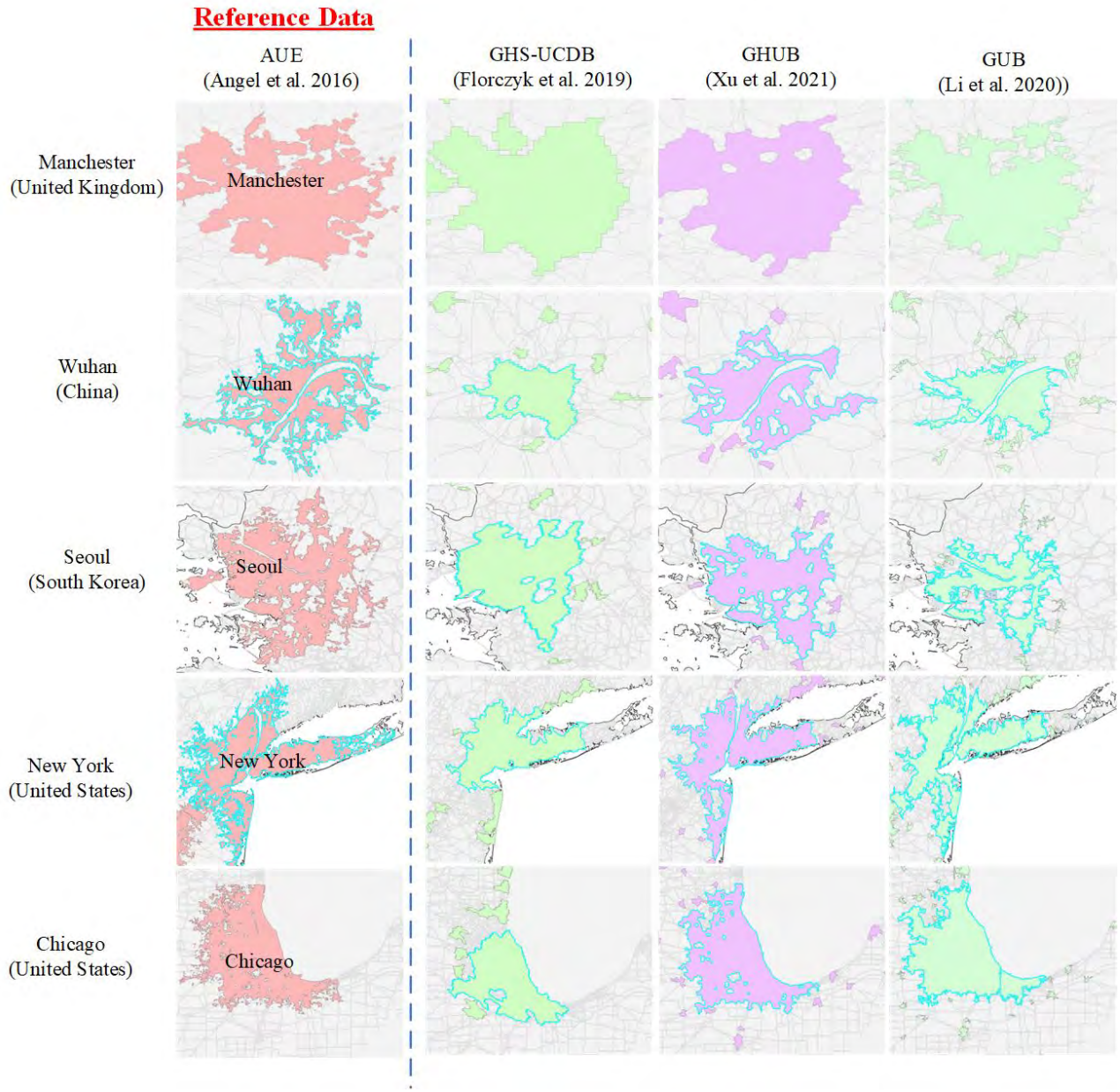

Figure S1-1. Accuracy assessment between the three urban datasets (UCDB, GHUB and GUB) with reference dataset (AUE). Several sample cities are displayed: Manchester (UK), Wuhan (China), Seoul (South Korea), New York (USA), Chicago (USA).

Note: Authors processing of original shapefiles.

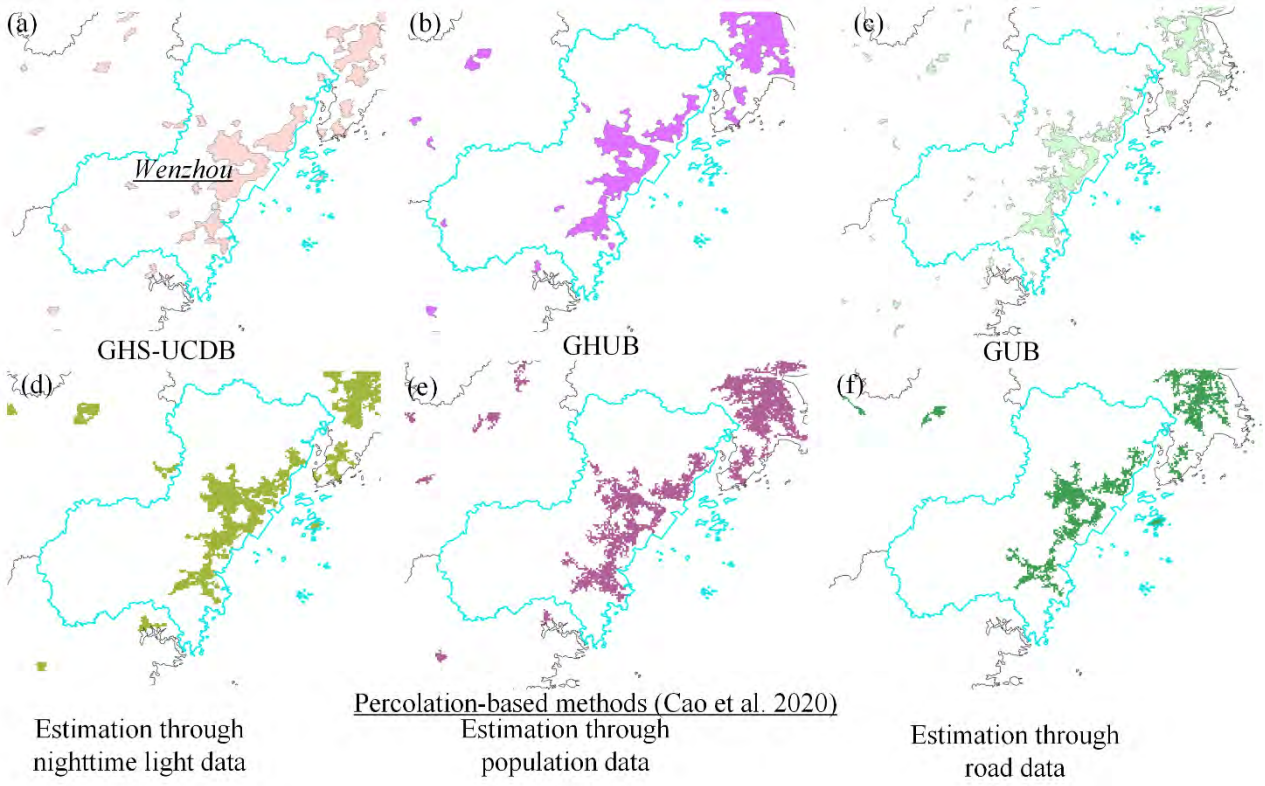

Figure S1-2. Accuracy assessment between the three urban datasets (UCDB, GHUB and GUB) used in this study and PCCA datasets.

Table S1-2. Accuracy evaluation between three global city datasets with AUE dataset.

Dataset	AUE	UCDB	GHUB (UE)	GUB
Test points	Reference	26, 351	30, 491	27, 620
Overall accuracy (%)	Reference	78.5	80.3	84.1

Table S1-3. Accuracy evaluation between three global city datasets with PCCA dataset.

Dataset	PCCA	UCDB	GHUB (UE)	GUB
Number of Cities	Reference	43	57	54
Overall accuracy (%)	Reference	76.2	82.1	88.3

1.3 Original and Derived Demographic and Socioeconomic Attributes in the datasets

The GHS-UCDB R2019A dataset is already rich in socio-economic attributes, encompassing many of the variables that we utilize in our analyses and the regression models specific to that dataset. The research of Freire et al. (2016) has validated the high accuracy of GHSL population products using the GEOSTAT 2011 resident population data on 18 European countries, finding a correlation of 0.83 (Freire et al., 2016). The quality of the GHSL built-up area data has also been confirmed: in the JRC technical report (Pesaresi et al., 2016), GHSL built-up data are compared with two external data sources—a systematic field survey by EUROSTAT and a set of digital cartographic products containing building footprints—finding total accuracies of 0.96 and 0.89 respectively; Liu et al. (2020) has evaluated the accuracy of GHSL build-up data using urban footprint information of 20 Chinese cities derived from Baidu Data (a high-accuracy mapping platform in China), setting the correlations of 250m and 1000m resolution products at 0.76 and 0.82, respectively. After our first draft of the paper, the Global Human Settlement Layer (GHSL) Database released a new data package—GHSL Data Package 2023 (Pesaresi et al., 2023), which we have processed to update the population data and build-up data in the original UCDB R2019A dataset.

In Table 1, country-level data are drawn from the World Development Indicators (WDI) (Bank, 2016) and from the Freedom House index (House, 2022). WDI is a comprehensive source for global economic conditions spanning six dimensions: World View, People, Environment, Economy, States and Markets, and Global Linkages. It contains over 900 variables for 208 economies, ranging from 1960 to the present. WDI offers a variety of measures reflecting social progress, quality of life, economic development, physical infrastructure, environmental conditions, and government performance. The database is compiled by the World Bank in collaboration with international partners. The WDI indicators used in this study include the ratio of urban population (as a % of total population), the ratio of population ages 0-14 (as a % of total population), the literacy rate (as a % of adult people), government expenditure (% of GDP), and country-level GDP per capita in 2015. In addition, we use the Political Rights freedom score, ranging from 0 to 7, with higher scores representing less political rights. This index is derived from the Freedom House dataset, which rates political rights and civil liberties in 210 countries and territories through its annual "Freedom in the World" report. The dataset provides a variety of quantitative measurements, such as the annual Political Rights freedom score and annual Civil Rights freedom score, from 1973 to the current year. In this study, we use the average annual Political Rights freedom score between 1973-2015 as the country-level freedom indicator.

For the other three city boundary delimitations—which primarily provide information about the urban extent of global cities but lack detailed socio-economic data—we have sourced the following attributes from external datasets. The specifics of these attributes, along with the data sources and how they correlate with each dataset, are further elaborated in Table S2.

- **Population:** The AUE dataset comprehensively includes population data for the urban areas of all its 200 cities, sourced from local census data corresponding approximately to the years 1990, 2000, and 2010. In contrast, the GHUB and GUB datasets do not directly provide population figures. Therefore, we proceeded to overlay their urban area shapefiles with most recent version of the 2023 vintage of the GHSL population raster data estimated for the year 2015 (Pesaresi et al., 2023). This integration enables us to calculate the total population within each urban unit, thus supplementing the GHUB and GUB boundary files.
- **Urban build-up area:** The AUE dataset directly contains the urban build-up area in the years 1990, 2000,

and 2010. The GHUB datasets contain a 'UB (urban build-up)' layer, which contains the coverage of the urban build-up area within each urban center. Regarding the GUB boundary files—which do not contain any build-up area information directly—we overlay them with GHSL build-up raster data at 100 m resolution for the year 2015 (Pesaresi et al., 2023) and calculate the total build-up area within each urban unit. We used the same method to reconstruct built-up area for the UCDB using their shapefiles.

- **Dummy variable for Ocean or Lake Intersection:** If the urban units is within 10 km distance of the ocean (calculated based on the Global 3 arc-second Water Body Map (G3WBM) (Yamazaki et al., 2015) we set the value of this variable to 1, while other cities are set to 0. The accuracy of G3WBM has been evaluated in (Lamarche et al., 2017) through a database of 2110 samples throughout the globe, and the overall accuracy is 0. 99, among which the accuracy for error-prone areas can also reach 0. 88.
- **Dummy variable for capital cities:** If the urban unit is the capital city of its country, the value is set to 1, and otherwise to 0.
- **Dummy of City not in Major River Watershed:** We overlay urban area boundaries with the Major River Basins (GRDC, 2020) global dataset. Cities that do not intersect with any river basins have this variable set to 1 and others to 0.
- **Precipitation and Temperature:** We derive the data from CRU TS v. 4.07 gridded time-series dataset (Harris et al., 2020) and overlay their rasters with urban unit boundaries from both the GHUB and GUB datasets, to calculate the average precipitation and temperature of the raster pixels within each urban unit in 2015. The CRU TS precipitation products have been proven to have high correlation with point-based observations (r=0.87) (Mutti et al., 2020), and even higher accuracy for temperatures as capture by the CRU's CRUTEM dataset (r=0.99) (Jones et al., 2012).
- **Elevation:** We derive elevation data mainly from NASA's Shuttle Radar Topographic Mission 90m-resolution Digital Elevation Database v4.1 (SRTM version 4.1) (Jarvis et al., 2008), which is notable for its coverage and accuracy—with absolute height error ranges from 5.6 m to 9.0 m (90% confidence) in major continent (Rodriguez et al., 2006). For places beyond its coverage (latitudes above 60 degrees north and below 56 degrees south), we adopt the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) datasets at 7.5 arc-second (~225m) resolutions instead (Danielson and Gesch, 2011). We overlay these raster layer with the urban boundaries from the GHUB and GUB datasets, respectively, to calculate the average elevation of pixels within each urban unit in 2015.
- **Local Gross Domestic Product (GDP):** We have updated the global gridded GDP raster, originally from (Kummu et al., 2018), by incorporating the 2023 vintage GHSL population raster data, which estimates the population for the year 2015. The original GDP raster, based on Gross Domestic Product Per Capita (GDPPC) and the earlier version of population data in GHS-POP R2015A, is less accurate compared to the 2023 GHSL data. To improve accuracy, we initially calculate grid-level GDPPC using GDP divided by GHS-POP R2015A. Then, we multiply this grid-level GDPPC by the 2023 vintage GHSL population raster data to produce an updated GDP raster product with an approximately 1km resolution. This revised GDP product is now used as the data source for all city datasets.
- **Countries:** We use the Database of Global Administrative Areas (GADM) data (GADM, 2022) to identify the country to which each urban unit in GHUB and GUB datasets belong.
- **Build-up volume and average height:** We derive information on build-up volumes from the GHSL Data

Package 2023 (Pesaresi et al., 2023), expressed as the number of building cubic meters. These data report the total built-up volumes for 3 arc-second (~100m) grids in year 2015. We also use these rasters to calculate the average build-up height in each urban unit—the total build-up volume divided by total urban build-up area.

- **Soil Type, Climate, and Biome:** These three kinds of geographic characteristics are relatively coarse and consistent at large scales. Therefore, we infer these characteristics using data from the nearest UCDB urban unit. A spatial analysis process is conducted for each urban unit in the GHUB and GUB datasets, respectively. For cities too far away (<50 km) from its nearby UCDB unit, these characteristics are set as *null*.

Table S2. The data sources of the major socioeconomic attributes in the three datasets (UCDB, AUE, GHUB and GUB).

Attributes	UCDB 2023	AUE	GHUB	GUB	Definition	Accuracy
Population	GHSL Data Package 2023 (Pesaresi et al., 2023)	AUE attributes	GHSL Data Package 2023	GHSL Data Package 2023	The total population within urban unit	R = 0.83 (Freire et al., 2016)
Urban build-up area	GHSL Data Package 2023	AUE attributes	GHSL Data Package 2023	GHSL Data Package 2023	The total build-up area within urban unit	R = 0.96 (Pesaresi et al., 2016)
Dummy variable for Ocean or Lake Intersection	Global 3 arc-second Water Body Map-G3WBM (Yamazaki et al., 2015)	G3WBM	G3WBM	G3WBM	Whether the urban units is within 10 km distance of the ocean (yes 1, no 0)	R= 0.99 (Lamarche et al., 2017)
Dummy of City not in Major River Watershed	UCDB 2019	GRDC Major River Basins of the World (GRDC, 2020)	GRDC Major River Basins of the World	GRDC Major River Basins of the World	Whether the urban units intersected with major river watershed (yes 0, no 1)	-
Precipitations and Temperatures	UCDB 2019	CRU TS datasets (Harris et al., 2020)	CRU TS datasets	CRU TS datasets	Average precipitation and temperature within each urban unit	R= 0.87 (Mutti et al., 2020) R= 0.99 (Jones et al., 2012)
Elevations	SRTM (Jarvis et al., 2008)	SRTM	SRTM	SRTM	Average elevation within each urban unit	5.6-9.0 m (Rodriguez et al., 2006)
Gross Domestic Product (GDP)	UCDB 2019	Gridded global datasets for GDP (Kummu et al., 2018)	Gridded global datasets for GDP	Gridded global datasets for GDP	Total GDP within each urban unit	-
Countries	UCDB 2019	-	(GADM, 2022)	(GADM, 2022)	-	-
Build-up volume and height:	GHSL Data Package 2023	GHSL Data Package 2023	GHSL Data Package 2023	GHSL Data Package 2023	Total build-up volume and average build-up height within urban unit	-
Soil Type, Climate, and Biome	UCDB 2019	UCDB 2019	UCDB 2019	UCDB 2019	Dummy variables for different kinds of characteristics	-

Section 2: Geographic indexes

2.1 Flowchart on the calculation of the geographic indexes

We measure geographic indexes for the four urban-boundary datasets, including the share of barriers, nonconvexity, and average detour. We proceed according to the following steps: (1) We first calculate a central point of each urban unit. AUE provides the Central Business District (CBD) geolocation, which we use directly. The GHUB contains the spatial boundaries of urban centers (UC) within each urban unit, which refers to the highest-density areas of surface agglomeration within each urban settlement. Therefore, for this dataset, we calculate the centroid of the UC. For the GUB and UCDB, which do not contain any information about their CBD, we use the centroid of each urban settlement instead. (2) We then create two circular buffers—of 5km and 10km—around each city center. For the UCDB boundary files, we generate an extra 5 random radiuses for robustness checks, as we will describe later. (3) We overlay each circular buffer with shapefiles capturing geographic barriers (water, steep slopes, and national boundaries) as well as road networks, and then calculate the three indexes for areas in and around each city within the specific radius. The flowchart of this calculation process is shown in Figure S2.

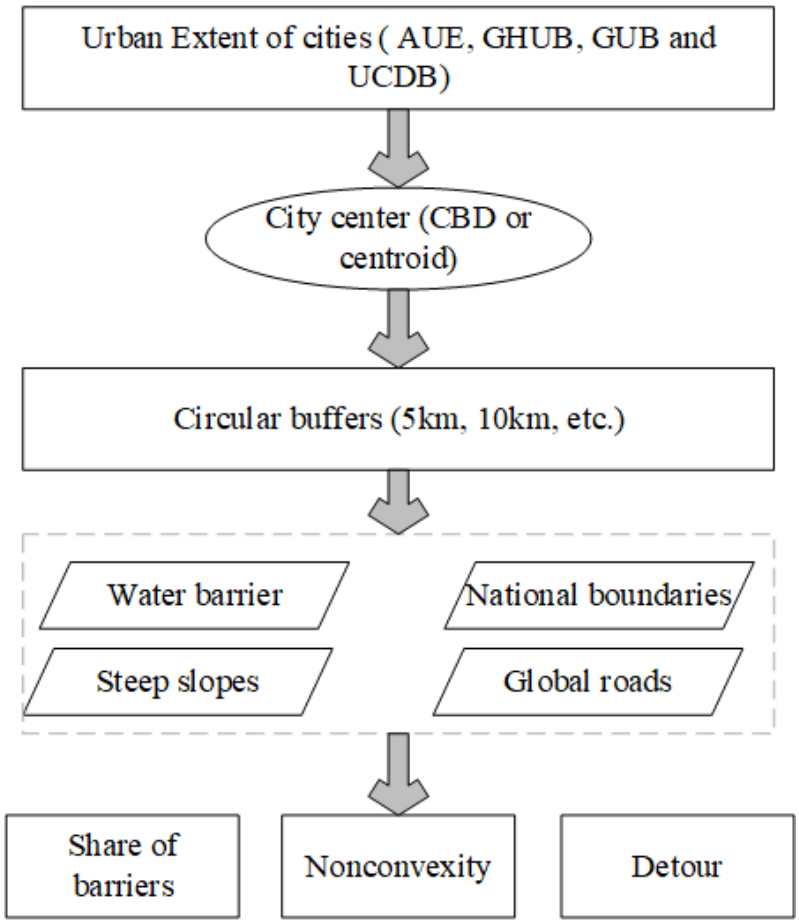

Figure S2. Flowchart for the calculation process of geographic indexes

2.2 Choice of radius to capture the environment around cities.

For our study, we needed to establish the radiuses used to estimate the surrounding urban barriers around each urban unit. We have chosen radiuses of 5km and 10km to ensure a uniform measurement. As indicated in Table S3-1, a 10km buffer exceeds the urban area of more than 96% of the urban units in the three global datasets (UCDB, GHUB, and GUB). Similarly, a 5km buffer encompasses an area equivalent to the urban extents of nearly 90% of cities in the UCDB and GUB datasets, and 85% of cities in the GHUB dataset. This indicates that these radiuses, are sufficiently large to cover most urban areas and yield consistent measurements of the surroundings of these cities.

Table S3-1. Distribution of urban areas by definition and comparison with area of 5km/10km buffer.

Indices	UCDB (km^2)	GHUB (km^2)	GUB (km^2)
Percentiles1	1	9	5
Percentiles5	3	10	5
Percentiles10	4	11	5
Percentiles25	8	15	6
Percentiles50	18	24	10
Percentiles75	37	49	22
Percentiles90	85	116	57
Percentiles95	156	224	112
Percentiles99	586	879	541
Area range	1 - 6622	1-7255	1 - 8293
Mean	50	70	39
Standard deviation	189	228	184
Fraction of areas larger than a 10 km buffer (314 km2)	2.2% (299 cities)	3.4% (349)	1.7% (244)
Fraction of areas larger than a 5 km buffer (79 km2)	10.7% (1416 cities)	15.1% (1540)	7.1% (1015)
Observations	13135	10169	14098

We use fixed buffers rather than the urban shapefiles themselves because the shape of cities is affected by geographic barriers not only within the city but also by those surrounding it. Relying exclusively on the ex-post shape of urban areas could result in a loss of information about the surrounding external barriers that may restrict a city's expansion space or direction, potentially impacting its socio-economic development. More importantly, the current urban footprint itself is shaped and constrained by surrounding geographic barriers. Therefore, using geographic indices that focus only on urban shapefiles—as currently developed—introduces direct endogeneity into the measurements.

In order to illustrate how we calculate our geographic variables, we select the city of Lugano, Switzerland, which is located in Switzerland, bordering Italy. It suffers impacts from all the three kinds of barriers, as shown in Figure S3. We exclude these barriers from its 10 km buffer and the left part (gray area) is then used to calculate the three geographic indexes.

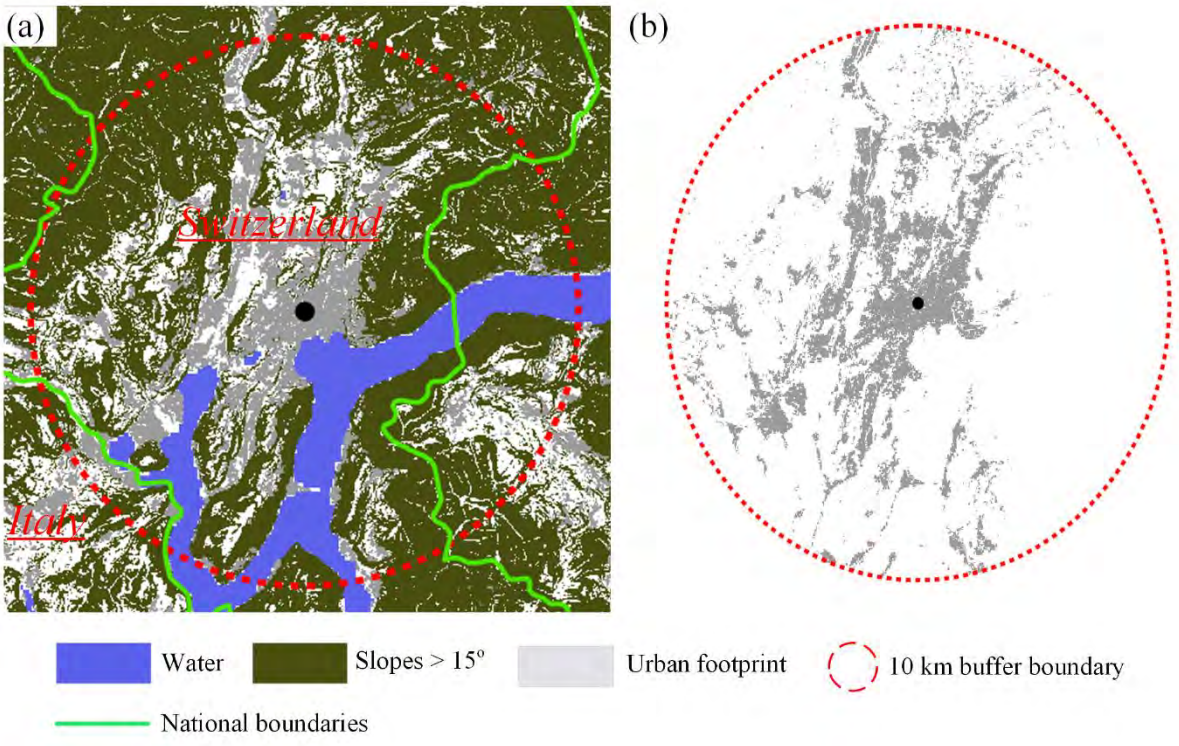

Figure S3. (a) The location of Lugano and its surrounding geographical environment. The blue areas denote water bodies, brown/green areas denote steep slopes, and the bright green line denotes the national boundary between Italy and Switzerland. The grey area denotes the build-up area within 10 km of city center (black point). **(b)** The built-up area of Lugano within 10 km of city center (black point) represented in grey. If we studied it in isolation, one might erroneously attribute its urban fragmentation to human agency.

2.3 Calculation of the geographic indexes

(1) Share of barriers

The Share of barriers index quantifies the proportion of the three types of geographic barriers relative to the total area within a *i* km (10km or 5 km) radius from the city center, with simply definition in Table S6. Importantly, this calculation is independent of the actual city shape, which is itself endogenous to the surrounding environment. In our methodology, the share of barriers within an *i* km distance from each urban center is determined by the following calculation:

$$S_i = \frac{\sum P_i + \sum W_i + \sum \hat{B}_i}{\pi \times i^2} \quad (1)$$

where W_i and P_i denote the areas of water coverage and steep slope areas located within the *i* km circle around the city center, respectively. $\hat{B}_i$ represents the area outside the national boundary excluding its overlaps with mountains and water. Finally, S_i denotes the share of barriers index of the urban unit.

Table S3-2. Simply definition of the three geographic indexes

Geographic Indexes	Definition	Value Range
Share of barriers	Proportion of geographic barriers within a *i* km (10km or 5 km) radius from the city center	0 – 1, with smaller values representing less geographic barriers
Nonconvexity	Proportion of straight lines intersected by geographic barriers. The straight lines are connections between each pair of points located within the non-barrier area around the *i* km buffer.	0 – 1, with smaller values representing higher connectivity in and around cities.
Detour	Proportion of minimum additional driving distances through street/ road network connecting the origin and destination points, which are located within the non-barrier area around the *i* km buffer.	0 – 1, with smaller values representing less additional driving distances compared to Euclidean distance.

(2) Nonconvexity

The nonconvexity index is designed to assess the impact of geographic barriers on the straight-line connectivity between various locations in and around the urban area. To calculate this index, we initially generate a set of points, evenly distributed across the internal circle footprint—this excluding the points that fall within high-sloped areas, foreign countries, or on water. These points are spaced at intervals greater than 0.5 km for the 10 km radius (0.25 km for the 5 km radius), to ensure comprehensive coverage of the study area. Subsequently, we create straight lines connecting each pair of points. These lines are then overlaid with the geographic barriers within the designated *i* km buffer zone. This process allows us to calculate the length of each line that intersects with barriers. We then determine the proportion of intersected length for each line. The average proportion of line lengths intersected by barriers is calculated to establish the nonconvexity index for each urban area.

$$NC = \frac{1}{n^2} \sum \sum \frac{SLDI_{ij}}{SLD_{ij}} \quad (2)$$

where SLD_{ij} denotes the length of straight line distance between points *i* and *j*, $SLDI_{ij}$ denotes the length of the line between points *i* and *j* intersected by geographic barriers, *n* denotes the number of internal points located within the radius at an interval of 5 or 10 km from the city center, and *NC* denotes the nonconvexity

index. Note that n^2 captures the total number of dyadic relationships included in the calculations.

(3) Detour

The detour index is designed to estimate the minimum additional driving distances incurred due to both geographic barriers and/or the layout of the street/road network. To calculate this index, we start by randomly generating points within the urban area at 0.5 km intervals from the 10 km radius (0.25 km for the 5 km radius), excluding areas classified as barriers. For each pair of points, we calculate both the Euclidean distance (the straight-line distance) and the minimum road distance. The minimum road distance is defined as the length of the shortest possible route via roads (such as streets, bridges, and tunnels) connecting the origin and destination points. The difference between the straight-line distance and the minimum road distance represents the detour for each pair of points. The detour index for a city is then determined by calculating the average detour across all pairs of points:

$$detour_{ij} = \frac{d_{ij}}{D_{ij}} - 1 \quad (3)$$

$$Detour = \frac{\sum detour_{ij}}{n^2} \quad (4)$$

where d_{ij} is the minimum road distance between points i and j, D_{ij} is the Euclidean distance between them, and $detour_{ij}$ is the detour index between the two points. n represents the number of points within 10 km (or 5 km) of the city center and Detour represents the average detour index of each city. It is noted that we have excluded points located 0.5 km away from any roads, to reduce the impacts of exurban areas where the distribution of roads is extremely parse. Illustrations about the calculation process for the three indexes are shown in Figures S4-S6, respectively. And the final distribution of geographic indicators is shown in Figure S7-S19.

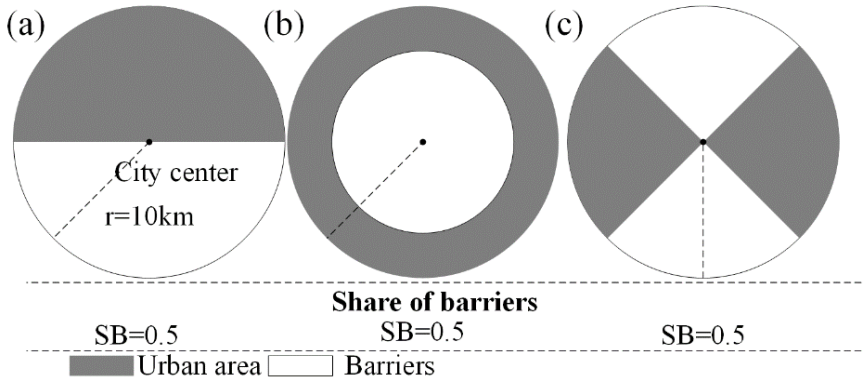

Figure S4. Three hypothetical urban forms with same *share of barriers* (0.5) and same city size, but different convexity properties (the grey area denote the urban area and blank areas denote geographic barriers). **(a)** Semicircular urban form, where geographic barriers occupy the bottom half of the disk: nonconvexity here is zero. **(b)** Annulus urban form, where geographic barriers are concentrated around the disk's center: this form displays the highest non-convexity. **(c)** Fan-shaped urban form, where geographic barriers are symmetrically distributed between the north and south parts of the city: large non-convexity.

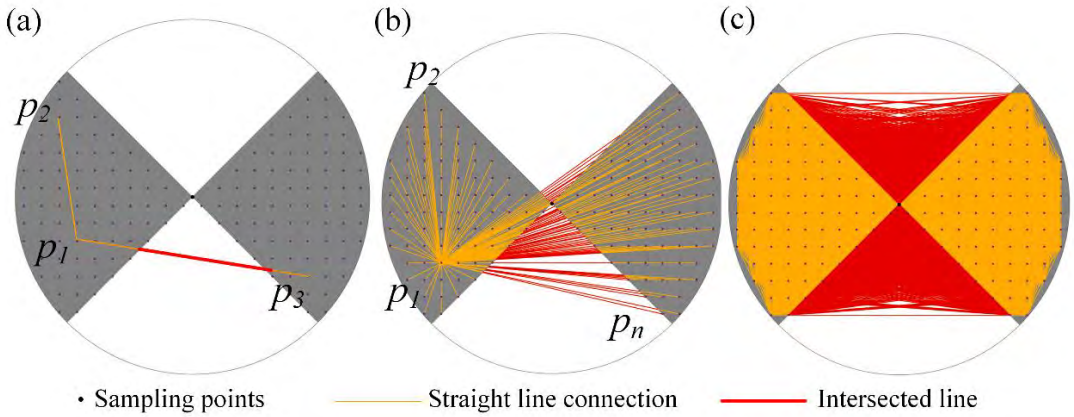

Figure S5. The process for the calculation of the dyadic nonconvexity index in a fan-shaped urban form (grey denotes urban areas and white geographic barriers). **(a)** Segments between the points P_1 and P_2, P_3. P_1 is located within the same sector as P_2 , whereas P_3 is in the opposite sector. Therefore, the straight-line connection between P_1 and P_3 is impeded by geographic barriers, and the impeded segments are shown in red. **(b)** The intersection between point P_1 and all other points (P_n) within the urban area. **(c)** Intersections between each pair of points within the urban unit.

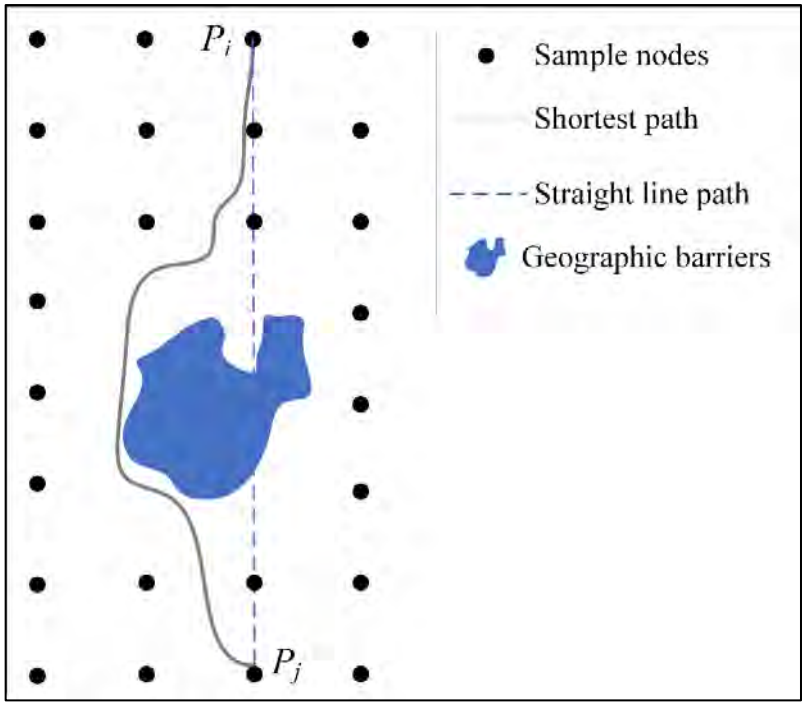

Figure S6. A hypothetical region containing geographic barriers, and a representation of a road. Sample nodes: the representative nodes distributed evenly in each city at 0.5 km spatial intervals for 10 km radius (p_i and p_j). Straight-line path: the straight-line connection between each pair of sample nodes. Shortest path: the road path with the shortest distance between each pair of sample nodes. Geographic barriers: the geographic barriers in and around the urban center.

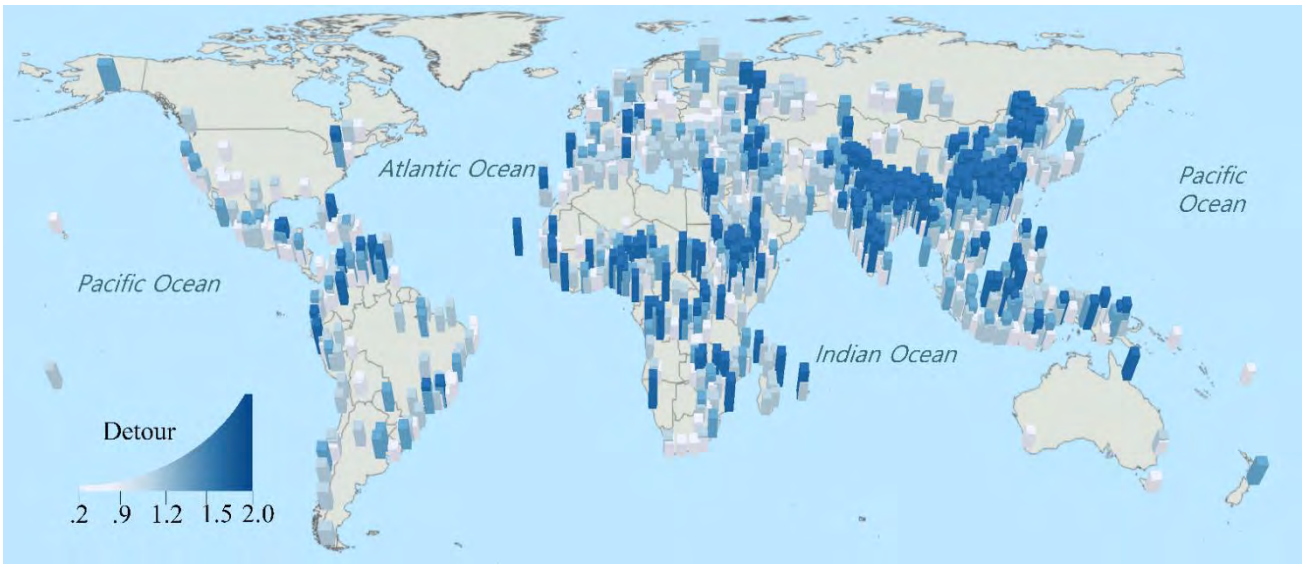

Figure S7-1. The global *detour* index of cities in UCDB dataset. The color and height of the bars represent its value.

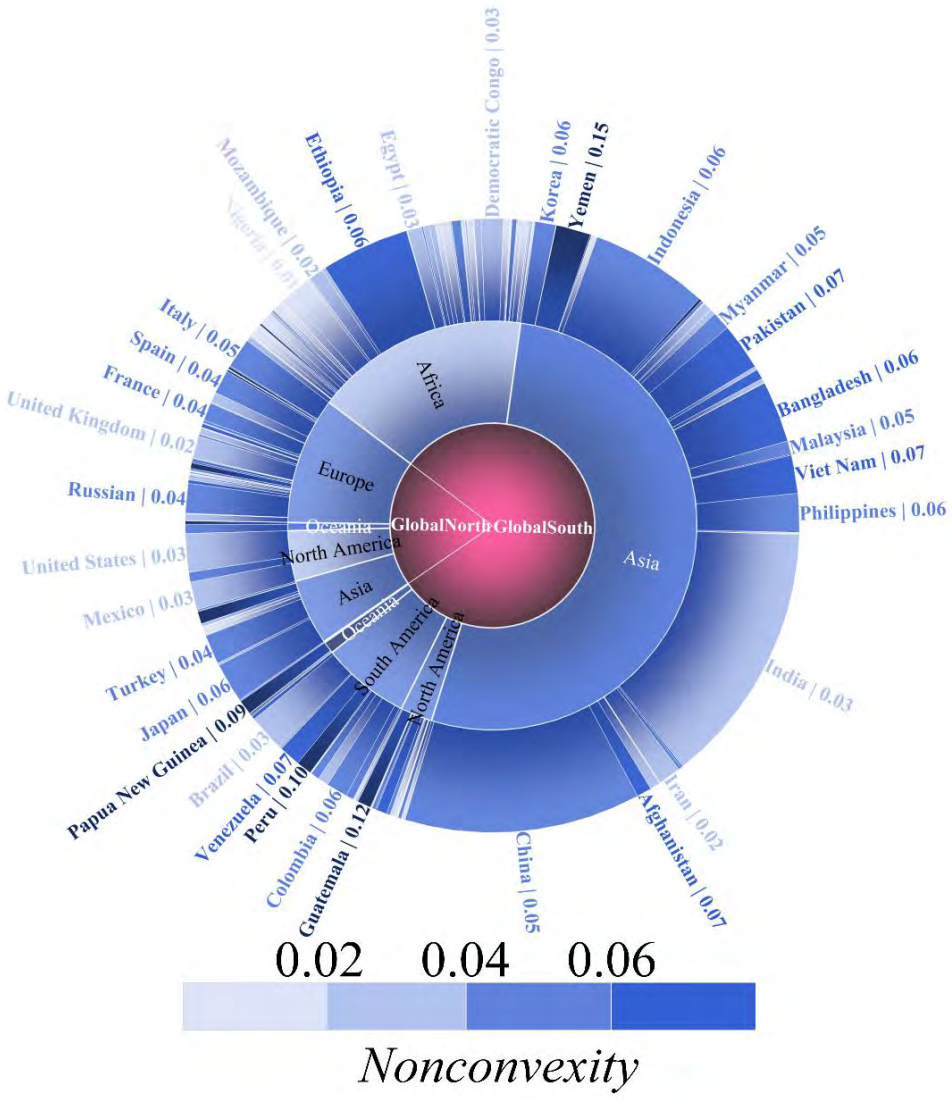

Figure S7-2. Average country-level nonconvexity index in and around urban areas obtained from the UCDB dataset.

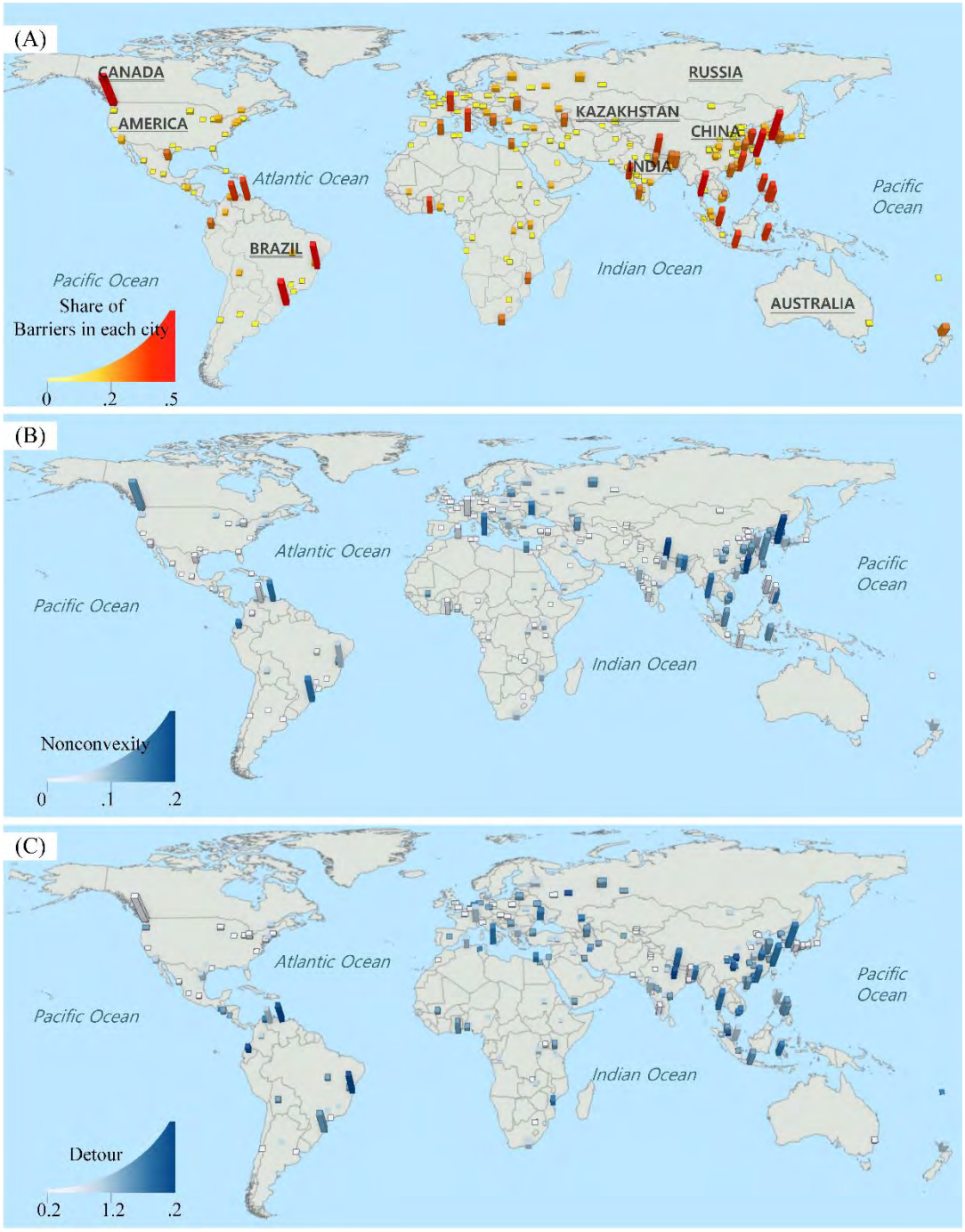

Figure S8. Geographic indexes in AUE dataset. **(A)** The global *share of barriers* in and around cities. **(B)** Global *nonconvexity* in and around cities. **(C)** Global *detour* in and around cities. Color and height of the bar represent the value of the indexes.

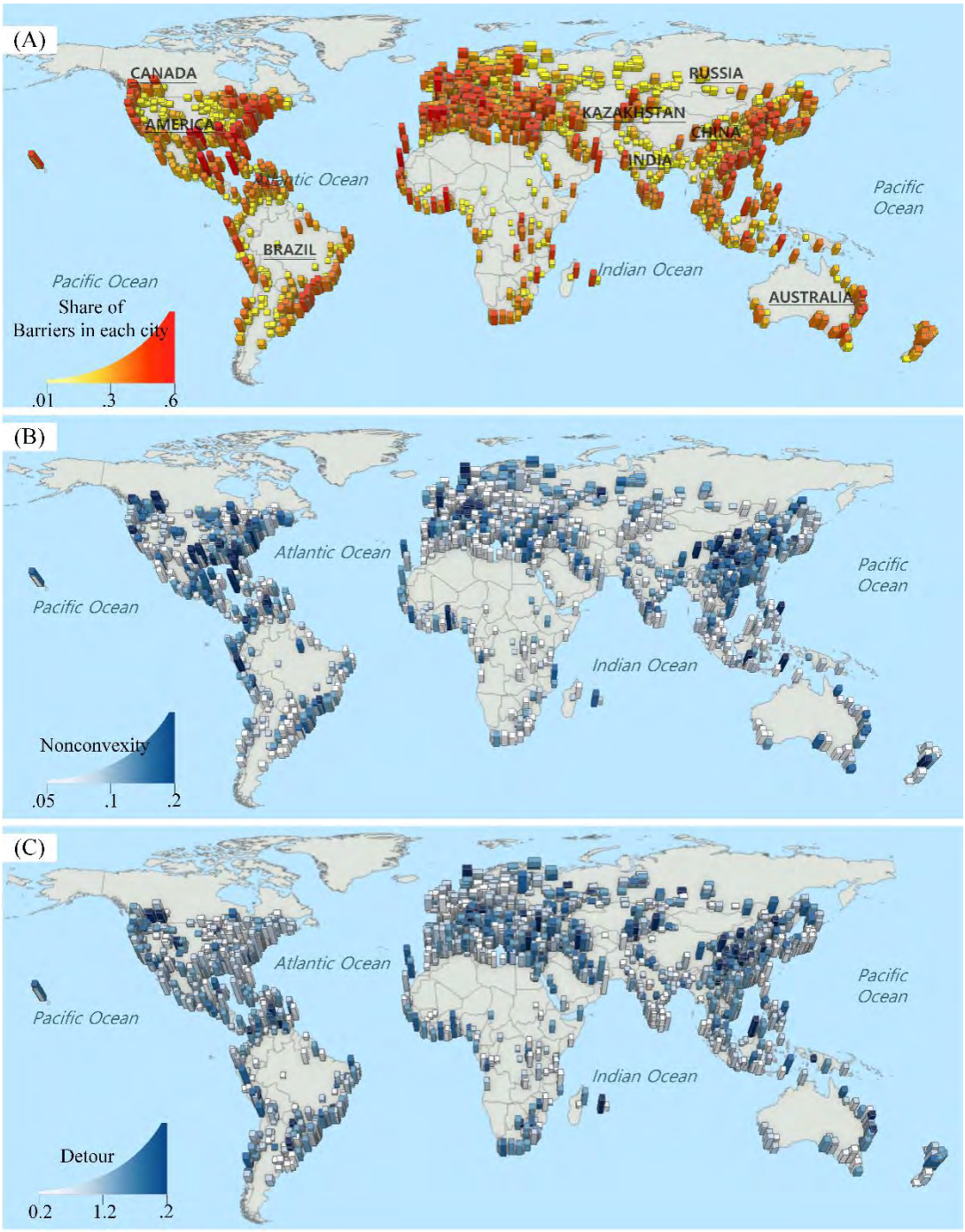

Figure S9. Geographic indexes in GHUB dataset. **(A)** Global *share of barriers* in and around cities. **(B)** Global *nonconvexity* in and around cities. **(C)** Global average road detour in and around cities. The color and height of the bars represent the value of the index. Cities with *nonconvexity* < 0.05 are undisplayed.

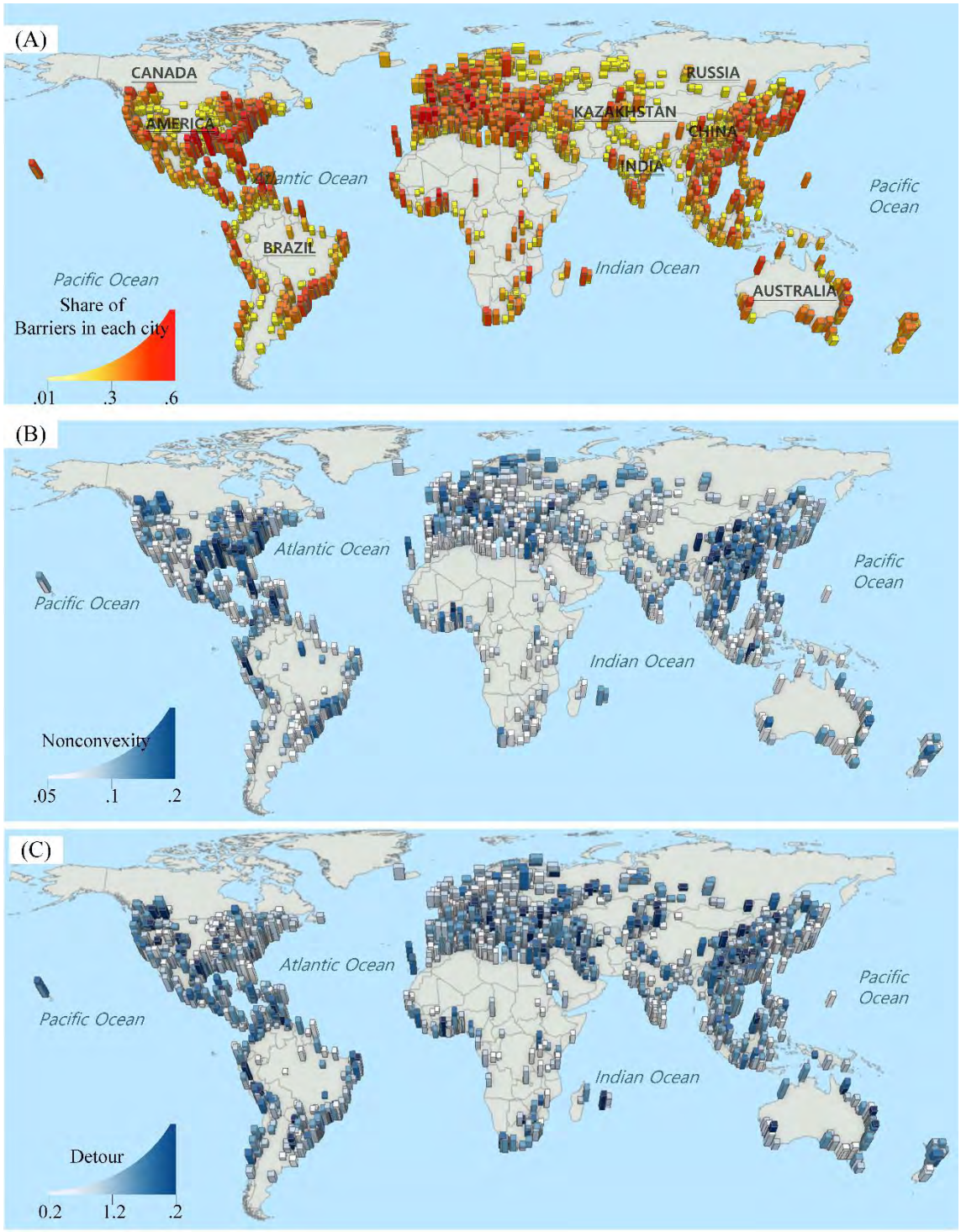

Figure S10. Geographic indexes in GUB dataset. **(A)** Global *share of barriers* in and around cities. **(B)** Global *nonconvexity* in and around cities. **(C)** Global road *detour* in and around cities. Color and height of the bars represent the value of the index, and cities with *nonconvexity* < 0.05 are undisplayed.

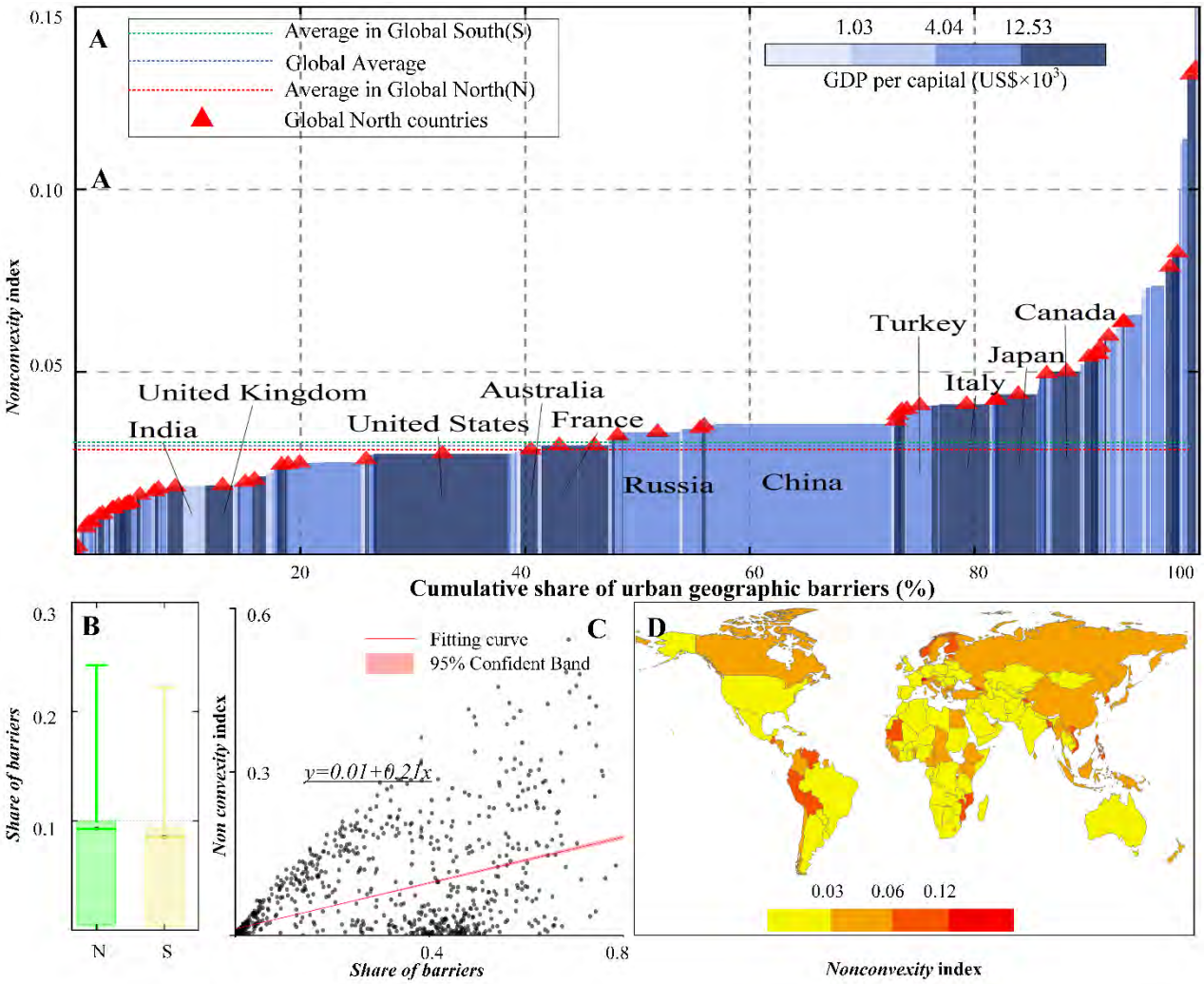

Figure S11. Country-level nonconvexity in the GHUB dataset. **(A)** Average *nonconvexity* index of each country. The dashed lines with red, green, and blue colors denote the average *nonconvexity* index in, respectively, the Global North countries (0.029), Global South countries (0.031) and all countries (0.030). The bar width represents the portion of urban geographic barriers in each country relative to the global total. Global North countries are marked with red triangles. **(B)** Comparison of the average urban share of barriers in Global North and Global South countries (N: Global North, S: Global South). **(C)** Relationship between the average urban nonconvexity index by country and the corresponding share of barriers. **(D)** Distribution of the average urban nonconvexity index across global countries.

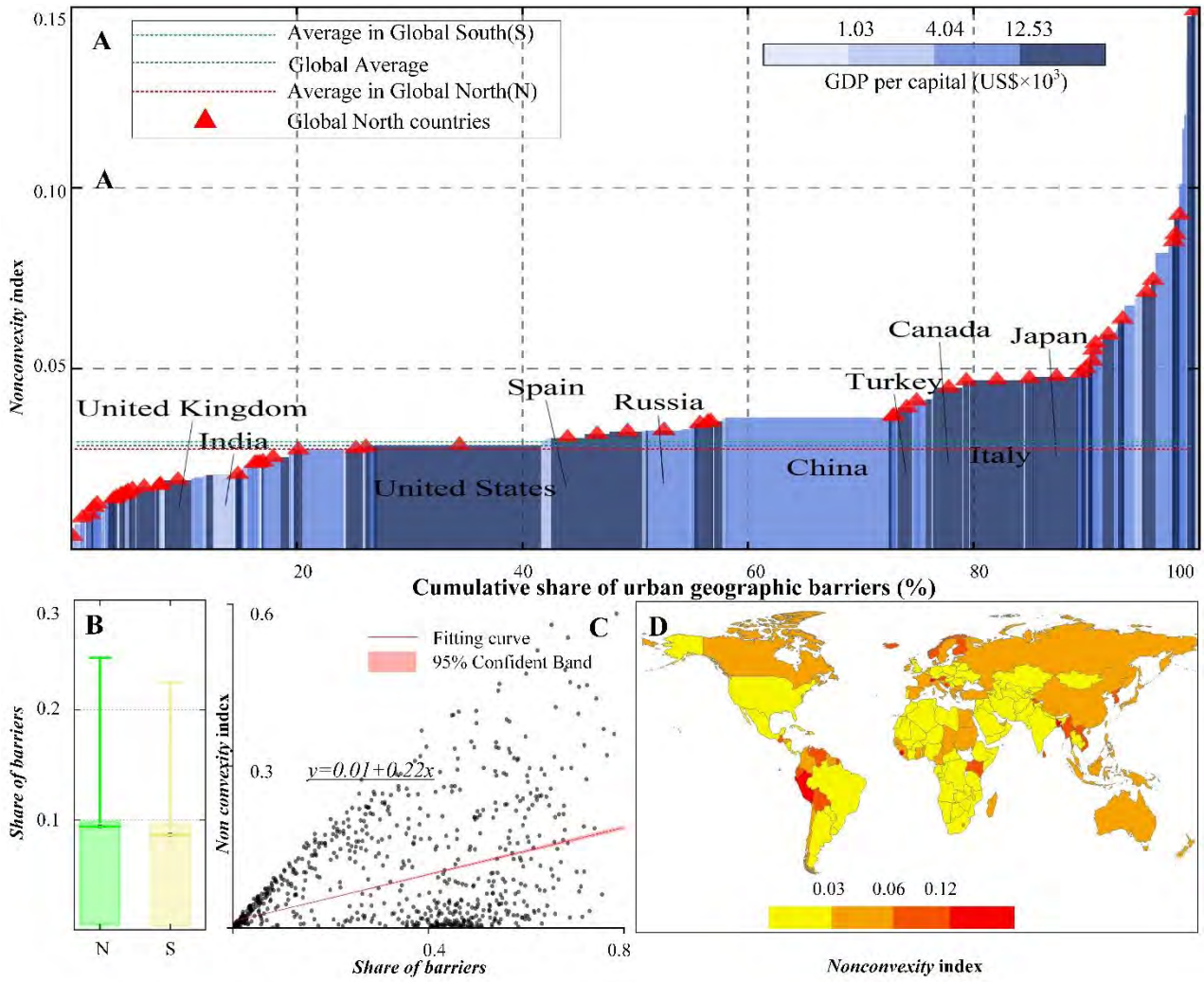

Figure S12. Country-level nonconvexity in the GUB dataset. **(A)** Average *nonconvexity* index in each country. The dashed lines with red, green, and blue colors denote the average *nonconvexity* index in, respectively, the Global North countries (0.029), Global South countries (0.031) and all countries (0.030). The bar width represents the portion of urban geographic barriers in each country relative to the global total. Global North countries are marked with red triangles. **(B)** Comparison of the average urban share of barriers in Global North and Global South countries (N: Global North, S: Global South). **(C)** Relationship between the average urban nonconvexity index by country and the corresponding share of barriers. **(D)** Distribution of the average urban nonconvexity index across global countries.

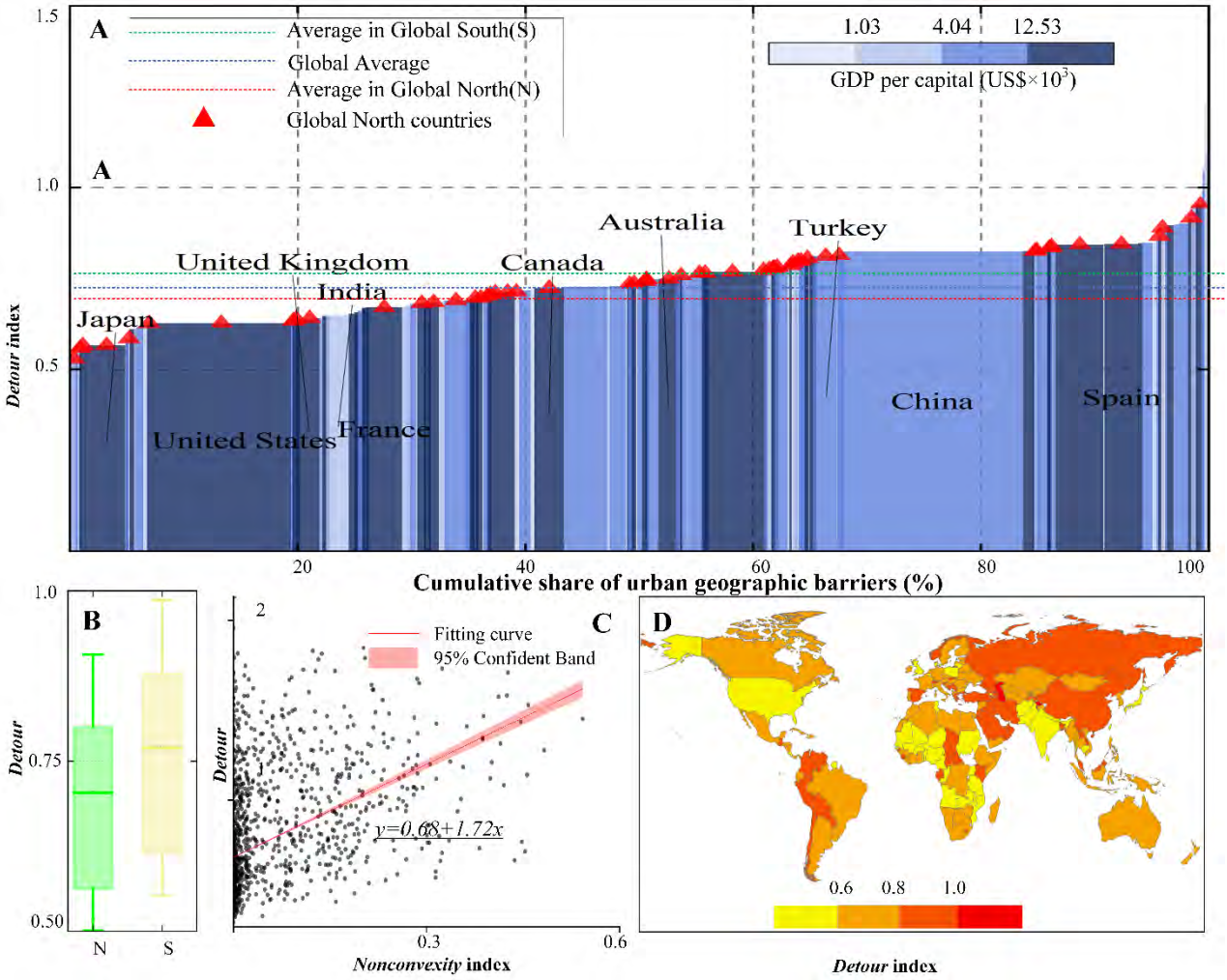

Figure S13. Country-level detour in the GHUB dataset. **(A)** Average *detour* index of each country. The dashed lines in red, green, and blue colors denote the average *detour* index of, respectively the Global North (0.70), Global South (0.77) and all countries (0.73). The bar width represents the portion of urban geographic barriers of each country to global urban barriers. Global North countries are marked with red triangles. (B) Comparison of the average *detour* index across Global North and Global South countries. (C) Relationship between country-level *nonconvexity* index and *detour* index. (D) Distribution of the *detour* index across global countries

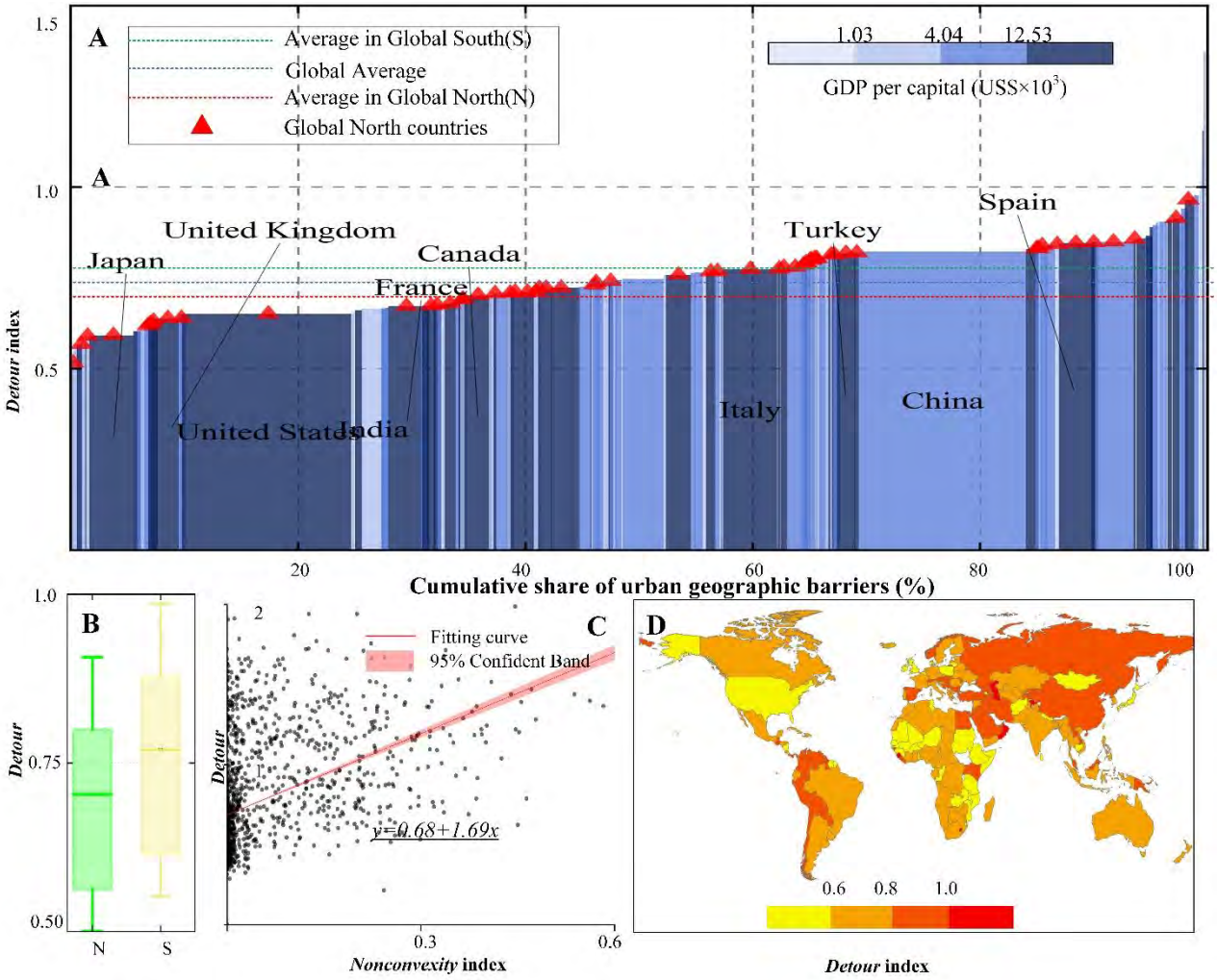

Figure S14. Country-level detour in GUB dataset. **(A)** Average *detour* index of each country. The dashed lines in red, green, and blue colors denote the average *detour* index of, respectively the Global North (0.70), Global South (0.77) and all countries (0.73). The bar width represents the portion of urban geographic barriers of each country to global urban barriers. Global North countries are marked with red triangles. (B) Comparison of the average *detour* index across Global North and Global South countries. (C) Relationship between country-level *nonconvexity* index and *detour* index. (D) The distribution of the *detour* index across global countries.

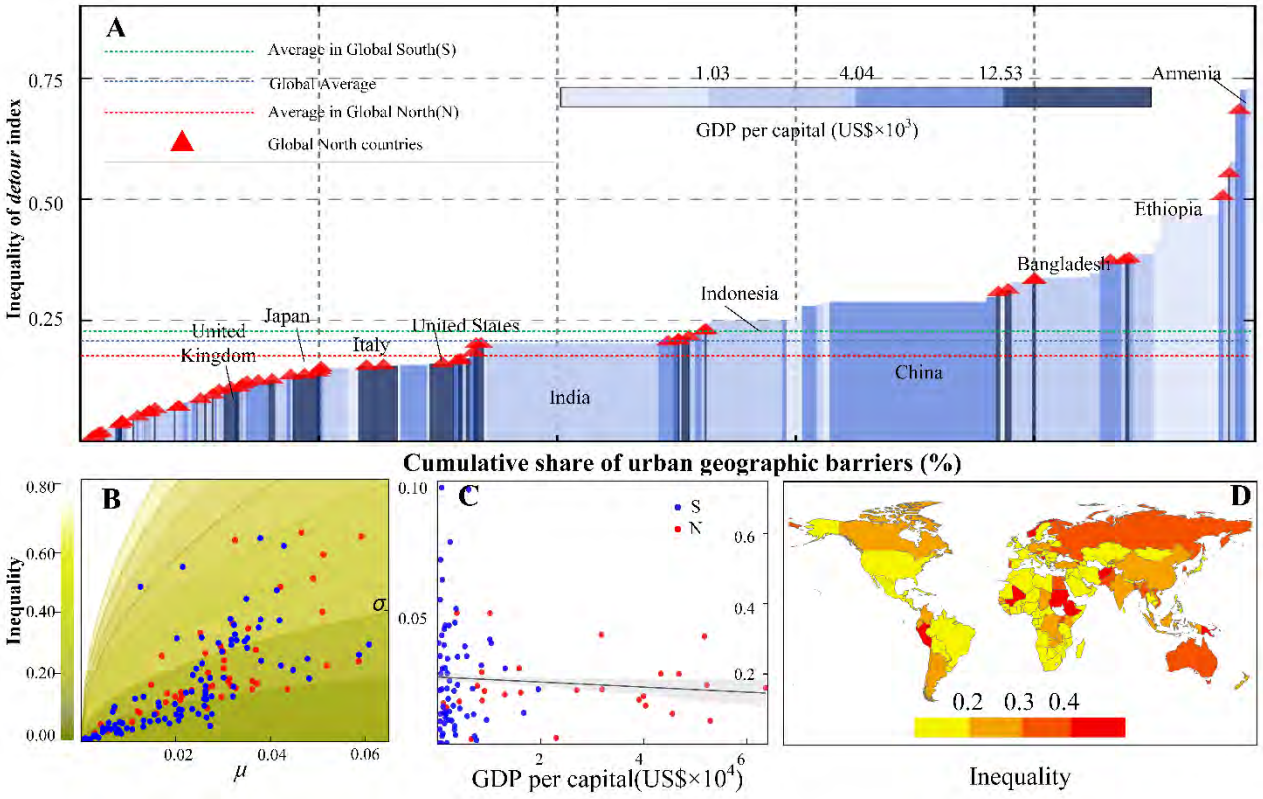

Figure S15. Measurement of road detour inequalities in the UCDB dataset (see note for calculations). **(A)** Inequality of the *detour* index within each country. The dashed lines in green, red, and blue colors denote the average inequality within Global North (0.14), Global South (0.23) and all countries (0.20), respectively. The bar width represents the portion of urban detour inequality of each country as a share of global urban barriers inequality. Global North countries are marked with red triangles. **(B)** The inequality of the *detour* index in Global South and North countries derived from means (μ) and SDs (σ) of the *detour* index. **(C)** The relationship between GDP per capita (GDPPC) and the inequality in the *detour* index. **(D)** Distribution of inequality in the *detour* index across worldwide countries.

Note: The inequality index measuring the spatial differences in cities' detour values within each country is computed through $Inequality = \frac{\sigma}{\sqrt{\mu(1-\mu)}}$, referring to (Pandey et al., 2022), where μ and σ represent the mean and SD of the *detour* indexes in the cities of each country. The inequality value ranges from 0 to 1, with a larger value indicating a higher degree of inequality in that country. The results suggest that economic development alone may not resolve the issue of infrastructure inequality. This finding is consistent with other recent research, which posits that urban infrastructure inequalities are an inherent aspect of urbanization (Pandey et al., 2022).

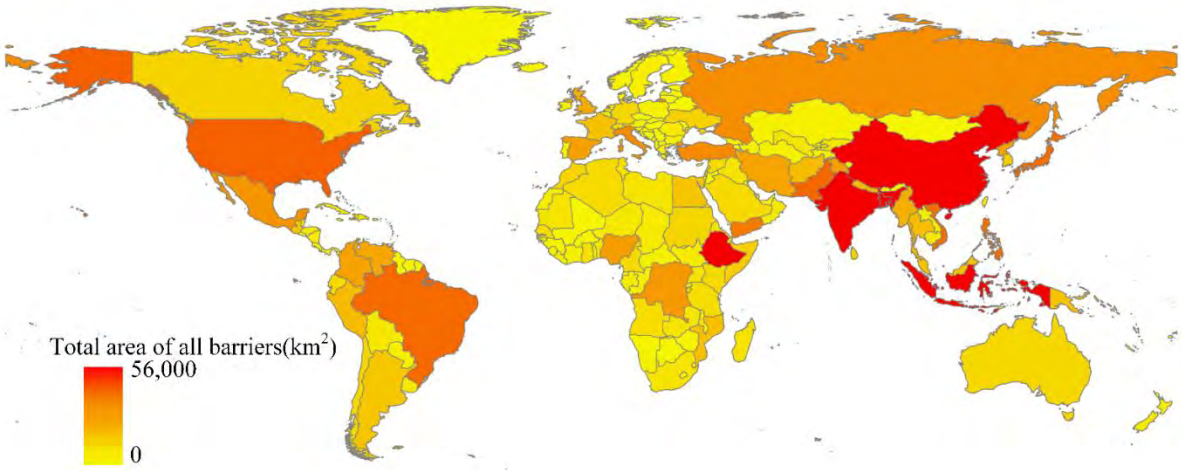

Figure S16. Distribution of total area of urban barriers (UCDB dataset).

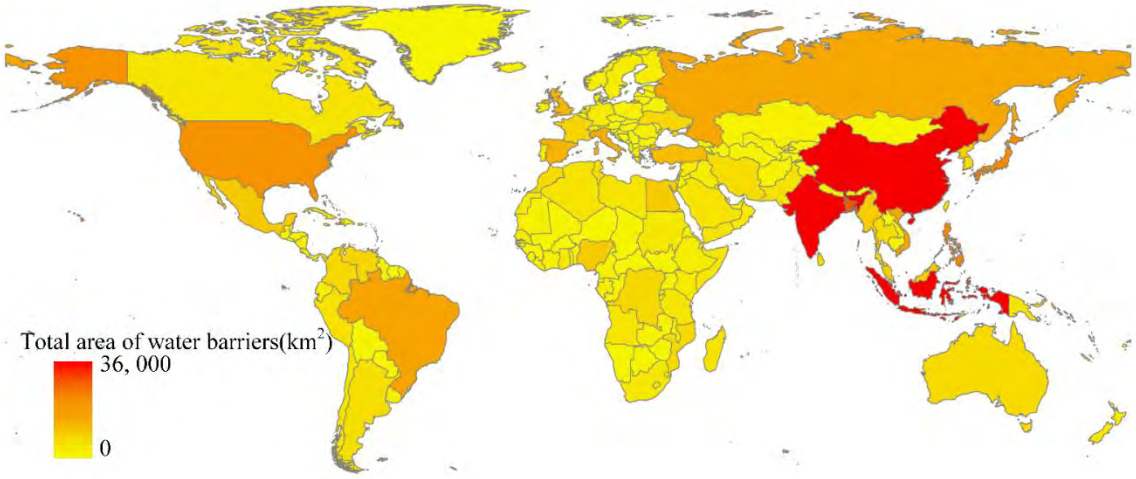

Figure S17. Distribution of the total area of urban water barriers (UCDB dataset).

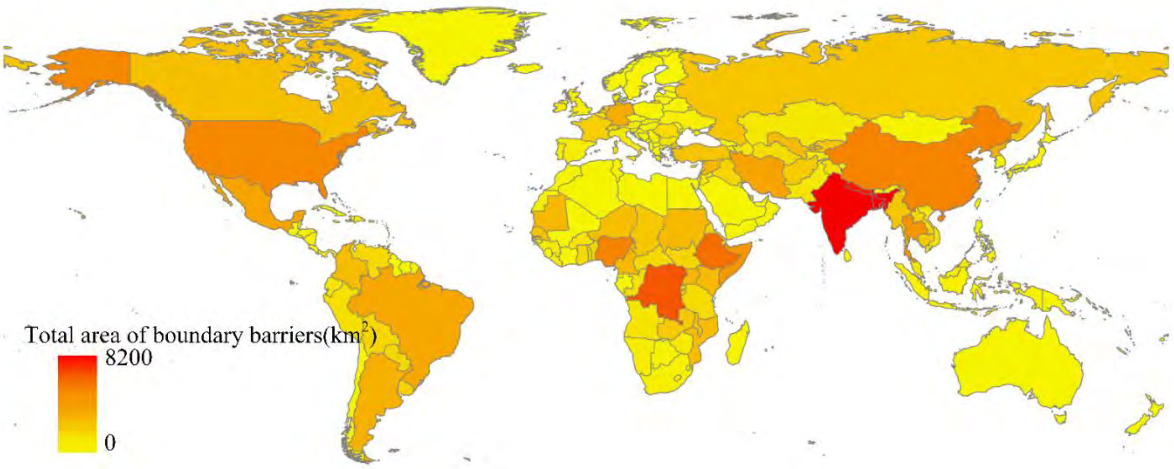

Figure S18. Distribution of the total area of urban national-border barriers (UCDB dataset).

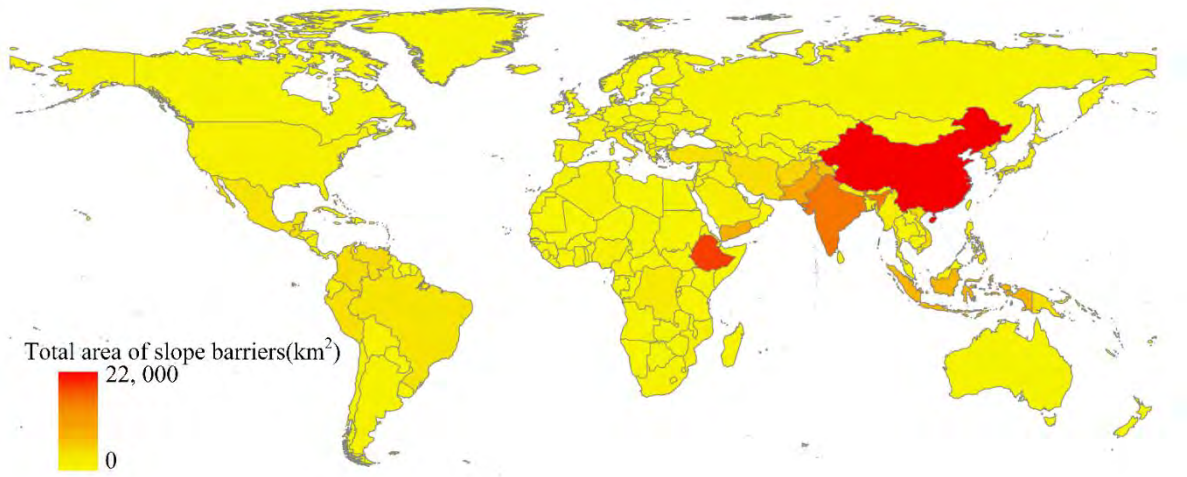

Figure S19. Distribution of the total area of slope barriers (UCDB dataset).

Section 3: Regression analysis

3.1 Description of variables

Table S4-1. Dimensions and city-level variables in adopted in regressions.

Dimension	Variable	Description
Basic information	Area	Area of the Urban Centre (km2)
	Name	Assigned name of the Urban Centre
	Unique ID	Urban Centre unique ID and quality control code
	Country	The country of the city
Socioeconomic variables-all (Four datasets have these variables)	Temperature	Average temperature for epoch 2015 (℃)
	Precipitation	Average Precipitation for epoch 2015 (mm)
	Resident population	Total resident population in 2015
	Built-up surface	Total built-up area in 2015 (km^2)
	Gross Domestic Product	Sum of GDP PPP values for year 2015 ($)
	Dummy: City not in Major River Watershed	If the city is in river basins (yes:0, no:1)
	Dummy: Capital city	If the city is capital city (yes: 1, no: 0)
	Dummy: Ocean or Lake Intersection	If the city is costal city (yes: 1, no: 0)
	Build-up volume	Total build-up volumes within the urban unit (m^3)
	Soil type	Major soil type in the city
	Climate	Major climate type in the city
	Biome	Major biome type in the city
Socioeconomic variables-UCDB (Only UCDB dataset has these variables)	Greenness value	Average greenness estimated for 2015
	Residential CO_2 emission	Total emission of CO_2 from the residential sector in 2015 (tonnes/year)
	Transportation CO_2 emission	Total emission of CO_2 from the transport sector in 2015 (tonnes/year)
	Nighttime light emission	Average nighttime light emission in 2015 ($nW\ cm^{-2}sr^{-1}$)
Geographic indexes	Share of barriers	Share of geographic barriers within 10 km (5km) buffer
	Nonconvexity	Nonconvexity index at 10 km (5km) buffer
	Detour	Detour index at 10 km (5km) buffer

Table S4-2. Country-level variables used in this study (for the UCDB dataset).

Sources	Category	Indicator
World Development Indicators (Bank, 2016)	Economy & Growth	GDP per capita, PPP (current international $)
		General government final consumption expenditure (% of GDP)
	Population	Population ages 0-14 (% of total population)
	Education	Literacy rate, adult female (% of females ages 15 and above)
	Urban Development	Urban population (% of total population)
Freedom Indicator (House, 2022)	Political democracy	Political Rights index Civil Liberties index

Table S4-3. Description of city-level variables in the UCDB-boundary dataset

Variables	Obs.	Mean	Std. Dev.	Min	Max
Average Elevation (m)	13,131	422.7	596.7	1	4,510
Average Precipitation for Epoch 2015 (mm)	13,135	1,108	665.9	0	6,282
Average Temperature for Epoch 2015 (°C)	13,135	21.00	6.667	-8.735	30.85
Average Night Time Light Emission in 2015 (nW $cm^{-2}sr^{-1}$)	13,135	11.64	16.23	0	312.2
Average greenness	12,545	0.409	0.111	0.0246	0.773
Residential CO_2 Emissions non-organic 2015 (tonnes/year)	13,133	99,938	802,334	0.0864	3.208×10^7
Transportation CO_2 Emissions non-organic 2015 (tonnes/year)	12,843	97,863	652,237	0.000202	2.504×10^7
Share of barriers within 10km of City Center	13,134	0.0955	0.152	0	0.991
Share of barriers within 5km of City Center	13,133	0.0708	0.128	0	0.918
Average detour within 10km of City Center	13,093	0.761	0.424	0.00332	5.788
Average detour within 5km of City Center	13,062	0.793	0.642	0.00475	13.38
Average Local Non-convexity within 5km of City Center	13,134	0.0299	0.0626	0	0.791
Average Local Non-convexity within 10km of City Center	13,134	0.0390	0.0693	0	0.788
Build-up volume (m^3)	13,135	8.883×10^7	4.464×10^8	0	1.970×10^{16}
Sum of GDP PPP value in 2015 ($)	13,135	4.053×10^9	2.853×10^{10}	0	1.400×10^{12}
Total built-up area in 2015 (km^2)	13,135	9.470	40.72	0	1,526
Total Resident Population in 2015	13,135	249,734	1.086×10^6	0	3.843×10^7

Table S4-4. Description of city-level variables in the AUE dataset

Variables	Obs.	Mean	Std. Dev.	Min	Max
Total Resident Population in 2014	200	3.936×10^{6}	5.735×10^{6}	97,888	3.477×10^{7}
Share of build-up area within urban unit	200	0.657	0.0759	0.480	0.840
Total built-up area in 2014 (km^2)	200	530.6	948.2	4.450	7,479
Density of Resident Population (people/ km^2)	200	10,056	8,243	1,274	55,151
Total area of urban extend (km^2)	200	786.9	1,345	8.490	9,511
Average Elevation (m)	200	373.3	574.2	1	3,071
Average Precipitations in 2015 (mm)	200	1,101	648.4	36.48	3,264
Average Temperature in 2015 (°C)	200	18.64	6.717	-0.356	30.34
Sum of GDP PPP value in 2015 ($)	200	9.581×10^{10}	1.954×10^{11}	6.472×10^{7}	1.43×10^{12}
Share of barriers within 5km of City Center	200	0.0413	0.0810	0	0.492
Average Local Non-convexity within 5km of City Center	200	0.0136	0.0275	0	0.193
Share of barriers within 10km of City Center	200	0.0752	0.118	0	0.565
Average Local Non-convexity within 10km of City Center	200	0.0273	0.0439	0	0.330
Average detour within 5km of City Center	200	0.639	0.169	0.327	1.264
Average detour within 10km of City Center	200	0.618	0.177	0.348	1.255

Note: We have transformed units of some variables from *ha* (in original AUE dataset) to km^2.

Table S4-5. Description of city-level variables in the GHUB-boundary dataset

Variables	Obs.	Mean	Std. Dev.	Min	Max
Average Precipitations in 2015 (mm)	10,241	919.2	556.1	0	3,697
Average Temperature in 2015 (℃)	10,241	15.25	6.664	-3.854	30.57
Average Elevation (m)	10,241	339.2	473.7	1	4,576
Total Resident Population in 2015	10,241	244,256	960,865	0	2.622×10^{7}
Total Built-up area within urban unit of 2015 (km^2)	10,241	12.67	44.15	0	1,751
Total Build-up area volume (m^3)	10,241	1.12×10^{8}	4.239×10^{8}	0	1.592×10^{10}
Sum of GDP PPP value in 2015 ($)	10,241	5.362×10^{9}	3.592×10^{10}	0	1.469×10^{12}
Share of barriers within 5km of City Center	10,169	0.0639	0.113	0	0.927
Average Local Non-convexity within 5km of City Center	10,169	0.0211	0.0362	0	0.463
Share of barriers within 10km of City Center	10,169	0.0890	0.144	0	0.972
Average Local Non-convexity within 10km of City Center	10,169	0.0306	0.0484	0	0.543
Average detour within 5km of City Center	10,165	0.728	0.213	0.0893	2.438
Average detour within 10km of City Center	10,168	0.734	0.212	0.294	2.134

Table S4-6. Description of city-level variables in the GUB dataset

Variables	Obs.	Mean	Std. Dev.	Min	Max
Average Elevation (m)	14,098	323.6	463.4	1	4,576
Average Precipitations in 2015 (mm)	14,098	935.9	556.8	0.386	9,729
Average Temperature in 2015 (℃)	14,098	15.30	6.395	-4.358	30.57
Total Built-up area within urban unit in 2015 (km^2)	14,098	9.003	40.27	0	1,742
Total Build-up area volume (m^3)	14,098	8.075×10^7	4.248×10^8	0	2.143×10^{10}
Total Resident Population in 2015	14,098	179,929	935,326	0	4.171e+07
Sum of GDP PPP value in 2015 ($)	14,098	3.557×10^9	2.774×10^{10}	0	1.388×10^{12}
Share of barriers within 5km of City Center	14,098	0.0664	0.118	0	0.928
Average Local Non-convexity within 5km of City Center	14,098	0.0220	0.0383	0	0.461
Share of barriers within 10km of City Center	14,098	0.0906	0.147	0	0.972
Average Local Non-convexity within 10km of City Center	14,098	0.0317	0.0506	0	0.623
Average detour within 5km of City Center	14,095	0.748	0.220	0.250	2.444
Average detour within 10km of City Center	14,092	0.740	0.212	0.00305	2.204

Table S4-8. Association between country-level detour and GDPPC in UCDB-2023 datasets.

	Country-level Detour index within 10km of city center
Country-level nonconvexity index within 10km of city center	2.475***
	(9.54)
Log GDP per Capita in 2015	-0.041***
	(-3.92)
Observations	178
R-squared	0.36

t statistics in parentheses

* p<0.05, ** p<0.01, *** p<0.001

3.2 Detour regressions at 10 kilometers from the center, as In Table 2, for the other three dataset boundaries (GHUB, GUB and AUE)

Table S5-1. Impacts of geographic barriers on street/road patterns using GHUB boundaries

	Detour Index within 10 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 10km		1.728***		1.577***	1.425***
		(0.0400)		(0.0377)	(0.0463)
Share of barriers, 10km	0.384***		0.401***		0.0960***
	(0.0141)		(0.0145)		(0.0170)
Dummy of Coastal City			0.0413***	0.0398***	0.0356***
			(0.00522)	(0.00494)	(0.00499)
Dummy of Capital City			0.0174	0.0141	0.0157
			(0.0110)	(0.0105)	(0.0105)
City not in Major River Watershed			-0.0342***	-0.0203***	-0.0247***
			(0.00503)	(0.00476)	(0.00481)
Log of Average Precipitations in 2015			0.0228***	0.0132**	0.0128**
			(0.00583)	(0.00559)	(0.00558)
Log of Average Temperature in 2015			0.0556***	0.0621***	0.0657***
			(0.0161)	(0.0154)	(0.0154)
Log of Average Elevation			0.0259***	0.0202***	0.0214***
			(0.00182)	(0.00173)	(0.00174)
Log of GDP per capital			0.00203	0.000675	0.00153
			(0.00268)	(0.00256)	(0.00256)
Log of total Resident Population in 2015			-0.00541	-0.0139***	-0.0132***
			(0.00335)	(0.00321)	(0.00321)
Log of Build-up Area in 2015			-0.0469***	-0.0377***	-0.0381***
			(0.00408)	(0.00391)	(0.00391)
Constant	0.700***	0.681***	0.350***	0.497***	0.466***
	(0.00239)	(0.00229)	(0.0736)	(0.0703)	(0.0704)
Country Fixed Effects	NO	NO	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES
Observations	10,168	10,168	10,082	10,082	10,082
R-squared	0.068	0.155	0.394	0.445	0.447

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S5-2. Impacts of geographic barriers on street/road patterns using GUB boundaries

	Detour Index within 10 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 10km		1.689***		1.566***	1.460***
		(0.0324)		(0.0302)	(0.0376)
Share of barriers, 10km	0.372***		0.391***		0.0661***
	(0.0117)		(0.0118)		(0.0140)
Dummy of Coastal City			0.0376***	0.0344***	0.0317***
			(0.00441)	(0.00416)	(0.00419)
Dummy of Capital City			0.00660	0.000466	0.00224
			(0.00978)	(0.00929)	(0.00929)
City not in Major River Watershed			-0.0362***	-0.0225***	-0.0256***
			(0.00426)	(0.00400)	(0.00406)
Log of Average Precipitations in 2015			0.0245***	0.0148***	0.0144***
			(0.00520)	(0.00495)	(0.00494)
Log of Average Temperature in 2015			0.0365**	0.0421***	0.0449***
			(0.0144)	(0.0137)	(0.0137)
Log of Average Elevation			0.0237***	0.0180***	0.0189***
			(0.00155)	(0.00146)	(0.00147)
Log of GDP per capital			-0.00765***	-0.00724***	-0.00689***
			(0.00211)	(0.00200)	(0.00200)
Log of total Resident Population in 2015			-0.00435*	-0.0101***	-0.00987***
			(0.00262)	(0.00249)	(0.00249)
Log of Build-up Area in 2015			-0.0437***	-0.0360***	-0.0363***
			(0.00328)	(0.00313)	(0.00312)
Constant	0.706***	0.687***	0.473***	0.578***	0.561***
	(0.00203)	(0.00193)	(0.0634)	(0.0602)	(0.0603)
Country Fixed Effects	NO	NO	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES
Observations	14,092	14,092	13,957	13,957	13,957
R-squared	0.067	0.162	0.370	0.432	0.433

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S5-3. Impacts of geographic barriers on street/road patterns using AUE boundaries

	Detour Index within 10 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 10km		1.856***		1.545***	1.529***
		(0.254)		(0.240)	(0.313)
Share of barriers, 10km	0.551***		0.405***		0.0101
	(0.0993)		(0.103)		(0.126)
Dummy of Coastal City			0.0499*	0.0431	0.0427
			(0.0299)	(0.0276)	(0.0283)
Dummy of Capital City			0.0406	0.0446*	0.0445
			(0.0286)	(0.0269)	(0.0270)
City not in Major River Watershed			-0.0637**	-0.0406	-0.0411
			(0.0290)	(0.0267)	(0.0278)
Log of Average Precipitations in 2015			-0.00446	-0.00874	-0.00881
			(0.0149)	(0.0141)	(0.0141)
Log of Average Temperature in 2015			0.0229	0.0649	0.0645
			(0.0522)	(0.0497)	(0.0500)
Log of Average Elevation			0.00767	0.00727	0.00726
			(0.00832)	(0.00783)	(0.00785)
Log of GDP per capital			-0.00316	-0.00254	-0.00255
			(0.0130)	(0.0123)	(0.0123)
Log of total Resident Population in 2015			-0.00643	-0.0179	-0.0178
			(0.0181)	(0.0172)	(0.0173)
Log of Build-up Area in 2015			-0.0551***	-0.0425**	-0.0424**
			(0.0193)	(0.0184)	(0.0184)
Constant	0.576***	0.567***	0.891***	0.860***	0.860***
	(0.0138)	(0.0131)	(0.291)	(0.274)	(0.275)
Observations	200	200	200	200	200
R-squared	0.135	0.212	0.375	0.446	0.446

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

3.3 Full Table 3 parameters (Nonconvexity regressions) at 10 kilometers for all datasets (UCDB, GHUB, GUB and AUE)

Table S6-1. Impacts of geographic barriers on major socioeconomic outcome using UCDB boundaries.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-1.686***	1.575***	-0.133***	-0.904***	1.439***	-0.0970***	0.184***
	(0.174)	(0.0775)	(0.00763)	(0.187)	(0.0772)	(0.00721)	(0.0375)
Dummy of coastal city	0.400***	-0.101***	0.00316**	-0.0239	-0.0895***	-0.00592***	0.00636
	(0.0311)	(0.0138)	(0.00140)	(0.0319)	(0.0132)	(0.00123)	(0.00639)
Dummy of Capital City	2.986***	0.0167	0.0461***	-0.235**	-0.00980	-0.0204***	-0.160***
	(0.0943)	(0.0419)	(0.00427)	(0.0992)	(0.0410)	(0.00383)	(0.0199)
City not in Major River Watershed	0.0105	-0.0292**	0.00156	0.0734**	-0.0334***	0.00166	0.0379***
	(0.0283)	(0.0126)	(0.00128)	(0.0291)	(0.0120)	(0.00112)	(0.00583)
Log of Average Precipitation	0.0418*	-0.0408***	-0.000647	0.0928***	-0.0341***	-0.00188*	-0.00477
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.00521)
Log of Average Temperature	0.161	0.422***	-0.00576	-0.526***	0.384***	-0.00846**	0.124***
	(0.105)	(0.0468)	(0.00475)	(0.108)	(0.0446)	(0.00417)	(0.0217)
Log Average Elevation	0.00876	0.0312***	0.00126***	-0.0285***	0.0264***	0.00126***	0.0322***
	(0.0105)	(0.00469)	(0.000477)	(0.0108)	(0.00446)	(0.000416)	(0.00216)
Log of GDPPC					-0.0880***	0.00462***	0.0186***
					(0.00373)	(0.000348)	(0.00181)
Log of Total Population				0.365***	0.0373***	0.0206***	0.138***
				(0.00981)	(0.00427)	(0.000399)	(0.00207)
Constant	10.47***	9.219***	0.171***	5.362***	9.588***	-0.0797***	-0.413***
	(0.374)	(0.166)	(0.0169)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.273	0.665	0.550	0.611	0.684	0.650	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S6-2. Impacts of geographic barriers on major socioeconomic outcome using UCDB boundaries and old version UCDB-2019 attributes.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
-0.607***	2.997***	-0.223***	-1.070***	2.194***	-0.177***	0.0756*	-0.607***
(0.123)	(0.173)	(0.0164)	(0.143)	(0.148)	(0.0158)	(0.0409)	(0.123)
0.225***	-0.435***	0.0107***	0.0419	-0.329***	-0.000620	0.0304***	0.225***
(0.0226)	(0.0317)	(0.00301)	(0.0256)	(0.0265)	(0.00282)	(0.00711)	(0.0226)
2.767***	-0.460***	0.125***	0.162**	0.0371	0.00570	-0.0758***	2.767***
(0.0688)	(0.0968)	(0.00917)	(0.0818)	(0.0846)	(0.00903)	(0.0227)	(0.0688)
0.0230	-0.00311	0.00110	0.0322	-0.0103	-2.41e-05	0.0369***	0.0230
(0.0206)	(0.0289)	(0.00274)	(0.0233)	(0.0241)	(0.00258)	(0.00647)	(0.0206)
0.00542	-0.146***	0.00446*	0.0630***	-0.0862***	0.00217	-0.00629	0.00542
(0.0184)	(0.0259)	(0.00245)	(0.0209)	(0.0216)	(0.00231)	(0.00581)	(0.0184)
0.224***	0.472***	-0.0286***	-0.173**	0.402***	-0.0346***	0.136***	0.224***
(0.0765)	(0.108)	(0.0102)	(0.0869)	(0.0899)	(0.00960)	(0.0241)	(0.0765)
0.00245	0.0367***	-0.00515***	-0.0264***	0.0229**	-0.00455***	0.0338***	0.00245
(0.00768)	(0.0108)	(0.00102)	(0.00869)	(0.00899)	(0.000960)	(0.00241)	(0.00768)
				-0.494***	0.0252***	0.0628***	
				(0.00923)	(0.000985)	(0.00249)	
			0.194***	-0.0419***	0.0363***	0.109***	
			(0.00992)	(0.0104)	(0.00111)	(0.00279)	
10.88***	9.820***	0.366***	6.058***	14.09***	-0.222***	-0.504***	10.88***
(0.272)	(0.383)	(0.0362)	(0.326)	(0.341)	(0.0365)	(0.0916)	(0.272)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,109	13,102	13,109	12,796	12,789	12,796	12,761
R-squared	0.165	0.656	0.723	0.666	0.730	0.758	0.628

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S6-3. Impacts of geographic barriers on major socioeconomic outcomes using GHUB boundaries.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-0.481	0.511***	-0.0529***	-0.684***	0.717***	-0.0480***	0.425***
	(0.337)	(0.128)	(0.0129)	(0.149)	(0.0967)	(0.0112)	(0.0459)
Dummy of coastal city	0.333***	0.0615***	0.00399**	-0.0828***	-0.0230*	-0.00247*	0.0314***
	(0.0442)	(0.0168)	(0.00170)	(0.0196)	(0.0127)	(0.00147)	(0.00603)
Dummy of Capital City	1.145***	0.123***	4.73e-05	0.217***	-0.117***	-0.0209***	-0.0600***
	(0.0938)	(0.0357)	(0.00360)	(0.0418)	(0.0270)	(0.00313)	(0.0128)
City not in Major River Watershed	-0.125***	-0.0775***	-0.00633***	0.112***	-0.0359***	-0.00311**	-0.0175***
	(0.0426)	(0.0162)	(0.00164)	(0.0189)	(0.0122)	(0.00142)	(0.00581)
Log of Average Precipitation	0.206***	0.0726***	0.00847***	0.110***	0.0294**	0.00422**	0.0476***
	(0.0498)	(0.0189)	(0.00191)	(0.0222)	(0.0144)	(0.00166)	(0.00682)
Log of Average Temperature	0.823***	0.382***	-0.0100*	-0.356***	0.160***	-0.0273***	0.00782
	(0.137)	(0.0520)	(0.00525)	(0.0611)	(0.0396)	(0.00458)	(0.0188)
Log Average Elevation	0.0609***	0.0763***	-0.000455	-0.0533***	0.0588***	-0.00175***	0.0249***
	(0.0153)	(0.00583)	(0.000588)	(0.00681)	(0.00441)	(0.000511)	(0.00210)
Log of GDPPC				0.0734***	0.237***	0.0187***	0.114***
				(0.00466)	(0.00305)	(0.000353)	(0.00145)
Log of Total Population					-0.0951***	-0.00365***	0.00654**
					(0.00650)	(0.000753)	(0.00309)
Constant	6.624***	7.354***	0.157***	9.451***	6.747***	0.0777***	0.185**
	(0.536)	(0.204)	(0.0205)	(0.241)	(0.168)	(0.0194)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	10,135	10,135	10,142	10,082	10,082	10,082	10,082
R-squared	0.270	0.601	0.455	0.604	0.765	0.588	0.599

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S6-4. Impacts of geographic barriers on major socioeconomic outcomes using GUB boundaries.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-0.936***	0.354***	-0.0490***	-0.821***	0.602***	-0.0298***	0.373***
	(0.276)	(0.113)	(0.0124)	(0.130)	(0.0820)	(0.0110)	(0.0469)
Dummy of coastal city	0.281***	0.0597***	0.00119	-0.0584***	-0.0193*	-0.00425***	0.0513***
	(0.0382)	(0.0157)	(0.00172)	(0.0179)	(0.0113)	(0.00152)	(0.00648)
Dummy of Capital City	1.101***	0.148***	0.00144	0.200***	-0.106***	-0.0209***	-0.0602***
	(0.0850)	(0.0349)	(0.00382)	(0.0400)	(0.0253)	(0.00341)	(0.0145)
City not in Major River Watershed	-0.0929**	-0.0470***	-0.00842***	0.0957***	-0.0104	-0.00634***	-0.0350***
	(0.0368)	(0.0151)	(0.00165)	(0.0173)	(0.0109)	(0.00147)	(0.00625)
Log of Average Precipitation	0.176***	0.0782***	0.00839***	0.113***	0.0371***	0.00331*	0.0369***
	(0.0452)	(0.0185)	(0.00203)	(0.0213)	(0.0135)	(0.00181)	(0.00772)
Log of Average Temperature	0.513***	0.252***	-0.00559	-0.462***	0.0617*	-0.0152***	-0.0332
	(0.126)	(0.0515)	(0.00564)	(0.0591)	(0.0374)	(0.00503)	(0.0214)
Log Average Elevation	0.0684***	0.0742***	-0.00166***	-0.0472***	0.0531***	-0.00300***	0.0307***
	(0.0133)	(0.00546)	(0.000599)	(0.00625)	(0.00396)	(0.000532)	(0.00227)
Log of GDPPC				0.123***	0.270***	0.0189***	0.129***
				(0.00416)	(0.00271)	(0.000365)	(0.00155)
Log of Total Population					-0.107***	-0.000981	0.0586***
					(0.00539)	(0.000724)	(0.00308)
Constant	7.303***	7.693***	0.210***	9.058***	6.865***	0.0935***	-0.320***
	(0.485)	(0.199)	(0.0218)	(0.230)	(0.153)	(0.0206)	(0.0877)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	14,055	14,055	14,073	13,963	13,963	13,963	13,963
R-squared	0.254	0.561	0.407	0.568	0.759	0.521	0.556

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S6-5. Impacts of geographic barriers on major socioeconomic outcome using the AUE.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-4.634**	2.672**	-0.108	-2.094	2.603***	-0.0211	0.773
	(2.318)	(1.037)	(0.123)	(1.522)	(0.926)	(0.118)	(0.547)
Dummy of coastal city	0.972***	-0.0613	0.0296**	-0.0724	-0.192*	0.0128	0.00427
	(0.265)	(0.118)	(0.0140)	(0.178)	(0.108)	(0.0138)	(0.0637)
Dummy of Capital City	1.460***	0.397***	0.00833	-0.00892	0.229**	-0.0173	0.0615
	(0.245)	(0.109)	(0.0130)	(0.173)	(0.105)	(0.0134)	(0.0620)
City not in Major River Watershed	-0.442*	0.0301	0.0217	0.233	0.146	0.0288**	0.0726
	(0.263)	(0.118)	(0.0140)	(0.172)	(0.105)	(0.0134)	(0.0620)
Log of Average Precipitation	-0.172	-0.0544	-0.0338***	0.0240	-0.0281	-0.0309***	-0.0110
	(0.140)	(0.0627)	(0.00745)	(0.0915)	(0.0554)	(0.00709)	(0.0327)
Log of Average Temperature	0.418	1.604***	0.00245	-2.212***	0.929***	0.000898	0.290***
	(0.397)	(0.178)	(0.0211)	(0.259)	(0.184)	(0.0236)	(0.109)
Log Average Elevation	0.0756	0.0539	0.00618	-0.170***	-0.00283	0.00529	0.0181
	(0.0760)	(0.0340)	(0.00403)	(0.0495)	(0.0309)	(0.00395)	(0.0182)
Log of GDPPC				0.147***	0.155***	0.0172***	0.0947***
				(0.0469)	(0.0291)	(0.00373)	(0.0172)
Log of Total Population					-0.284***	0.00262	-0.0609**
					(0.0438)	(0.00560)	(0.0259)
Constant	13.00***	3.588***	0.831***	15.44***	6.500***	0.563***	-0.667
	(1.507)	(0.674)	(0.0800)	(1.154)	(0.973)	(0.124)	(0.575)
Observations	200	200	200	200	200	200	200
R-squared	0.239	0.361	0.135	0.347	0.510	0.231	0.283

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S6-6. Impacts of geographic barriers on environmental outcomes using the UCDB (all parameters).

VARIABLES	(1) Log Average Nighttime Light Emission in 2015	(2) Average greenness	(3) Log Residential CO2 Emissions non-organic 2015	(4) Log Transportation CO2 Emissions non-organic 2015
Average Local Non-convexity, 10km	-0.772***	0.0174*	-0.896***	-1.598***
	(0.154)	(0.0106)	(0.134)	(0.271)
Dummy of coastal city	0.119***	-0.0161***	0.209***	0.489***
	(0.0268)	(0.00178)	(0.0238)	(0.0473)
Dummy of Capital City	0.00706	0.00539	0.201***	0.364**
	(0.0852)	(0.00561)	(0.0762)	(0.150)
City not in Major River Watershed	0.0953***	-0.0102***	-0.0383*	0.0637
	(0.0244)	(0.00162)	(0.0217)	(0.0433)
Log of Average Precipitation	-0.109***	0.0261***	0.0340*	0.0288
	(0.0219)	(0.00146)	(0.0195)	(0.0387)
Log of Average Temperature	0.296***	-0.0429***	-0.539***	-0.280*
	(0.0907)	(0.00602)	(0.0810)	(0.161)
Log Average Elevation	0.0296***	-0.00613***	-0.121***	-0.0947***
	(0.00908)	(0.000603)	(0.00810)	(0.0161)
Log of GDPPC	0.367***	-0.0177***	1.338***	1.372***
	(0.0104)	(0.000695)	(0.00938)	(0.0186)
Log of Total Population		-0.00300***	0.226***	0.405***
		(0.000695)	(0.00831)	(0.0170)
Constant	-3.237***	0.648***	-6.852***	-9.994***
	(0.339)	(0.0229)	(0.308)	(0.610)
Country Fixed Effects	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES
Observations	12,524	12,374	12,795	12,541
R-squared	0.696	0.690	0.900	0.678

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Section 4: Robustness checks

4.1 Detour regressions at 5 kilometers for all datasets (UCDB, GHUB, GUB and AUE)

Table S7-1. Impacts of geographic barriers on street/road patterns of UCDB (5 km radius).

	Average detour within 5 km of City Center					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Average Local Non-convexity, 5 km		3.250***		2.451***	2.442***	2.458***
		(0.0880)		(0.103)	(0.128)	(0.111)
Share of barriers, 5 km	0.936***		0.729***		0.00705	
	(0.0433)		(0.0516)		(0.0634)	
Dummy of coastal cities			0.0202	0.0214	0.0207	0.0153
			(0.0166)	(0.0152)	(0.0164)	(0.0165)
Dummy of Capital City			0.203***	0.186***	0.186***	0.157***
			(0.0489)	(0.0482)	(0.0482)	(0.0511)
City not in Major River Watershed			-0.0820***	-0.0668***	-0.0669***	-0.0822***
			(0.0144)	(0.0142)	(0.0142)	(0.0144)
Log of Average Precipitations in 2015			0.0236*	0.00278	0.00285	-0.00453
			(0.0128)	(0.0127)	(0.0127)	(0.0120)
Log of Average Temperature in 2015			0.0165	0.0394	0.0397	0.00164
			(0.0534)	(0.0525)	(0.0526)	(0.0515)
Log Average Elevation			0.0130**	0.00370	0.00376	-0.00387
			(0.00532)	(0.00524)	(0.00527)	(0.00528)
Log of GDP per capital			0.00931**	0.0105**	0.0105**	0.0183***
			(0.00458)	(0.00451)	(0.00451)	(0.00469)
Log of Total Resident Population in 2015			0.00982	-0.0161	-0.0161	-0.00419
			(0.0116)	(0.0115)	(0.0115)	(0.0114)
Log of Build-up Area			-0.115***	-0.0882***	-0.0882***	-0.103***
			(0.0108)	(0.0107)	(0.0107)	(0.0105)
COUNTRY: Urban population (% of total population)						-0.00134*
						(0.000753)
COUNTRY: Average freehouse index from 1973-2015						0.0412***
						(0.00500)
COUNTRY: Population ages 0-14 (% of total population)						-0.00532***
						(0.00163)

COUNTRY: Literacy rate, adult total (% of people ages 15 and above)						-0.00149**
						(0.000741)
COUNTRY: General government final consumption expenditure (% of GDP)						-0.00251
						(0.00184)
COUNTRY: GDPPC						0.0816***
						(0.0167)
Constant	0.727***	0.697***	0.395*	0.742***	0.740***	0.336
	(0.00631)	(0.00595)	(0.222)	(0.219)	(0.220)	(0.271)
Country Fixed Effects	NO	NO	YES	YES	YES	NO
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES	NO
Observations	13,062	13,062	12,429	12,429	12,429	11,098
R-squared	0.035	0.095	0.166	0.190	0.190	0.158

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S7-2. Impacts of geographic barriers on street/road patterns using GHUB boundaries.

	Detour Index within 5 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 5 km		1.814***		1.749***	1.776***
		(0.0561)		(0.0539)	(0.0653)
Share of barriers, 5 km	0.369***		0.329***		-0.0168
	(0.0184)		(0.0192)		(0.0225)
Dummy of Coastal City			0.0553***	0.0516***	0.0522***
			(0.00569)	(0.00544)	(0.00549)
Dummy of Capital City			0.0684***	0.0680***	0.0677***
			(0.0121)	(0.0116)	(0.0116)
City not in Major River Watershed			-0.0129**	-0.00240	-0.00189
			(0.00548)	(0.00526)	(0.00530)
Log of Average Precipitations in 2015			0.0159**	0.00710	0.00711
			(0.00639)	(0.00617)	(0.00617)
Log of Average Temperature in 2015			0.0583***	0.0720***	0.0716***
			(0.0177)	(0.0170)	(0.0170)
Log of Average Elevation			0.0195***	0.0188***	0.0186***
			(0.00199)	(0.00192)	(0.00192)
Log of GDP per capital			0.00958***	0.00999***	0.00987***
			(0.00294)	(0.00283)	(0.00283)
Log of total Resident Population in 2015			-0.0175***	-0.0246***	-0.0247***
			(0.00368)	(0.00355)	(0.00356)
Log of Build-up Area in 2015			-0.0301***	-0.0236***	-0.0235***
			(0.00448)	(0.00433)	(0.00433)
Constant	0.705***	0.690***	0.450***	0.514***	0.518***
	(0.00238)	(0.00233)	(0.0808)	(0.0778)	(0.0780)
Country Fixed Effects	NO	NO	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES
Observations	10,165	10,165	10,079	10,079	10,079
R-squared	0.038	0.093	0.275	0.325	0.325

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S7-3. Impacts of geographic barriers on street/road patterns of GUB.

	Detour Index within 5 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 5 km		1.765***		1.629***	1.614***
		(0.0462)		(0.0432)	(0.0547)
Share of barriers, 5 km	0.386***		0.351***		0.00819
	(0.0154)		(0.0154)		(0.0189)
Dummy of Coastal City			0.0470***	0.0465***	0.0463***
			(0.00482)	(0.00463)	(0.00467)
Dummy of Capital City			0.0810***	0.0772***	0.0774***
			(0.0107)	(0.0104)	(0.0104)
City not in Major River Watershed			-0.0189***	-0.00835*	-0.00861*
			(0.00465)	(0.00448)	(0.00452)
Log of Average Precipitations in 2015			0.0195***	0.0127**	0.0127**
			(0.00571)	(0.00554)	(0.00554)
Log of Average Temperature in 2015			0.0388**	0.0433***	0.0435***
			(0.0158)	(0.0154)	(0.0154)
Log of Average Elevation			0.0164***	0.0151***	0.0152***
			(0.00170)	(0.00164)	(0.00165)
Log of GDP per capital			-0.00917***	-0.00785***	-0.00781***
			(0.00232)	(0.00225)	(0.00225)
Log of total Resident Population in 2015			-0.0228***	-0.0281***	-0.0281***
			(0.00287)	(0.00279)	(0.00279)
Log of Build-up Area in 2015			-0.0316***	-0.0253***	-0.0253***
			(0.00361)	(0.00350)	(0.00350)
Constant	0.723***	0.709***	0.738***	0.796***	0.794***
	(0.00208)	(0.00204)	(0.0696)	(0.0674)	(0.0676)
Country Fixed Effects	NO	NO	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	NO	NO	YES	YES	YES
Observations	14,095	14,095	13,962	13,962	13,962
R-squared	0.043	0.094	0.290	0.332	0.332

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S7-4. Impacts of geographic barriers on street/road patterns of AUE.

	Detour Index within 5 km of City Center				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Average Local Non-convexity, 5 km		2.550***		2.353***	2.044***
		(0.396)		(0.421)	(0.646)
Share of barriers, 5 km	0.755***		0.695***		0.145
	(0.138)		(0.154)		(0.229)
Dummy of Coastal City			0.0525	0.0574*	0.0538*
			(0.0318)	(0.0305)	(0.0311)
Dummy of Capital City			0.0254	0.0211	0.0214
			(0.0309)	(0.0302)	(0.0302)
City not in Major River Watershed			-0.0763**	-0.0495	-0.0541*
			(0.0309)	(0.0300)	(0.0309)
Log of Average Precipitations in 2015			-0.0234	-0.0286*	-0.0285*
			(0.0161)	(0.0158)	(0.0158)
Log of Average Temperature in 2015			0.0778	0.0939*	0.0929*
			(0.0565)	(0.0553)	(0.0554)
Log of Average Elevation			0.00337	0.00594	0.00560
			(0.00899)	(0.00878)	(0.00881)
Log of GDP per capital			0.00812	0.00710	0.00738
			(0.0141)	(0.0137)	(0.0138)
Log of total Resident Population in 2015			0.0149	0.00819	0.00895
			(0.0196)	(0.0192)	(0.0192)
Log of Build-up Area in 2015			-0.0339	-0.0288	-0.0285
			(0.0209)	(0.0204)	(0.0205)
Constant		2.550***		2.353***	2.044***
		(0.396)		(0.421)	(0.646)
Observations	200	200	200	200	200
R-squared	0.132	0.173	0.196	0.235	0.237

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.2 Nonconvexity regressions at 5 kilometers from all datasets (UCDB, GHUB, GUB and AUE)

Table S8-1. Impacts of geographic barriers on major socioeconomic outcomes using UCDB boundaries: 5 km radius.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 5km	-2.102***	1.723***	-0.150***	-0.957***	1.598***	-0.115***	0.165***
	(0.189)	(0.0840)	(0.00802)	(0.206)	(0.0851)	(0.00793)	(0.0413)
Dummy of coastal city	0.392***	-0.0827***	0.00185	-0.0359	-0.0737***	-0.00666***	0.0102
	(0.0304)	(0.0135)	(0.00137)	(0.0312)	(0.0129)	(0.00120)	(0.00625)
Dummy of Capital City	2.980***	0.0206	0.0457***	-0.236**	-0.00763	-0.0205***	-0.160***
	(0.0942)	(0.0419)	(0.00426)	(0.0992)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00472	-0.0242*	0.00117	0.0702**	-0.0282**	0.00130	0.0385***
	(0.0282)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0437*	-0.0407***	-0.000624	0.0924***	-0.0341***	-0.00182*	-0.00446
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.00521)
Log of Average Temperature	0.179*	0.398***	-0.00389	-0.511***	0.362***	-0.00721*	0.120***
	(0.105)	(0.0467)	(0.00473)	(0.108)	(0.0445)	(0.00415)	(0.0216)
Log Average Elevation	0.00523	0.0354***	0.000960**	-0.0311***	0.0304***	0.00101**	0.0328***
	(0.0105)	(0.00467)	(0.000474)	(0.0108)	(0.00445)	(0.000414)	(0.00216)
Log of GDPPC					-0.0880***	0.00462***	0.0185***
					(0.00373)	(0.000348)	(0.00181)
Log of Total Population				0.364***	0.0378***	0.0205***	0.138***
				(0.00981)	(0.00427)	(0.000398)	(0.00207)
Constant	10.41***	9.287***	0.166***	5.324***	9.642***	-0.0825***	-0.401***
	(0.373)	(0.166)	(0.0168)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.275	0.665	0.552	0.611	0.684	0.651	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S8-2. Impacts of geographic barriers on major socioeconomic outcomes using UCDB boundaries and old version UCDB-2019 attributes : 5 km radius.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 5km	-0.509***	3.226***	-0.220***	-1.172***	2.358***	-0.180***	-0.00411
	(0.130)	(0.182)	(0.0173)	(0.153)	(0.159)	(0.0169)	(0.0448)
Dummy of coastal city	0.211***	-0.398***	0.00707**	0.0294	-0.301***	-0.00332	0.0352***
	(0.0221)	(0.0310)	(0.00294)	(0.0250)	(0.0259)	(0.00276)	(0.00695)
Dummy of Capital City	2.766***	-0.452***	0.125***	0.158*	0.0461	0.00498	-0.0755***
	(0.0688)	(0.0967)	(0.00917)	(0.0818)	(0.0846)	(0.00904)	(0.0227)
City not in Major River Watershed	0.0214	0.00544	0.000486	0.0290	-0.00373	-0.000545	0.0370***
	(0.0206)	(0.0289)	(0.00274)	(0.0233)	(0.0241)	(0.00258)	(0.00647)
Log of Average Precipitation	0.00409	-0.145***	0.00423*	0.0627***	-0.0851***	0.00201	-0.00564
	(0.0184)	(0.0259)	(0.00245)	(0.0209)	(0.0216)	(0.00231)	(0.00581)
Log of Average Temperature	0.240***	0.423***	-0.0242**	-0.155*	0.363***	-0.0312***	0.132***
	(0.0764)	(0.107)	(0.0102)	(0.0867)	(0.0897)	(0.00959)	(0.0241)
Log Average Elevation	0.000346	0.0442***	-0.00579***	-0.0291***	0.0286***	-0.00505***	0.0342***
	(0.00766)	(0.0108)	(0.00102)	(0.00865)	(0.00896)	(0.000957)	(0.00240)
Log of GDPPC					-0.494***	0.0252***	0.0626***
					(0.00923)	(0.000986)	(0.00250)
Log of Total Population				0.195***	-0.0432***	0.0364***	0.109***
				(0.00991)	(0.0104)	(0.00111)	(0.00279)
Constant	10.84***	9.952***	0.355***	6.001***	14.20***	-0.232***	-0.491***
	(0.272)	(0.382)	(0.0362)	(0.325)	(0.341)	(0.0364)	(0.0915)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,109	13,102	13,109	12,796	12,789	12,796	12,761
R-squared	0.165	0.657	0.722	0.667	0.730	0.758	0.628

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S8-3. Impacts of geographic barriers on major socioeconomic outcome using GHUB boundaries: 5km radius.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 5km	0.0630	0.700***	-0.0886***	-1.078***	0.791***	-0.0929***	0.485***
	(0.432)	(0.164)	(0.0165)	(0.191)	(0.124)	(0.0143)	(0.0589)
Dummy of coastal city	0.322***	0.0632***	0.00401**	-0.0836***	-0.0179	-0.00233	0.0341***
	(0.0440)	(0.0167)	(0.00169)	(0.0195)	(0.0126)	(0.00146)	(0.00601)
Dummy of Capital City	1.148***	0.124***	-8.59e-05	0.215***	-0.117***	-0.0211***	-0.0598***
	(0.0938)	(0.0357)	(0.00360)	(0.0417)	(0.0270)	(0.00312)	(0.0128)
City not in Major River Watershed	-0.127***	-0.0748***	-0.00664***	0.108***	-0.0321***	-0.00338**	-0.0152***
	(0.0426)	(0.0162)	(0.00163)	(0.0189)	(0.0122)	(0.00141)	(0.00581)
Log of Average Precipitation	0.200***	0.0738***	0.00847***	0.109***	0.0323**	0.00432***	0.0493***
	(0.0497)	(0.0189)	(0.00191)	(0.0222)	(0.0144)	(0.00166)	(0.00682)
Log of Average Temperature	0.834***	0.382***	-0.0103**	-0.360***	0.158***	-0.0279***	0.00709
	(0.137)	(0.0520)	(0.00524)	(0.0611)	(0.0396)	(0.00457)	(0.0188)
Log Average Elevation	0.0603***	0.0776***	-0.000600	-0.0552***	0.0605***	-0.00191***	0.0259***
	(0.0153)	(0.00583)	(0.000588)	(0.00681)	(0.00442)	(0.000510)	(0.00210)
Log of GDPPC				0.0738***	0.237***	0.0188***	0.114***
				(0.00466)	(0.00305)	(0.000353)	(0.00145)
Log of Total Population					-0.0949***	-0.00378***	0.00667**
					(0.00651)	(0.000752)	(0.00309)
Constant	6.623***	7.337***	0.159***	9.475***	6.729***	0.0810***	0.174**
	(0.536)	(0.204)	(0.0205)	(0.241)	(0.168)	(0.0194)	(0.0797)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	10,135	10,135	10,142	10,082	10,082	10,082	10,082
R-squared	0.269	0.601	0.456	0.605	0.765	0.589	0.599

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S8-4. Impacts of geographic barriers on major socioeconomic outcome using GUB boundaries: 5 km radius.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 5km	-1.580***	0.531***	-0.0666***	-1.434***	0.885***	-0.0323**	0.502***
	(0.352)	(0.144)	(0.0157)	(0.165)	(0.105)	(0.0141)	(0.0599)
Dummy of coastal city	0.281***	0.0606***	0.000989	-0.0578***	-0.0177	-0.00447***	0.0529***
	(0.0380)	(0.0156)	(0.00171)	(0.0178)	(0.0113)	(0.00152)	(0.00645)
Dummy of Capital City	1.098***	0.149***	0.00138	0.198***	-0.105***	-0.0209***	-0.0601***
	(0.0850)	(0.0349)	(0.00382)	(0.0400)	(0.0253)	(0.00341)	(0.0145)
City not in Major River Watershed	-0.0973***	-0.0453***	-0.00865***	0.0918***	-0.00761	-0.00648***	-0.0332***
	(0.0368)	(0.0151)	(0.00165)	(0.0172)	(0.0109)	(0.00147)	(0.00624)
Log of Average Precipitation	0.177***	0.0783***	0.00830***	0.114***	0.0374***	0.00319*	0.0374***
	(0.0452)	(0.0185)	(0.00203)	(0.0213)	(0.0135)	(0.00181)	(0.00771)
Log of Average Temperature	0.511***	0.252***	-0.00542	-0.463***	0.0611	-0.0150***	-0.0343
	(0.125)	(0.0514)	(0.00564)	(0.0590)	(0.0374)	(0.00503)	(0.0214)
Log Average Elevation	0.0656***	0.0751***	-0.00178***	-0.0498***	0.0548***	-0.00307***	0.0317***
	(0.0133)	(0.00546)	(0.000599)	(0.00625)	(0.00396)	(0.000533)	(0.00227)
Log of GDPPC				0.122***	0.270***	0.0189***	0.129***
				(0.00416)	(0.00271)	(0.000365)	(0.00155)
Log of Total Population					-0.106***	-0.000998	0.0592***
					(0.00540)	(0.000725)	(0.00309)
Constant	7.322***	7.687***	0.211***	9.079***	6.842***	0.0940***	-0.332***
	(0.485)	(0.199)	(0.0218)	(0.230)	(0.153)	(0.0206)	(0.0878)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	14,055	14,055	14,073	13,963	13,963	13,963	13,963
R-squared	0.255	0.561	0.407	0.569	0.760	0.521	0.556

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S8-5. Impacts of geographic barriers on major socioeconomic outcomes suing the AUE: 5 km radius.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 5km	-8.468**	3.038*	-0.181	-2.908	3.150**	-0.0251	-0.217
	(3.650)	(1.652)	(0.194)	(2.417)	(1.480)	(0.188)	(0.871)
Dummy of coastal city	0.954***	-0.0312	0.0290**	-0.0902	-0.164	0.0125	0.0277
	(0.261)	(0.118)	(0.0139)	(0.176)	(0.108)	(0.0136)	(0.0634)
Dummy of Capital City	1.483***	0.383***	0.00887	0.00189	0.218**	-0.0172	0.0633
	(0.244)	(0.110)	(0.0130)	(0.174)	(0.106)	(0.0134)	(0.0624)
City not in Major River Watershed	-0.486*	0.0495	0.0207	0.216	0.166	0.0287**	0.0736
	(0.263)	(0.119)	(0.0140)	(0.173)	(0.106)	(0.0134)	(0.0624)
Log of Average Precipitation	-0.162	-0.0493	-0.0337***	0.0231	-0.0245	-0.0309***	-0.00388
	(0.140)	(0.0634)	(0.00746)	(0.0918)	(0.0560)	(0.00710)	(0.0330)
Log of Average Temperature	0.443	1.586***	0.00306	-2.198***	0.903***	0.00111	0.275**
	(0.395)	(0.179)	(0.0210)	(0.259)	(0.185)	(0.0235)	(0.109)
Log Average Elevation	0.0621	0.0601*	0.00588	-0.175***	0.00279	0.00525	0.0185
	(0.0757)	(0.0343)	(0.00403)	(0.0495)	(0.0312)	(0.00395)	(0.0184)
Log of GDPPC				0.147***	0.154***	0.0172***	0.0913***
				(0.0471)	(0.0295)	(0.00374)	(0.0174)
Log of Total Population					-0.288***	0.00265	-0.0651**
					(0.0441)	(0.00560)	(0.0260)
Constant	12.92***	3.601***	0.830***	15.43***	6.595***	0.562***	-0.567
	(1.503)	(0.680)	(0.0801)	(1.155)	(0.980)	(0.124)	(0.577)
Observations	200	200	200	200	200	200	200
R-squared	0.245	0.351	0.136	0.346	0.502	0.231	0.276

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.3 Checks for potential outliers

Here, we would like to establish the robustness of the results to potential outliers. The concern here is some observations with very high and low values of the density variable. We present the results using a trimming technique to the GHSL23-UCDB (our data processing of UCDB's boundaries), commonly used to address concerns about extreme observations. Concretely, conservatively, we remove observations that are at the top and bottom decile of the density distribution. The results based on this "clean" sample— now available in Table S9 —are almost identical to the ones using the full sample.

Table S9. Robustness check of potential outliers in UCDB.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Log of Population	Log of density	Share of build-up	Log of GDP PC	Log of density	Share of build-up
Average Local Non-convexity, 10km	-1.668***	0.805***	-0.111***	-0.730***	0.789***	-0.0726***
	(0.212)	(0.0648)	(0.00934)	(0.206)	(0.0655)	(0.00802)
Dummy of coastal city	0.426***	-0.0344***	0.00323**	-0.0660**	-0.0359***	-0.00670***
	(0.0350)	(0.0107)	(0.00154)	(0.0334)	(0.0106)	(0.00130)
Dummy of Capital City	3.118***	0.0807***	0.0473***	-0.389***	0.0655**	-0.0280***
	(0.101)	(0.0308)	(0.00444)	(0.1000)	(0.0319)	(0.00390)
City not in Major River Watershed	0.0564*	-0.0249**	0.00222	0.0575*	-0.0205**	0.000509
	(0.0319)	(0.00975)	(0.00140)	(0.0304)	(0.00967)	(0.00118)
Log of Average Precipitation	0.0691**	-0.0302***	0.00119	0.0565**	-0.0286***	-0.000223
	(0.0280)	(0.00856)	(0.00123)	(0.0267)	(0.00851)	(0.00104)
Log of Average Temperature	0.157	0.393***	-0.00921*	-0.466***	0.361***	-0.0130***
	(0.123)	(0.0374)	(0.00539)	(0.116)	(0.0370)	(0.00454)
Log Average Elevation	0.00932	0.0273***	0.00161***	-0.0310***	0.0251***	0.00150***
	(0.0119)	(0.00363)	(0.000523)	(0.0113)	(0.00360)	(0.000441)
Log of GDPPC					-0.0601***	0.00348***
					(0.00320)	(0.000392)
Log of Total Population				0.393***	0.0217***	0.0230***
				(0.0101)	(0.00346)	(0.000424)
Constant	10.30***	9.251***	0.169***	5.085***	9.594***	-0.0969***
	(0.425)	(0.130)	(0.0187)	(0.415)	(0.133)	(0.0163)
Country Fixed Effects	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	10,440	10,440	10,440	10,102	10,102	10,102
R-squared	0.279	0.620	0.500	0.571	0.632	0.641

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.4 Robustness to National Boundaries

National boundaries are not entitative ground objects like water and mountains, however, they may play a similar role as water barriers and steep slopes for urban development. For some bounder cities, like Maseru (Lesotho) and Dandong (China), national boundaries could be a substantial constraint to urban expansion (figure 20). That is also the reason we take national boundaries into consideration in our study. Nevertheless, the results are robust to excluding borders from our indexes. The results in Table S10 indicate that the impacts of geographic barriers on socioeconomic factors are consistent when national boundaries are excluded from the measurements. Our open-source data is also available in both versions: either including national boundaries or not.

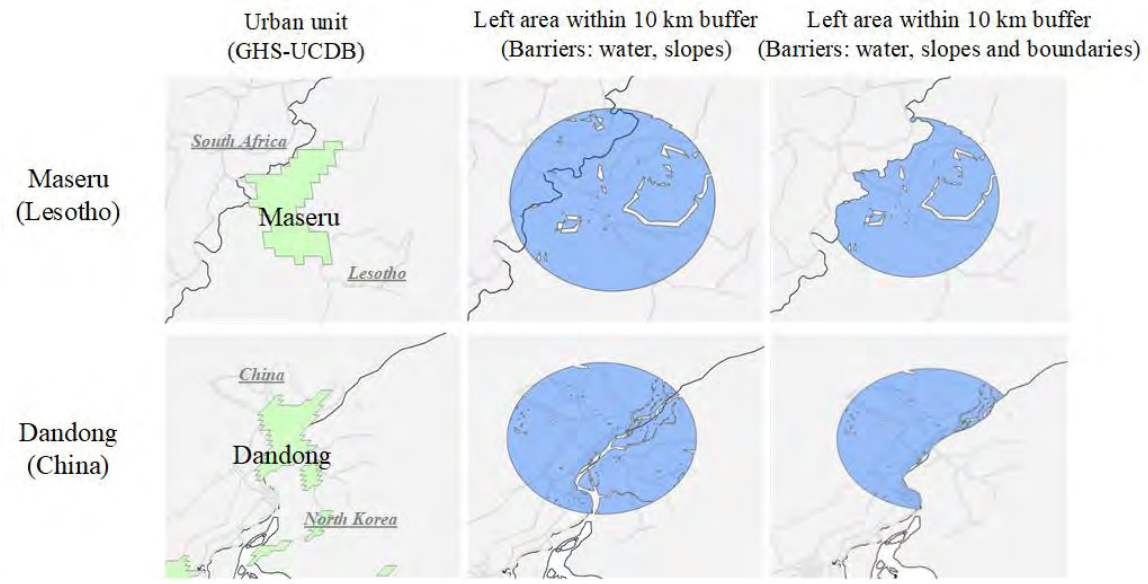

Figure 20. The effects on our measurements in two border cities when the national boundaries are considered (middle figures) and not considered (right figures) as geographic barriers. The blue area represents the 10 km radius with and without border considerations.

Table S10. *Ignoring the existence of borders*: Socioeconomic outcome of UCDB.

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km (borders ignored)	-1.696***	1.587***	-0.133***	-0.902***	1.447***	-0.0974***	-1.696***
	(0.175)	(0.0777)	(0.00765)	(0.187)	(0.0775)	(0.00723)	(0.175)
Dummy of coastal city	0.401***	-0.102***	0.00322**	-0.0238	-0.0903***	-0.00587***	0.401***
	(0.0311)	(0.0138)	(0.00140)	(0.0319)	(0.0132)	(0.00123)	(0.0311)
Dummy of Capital City	2.986***	0.0167	0.0461***	-0.235**	-0.00994	-0.0204***	2.986***
	(0.0943)	(0.0419)	(0.00427)	(0.0992)	(0.0410)	(0.00383)	(0.0943)
City not in Major River Watershed	0.0102	-0.0289**	0.00153	0.0731**	-0.0330***	0.00163	0.0102
	(0.0283)	(0.0126)	(0.00128)	(0.0291)	(0.0120)	(0.00112)	(0.0283)
Log of Average Precipitation	0.0420*	-0.0410***	-0.000634	0.0928***	-0.0342***	-0.00188*	0.0420*
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.0253)
Log of Average Temperature	0.163	0.421***	-0.00560	-0.525***	0.382***	-0.00835**	0.163
	(0.105)	(0.0467)	(0.00475)	(0.108)	(0.0446)	(0.00417)	(0.105)
Log Average Elevation	0.00884	0.0311***	0.00126***	-0.0285***	0.0264***	0.00126***	0.00884
	(0.0105)	(0.00469)	(0.000477)	(0.0108)	(0.00446)	(0.000416)	(0.0105)
Log of GDPPC					-0.0880***	0.00462***	
					(0.00373)	(0.000348)	
Log of Total Population				0.365***	0.0373***	0.0206***	
				(0.00981)	(0.00427)	(0.000399)	
Constant	10.46***	9.225***	0.171***	5.358***	9.594***	-0.0801***	10.46***
	(0.374)	(0.166)	(0.0169)	(0.394)	(0.164)	(0.0153)	(0.374)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	13,048
R-squared	0.273	0.665	0.550	0.611	0.684	0.650	0.273

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.5 Fixed Radiuses: Avoiding Measurement Endogeneity

4.5.1 Theorical discussion

Studying the current shapes of the existing urban footprint is important for urban planning, environmental outcomes, and the study of city expansion. However, the current size, density, and shape of cities are outcomes, or endogenous variables. In our paper, we focus on exogenous or primigenial factors as the independent variables, which impel us to use a constant radius—like 10 km or 5 km, as the 'environment' around cities rather than an endogenous buffer according to the shape or extent of each city.

In this paper we are not focused on marginal urban expansion: this is, the further expansion of large cities with respect to their current boundaries. Rather, we are interested in knowing why/where large human settlements exist at all. A way to frame the question is by thinking about the full ex ante distribution of potentially urbanizable environments, prior to the explosion of human urbanization in the past century. Naturally, we want to condition on places that had the minimum demand conditions to qualify as urban. Now the question becomes: which of these ex-ante environments have tended to engender larger, denser, richer, taller, and more "saturated" cities today? We want to see if potential constraints around a potential expansion of a city mattered. To answer this question, we need an exogenous and comparable definition of the environment around each potential city. Of course, some cities will end up smaller because of their surrounding geographic barriers, but that is precisely an outcome of such constraints. In other words, the actual urban footprint is an outcome, whereas a consistently measured geographic environment around a place that could potentially become a large city is, by definition, exogenous.

In the subsections below, we provide more examples and arguments on why a reasonable scientific approach to the questions that we ask requires comparable fixed radiuses across environments. We also show that the results are robust to any reasonable delimitation of radius, but that some of them may be biased if one adopted an endogenous radius.

Any measure of current city shape is endogenous to the existence of natural barriers around a potential city. To see this point, consider Figure S20-1 (a), depicting a hypothetical city. The urban area, as defined by a suitable method, is represented in gray color. Mountains are in brown, and the ocean is in blue. Flat unurbanized land conforms the white background. A 10 (or 5) kilometer radius is shown in dark green. Now, this city confronted a difficult geography. Per our findings, this implies a smaller denser footprint. Yet, if we use the shapefile that characterizes its current extent (gray area), we will measure a very convex set. But this is so precisely because the city was corseted in by undevelopable land and within a very nonconvex set of developable land. Considering a buffer around the city (in dark orange, Figure S20-1 (b)) does not help, as it follows the shape of city: the buffer is a convex set. Because this city had to avoid difficult terrain, the buffer does not overlap much with the geographical barriers. In contrast, the homogenous exogenous circle, in green, properly identifies this as a geographically challenging environment for a city to grow large (high share undevelopable land and presence of nonconvexities).

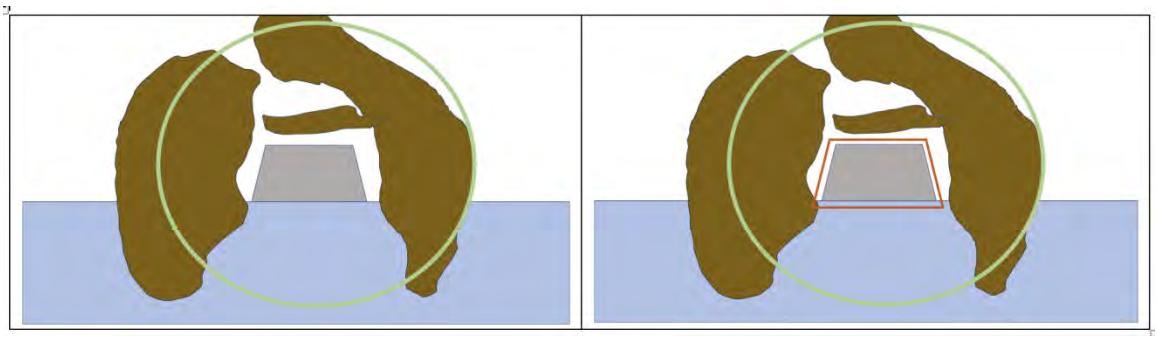

Figure S21-1. A hypothetical city surrounded by geographic barriers. Urban area is represented in gray color. Mountains are in brown, and the ocean is in blue. A 10 (or 5) kilometer radius is shown in dark green and dark orange in (b) represents a buffer around the urban area.

Note that using a buffer around the existing perimeter of the city would be indeed germane if we wanted to focus on further city expansion around the buffer, conditional on its current extent (say, how much room there is for population to grow by 10% in a contiguous fashion?). But that is not our intent, as we are providing evidence about why cities are at their current population/density/income/etc. levels to start with.

Problems may become more acute as one compares alternate sets of geographic environments. And the endogeneity problem affects not only urban shapes but also city size, population, density, and any measures of road network efficiency. Now consider the figure below, where we show two potential locations for a city to arise. The green circles are of the exact same radius. Clearly, environment B has larger share of barriers, and its developable footprint is very nonconvex, as would be captured by our **exogenous** measures.

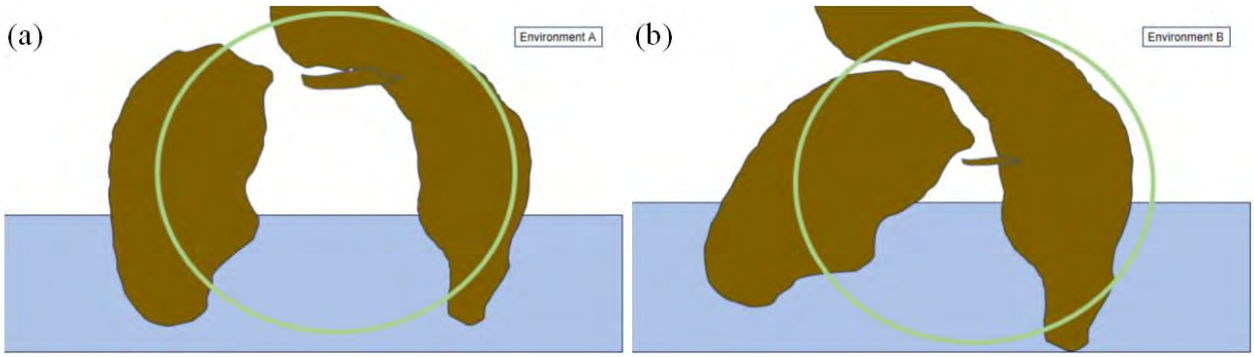

Figure S21-2. Two potential locations for urban growth. Environment A contains less geographic barriers at relevant radius (a), and Environment B has larger share of barriers and higher nonconvexity.

Given the results from our research, one would expect a larger and less dense city to develop in environment A, as illustrated below:

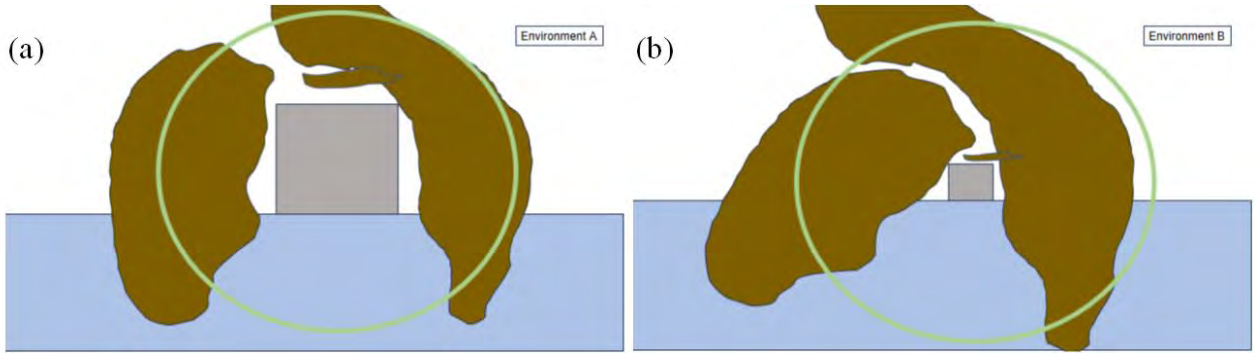

Figure S21-3. A larger and less dense city develops in environment A (a) and a smaller and dense city to develop in environment B (b).

While these scenarios are speculative, it is very clear that geography will impact the shape and extent of the urban perimeter, as captured by the corresponding GIS layers. Now, if one were to calculate non-convexity using endogenous city boundaries (gray areas) one would conclude that both cities are fully convex sets, and that there is no correlation between non-convexity (or other topography measures within the current footprint) and city size/density/etc: we would be **unfortunately drawing the wrong conclusion**.

In contrast, homogenously defined exogenous circles clearly and consistently capture relevant features of the *ex-ante* environment that equidistantly surround a potential city, regardless of its *ex-post* extent. As we will show below, the precise extent of the circle will end up not mattering much, as there is a very strong correlation in measurements across radiuses. Buffers may not help much mitigating this problem, as they follow the contours of the convex city, per the illustration below:

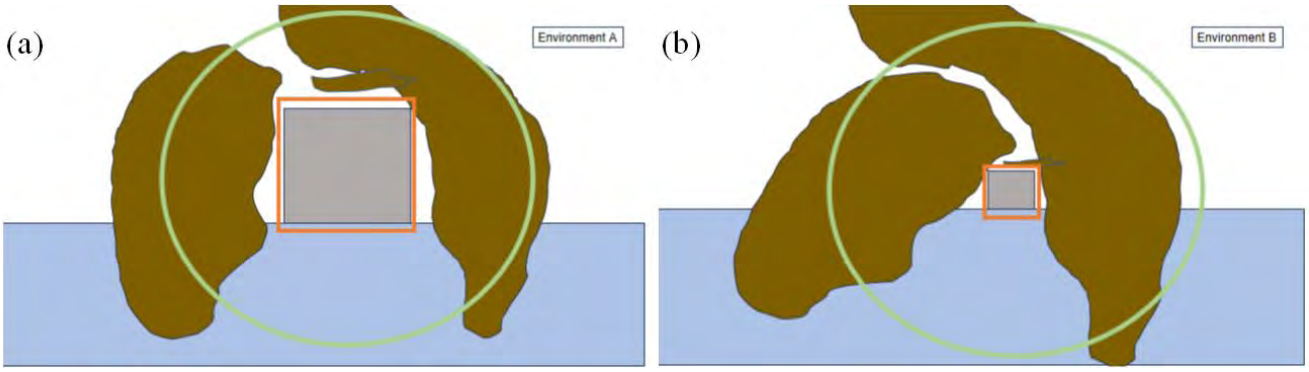

Figure S21-4. Buffers around the convex cities in Environment A (a) and Environment B (b).

An exogenous circle is also useful to study the impacts of ex ante surrounding conditions on urban transportation. A very nonconvex geography may have induced a localized transportation network—including national and regional roads going through the circle—to be more circuitous, which is what we find exactly. Because cities partially grow along roads, distances between potential

areas of expansion get larger, which may itself disincentivize urban expansion. Again, the current extent of the urban footprint is endogenous itself to the local transportation costs. This can be illustrated again using the example in Figure S20-5, where we added a stylized road systems for both environments A and B, signified by the red polylines. In this example, roads and streets within the urban footprint follow a grid. Therefore, if one simply focused on the network within the endogenous city footprint (red lines within gray area), one would conclude that connectivity between locations across environments is similar. While this is true conditioning on the current extent of the cities, a potential city in environment B had a much harder time growing, because marginal developable land at the same beeline distance is scarcer.

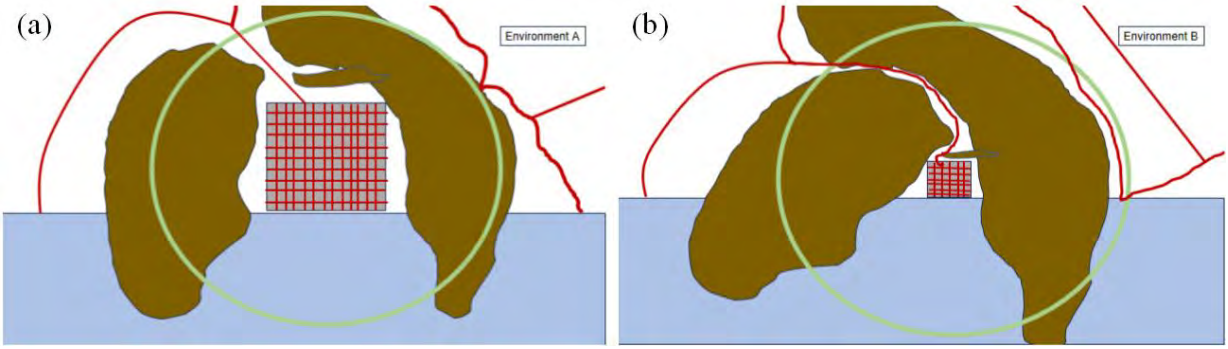

Figure S21-5. Distribution of road networks in Environment A (a) and Environment B (b).

4.5.2 Generating Endogenous measurements:

The arguments above can also be made statistically. Consider again the model below:

$$Y_i = \beta X_i + \varepsilon_i \qquad (5)$$

where X stands for a measure of the geographic environment, say dyadic nonconvexity. Y_i is the population in the city (or density, or any other variable closely related to population). Now, if we let the measurement of X vary with the population of the city—e.g. larger buffers for larger cities—we have $X_i = f(POP_i, Z_i)$, where Z_i are other determinants of nonconvexity. This introduces endogeneity bias into the regression.

$$POP_i = \beta f(POP_i, Z), + \varepsilon_i \qquad (6)$$

To show the statistical problems of having a measurement that is endogenous to the local population, we have designed an 'endogenous radius approach' for each city for the calculation of the share of barriers and the nonconvexity index. This endogenous radius considers the scale (population) of the cities. The endogenous radius is calculated based on a stylized ‘theoretical area of the urban extent'—essentially, the hypothetical area required to accommodate the city's population at the country's average population density (excluding the city itself). We then set the endogenous radius as **twice** the theoretical area’s radius. This serves to represent the city's surrounding environment and potential area for future expansion. The formula for calculating this endogenous radius is described as follows:

$$R_{ik}^{*} = 2 \times \sqrt{\frac{POP_i \times \left[\frac{A_{k \neq i}}{POP_{k \neq i}}\right]}{\pi}} \qquad (7)$$

where R_{ik}^{*} is the theoretical radios of city catchment area, POP_i is the population of the city i, $A_{k \neq i}$ is the total sum of urban areas in other cities in country k (excluding i), $POP_{k \neq i}$ is the total sum of urban population in other cities in country k (excluding i).

Within these endogenous radiuses around each city's centroid in the UCDB, we recalculated the indexes. Now, we focus on this endogenous nonconvexity index (where we adjust the area of the calculation with the population of each city). We start by calculating the difference between this "endogenous" convexity index and the homogeneous nonconvexity measure at fixed 5-kilometer radiuses. The average difference is close to zero (-.0040). But unfortunately, as can be seen in the panel below, it is very clear that the difference between the endogenous and homogenous measures is directly and positively related to the radius we use to measure nonconvexity. The same results are obtained via regression if one tries to explain the variation of endogenous nonconvexity with nonconvexity at the 5 or 10 km radius. These two homogenous measures capture almost 80 of the variance of the former, but large radiuses tend to imply more nonconvexities. By construction, larger radiuses correspond to larger cities, inducing clear endogeneity bias.

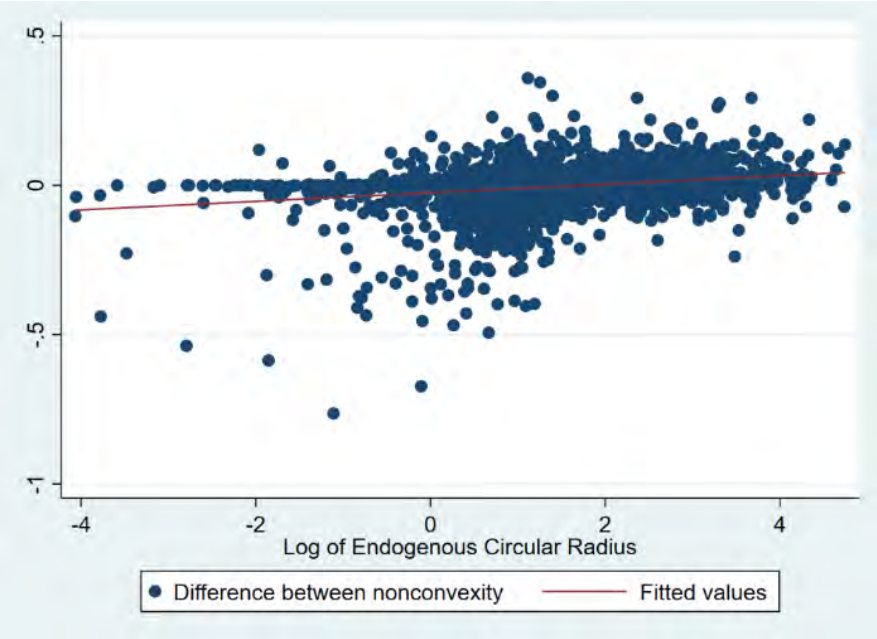

Figure S21-6. Relationship between nonconvexity at 5km and measurement using an endogenous radius. Vertical axis is the difference between the nonconvexity of endogenous radius and nonconvexity of 5 km radius.

Table S11-1. Correlation between local non-convexity and endogenous ratio.

VARIABLES	Average Local Non-convexity with endogenous ratio
Log of Endogenous Circular Radius	0.0140***
	(0.000454)
Average Local Non-convexity within 5km of City Center	0.600***
	(0.00849)
Average Local Non-convexity within 10km of City Center	0.187***
	(0.00725)
Constant	-0.0207***
	(0.000828)
Observations	9,938
Adjusted R-squared	0.795

Why is that? On inspection, we see that this a natural consequence of endogenously expanding the radius of measurement. Urban areas are invariable built in flat and dry land. As we expand the radius of measurement, we tend to mechanically encounter more barriers.

To see that this is, indeed, a mechanical effect, we calculated the distribution of the endogenous radius, as defined by each city's population in the equation above (R_{ik}^{*}). In our UCDB sample, the average radius using this formula—which depends on the city's population and the average residential density in the country—is 5.65, with a standard deviation of 5.87. We then use this mean and sample variance to randomly draw radiuses for each city from a lognormal distribution (remember that radiuses cannot be negative). We then use our geography maps to estimate the nonconvexity around each city's random radius. By construction, these random radiuses are now independent of the city's population or other characteristics. We repeat the process 5 times and each column in Table 13-2 shows the nonconvexity measure as a function of the random radius, controlling for nonconvexity measured at 10 and 5 kilometers. It is very clear that these new measurements are still a function of the random radiuses: the further away we extend the measurement the larger the measured nonconvexity mechanically becomes.

Nonetheless, the fact that we require homogeneous radiuses across cities does not mean that we cannot check on the robustness of the results to our definitions of radiuses: as exposited in next tables results are unchanged using 5 km radiuses.

Table S11-2. Correlation between each random radius and corresponding nonconvexity.

VARIABLES	(1) Nonconvexity of random radius 0	(2) Nonconvexity of random radius 1	(3) Nonconvexity of random radius 2	(4) Nonconvexity of random radius 3	(5) Nonconvexity of random radius 4
Log of random radius 0	0.0146***				
	(0.000405)				
Log of random radius 1		0.0136***			
		(0.000410)			
Log of random radius 2			0.0137***		
			(0.000398)		
Log of random radius 3				0.0144***	
				(0.000398)	
Log of random radius 4					0.0143***
					(0.000397)
Average Local Non-convexity, 5km	0.645***	0.661***	0.709***	0.719***	0.709***
	(0.0109)	(0.0111)	(0.0108)	(0.0108)	(0.0107)
Average Local Non-convexity, 10km	0.0710***	0.0799***	0.0372***	0.0517***	0.0452***
	(0.00929)	(0.00951)	(0.00922)	(0.00920)	(0.00912)
Constant	-0.0171***	-0.0165***	-0.0163***	-0.0178***	-0.0171***
	(0.000662)	(0.000670)	(0.000653)	(0.000648)	(0.000645)
Observations	9,938	9,938	9,938	9,938	9,937
Adjusted R-squared	0.668	0.669	0.682	0.699	0.694

4.5.3 Results using measurements arising from the five random radiuses (exogenous to ex post urban extend)

In the Tables below, we use the measurements from the five random radiuses we generated to estimate specifications as in Table 2. We find similar results in all five regressions.

Table S11-3. Impacts of geographic barriers on major socioeconomic outcome with random radius 0

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
(Random Radius 0) nonconvexity	-2.234***	1.830***	-0.167***	-0.749***	1.699***	-0.131***	0.108**
	(0.203)	(0.0904)	(0.00860)	(0.220)	(0.0908)	(0.00845)	(0.0441)
Dummy of coastal city	0.377***	-0.0703***	0.00103	-0.0520*	-0.0629***	-0.00711***	0.0136**
	(0.0300)	(0.0134)	(0.00135)	(0.0308)	(0.0127)	(0.00119)	(0.00618)
Dummy of Capital City	2.985***	0.0163	0.0460***	-0.236**	-0.0117	-0.0202***	-0.160***
	(0.0942)	(0.0419)	(0.00426)	(0.0993)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00398	-0.0236*	0.00112	0.0704**	-0.0275**	0.00123	0.0385***
	(0.0282)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0408	-0.0384***	-0.000829	0.0895***	-0.0317***	-0.00195*	-0.00382
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.001000)	(0.00521)
Log of Average Temperature	0.193*	0.386***	-0.00288	-0.499***	0.353***	-0.00680	0.117***
	(0.105)	(0.0466)	(0.00472)	(0.108)	(0.0445)	(0.00415)	(0.0216)
Log Average Elevation	0.00435	0.0361***	0.000929**	-0.0320***	0.0308***	0.000998**	0.0330***
	(0.0105)	(0.00467)	(0.000473)	(0.0108)	(0.00444)	(0.000414)	(0.00216)
Log of GDPPC					-0.0888***	0.00466***	0.0183***
					(0.00373)	(0.000347)	(0.00181)
Log of Total Population				0.365***	0.0381***	0.0205***	0.138***
				(0.00981)	(0.00427)	(0.000398)	(0.00208)
Constant	10.39***	9.307***	0.164***	5.292***	9.656***	-0.0826***	-0.393***
	(0.373)	(0.166)	(0.0168)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.275	0.665	0.552	0.611	0.684	0.652	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S11-4. Impacts of geographic barriers on major socioeconomic outcome with random radius 1

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
(Random Radius 1) nonconvexity	-2.212***	1.669***	-0.161***	-0.560***	1.567***	-0.123***	0.146***
	(0.197)	(0.0877)	(0.00843)	(0.213)	(0.0880)	(0.00819)	(0.0427)
Dummy of coastal city	0.375***	-0.0639***	0.000737	-0.0589*	-0.0576***	-0.00746***	0.0124**
	(0.0300)	(0.0134)	(0.00135)	(0.0308)	(0.0127)	(0.00118)	(0.00616)
Dummy of Capital City	2.982***	0.0182	0.0458***	-0.238**	-0.00873	-0.0204***	-0.160***
	(0.0942)	(0.0420)	(0.00426)	(0.0993)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00254	-0.0228*	0.00103	0.0706**	-0.0270**	0.00118	0.0386***
	(0.0282)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0413	-0.0379***	-0.000729	0.0886***	-0.0314***	-0.00196*	-0.00409
	(0.0253)	(0.0113)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.00521)
Log of Average Temperature	0.181*	0.391***	-0.00397	-0.497***	0.356***	-0.00710*	0.119***
	(0.105)	(0.0467)	(0.00473)	(0.108)	(0.0446)	(0.00415)	(0.0216)
Log Average Elevation	0.00339	0.0372***	0.000824*	-0.0326***	0.0316***	0.000937**	0.0329***
	(0.0105)	(0.00468)	(0.000473)	(0.0108)	(0.00445)	(0.000414)	(0.00216)
Log of GDPPC					-0.0894***	0.00470***	0.0184***
					(0.00374)	(0.000348)	(0.00181)
Log of Total Population				0.366***	0.0380***	0.0205***	0.138***
				(0.00982)	(0.00428)	(0.000398)	(0.00208)
Constant	10.43***	9.285***	0.168***	5.283***	9.648***	-0.0817***	-0.399***
	(0.373)	(0.166)	(0.0168)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.275	0.663	0.552	0.611	0.683	0.652	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S11-5. Impacts of geographic barriers on major socioeconomic outcome with random radius 2

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
(Random Radius 2) nonconvexity	-2.066***	1.735***	-0.156***	-0.845***	1.601***	-0.121***	0.145***
	(0.201)	(0.0895)	(0.00856)	(0.216)	(0.0895)	(0.00833)	(0.0434)
Dummy of coastal city	0.370***	-0.0665***	0.000627	-0.0488	-0.0590***	-0.00751***	0.0124**
	(0.0300)	(0.0134)	(0.00135)	(0.0308)	(0.0127)	(0.00119)	(0.00617)
Dummy of Capital City	2.990***	0.0126	0.0463***	-0.234**	-0.0123	-0.0201***	-0.160***
	(0.0943)	(0.0420)	(0.00426)	(0.0993)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00288	-0.0225*	0.00102	0.0697**	-0.0267**	0.00118	0.0386***
	(0.0283)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0406	-0.0384***	-0.000823	0.0901***	-0.0313***	-0.00199**	-0.00405
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.00521)
Log of Average Temperature	0.180*	0.398***	-0.00387	-0.507***	0.361***	-0.00730*	0.119***
	(0.105)	(0.0467)	(0.00473)	(0.108)	(0.0446)	(0.00415)	(0.0216)
Log Average Elevation	0.00209	0.0378***	0.000760	-0.0324***	0.0324***	0.000875**	0.0330***
	(0.0105)	(0.00467)	(0.000474)	(0.0107)	(0.00445)	(0.000414)	(0.00216)
Log of GDPPC					-0.0886***	0.00465***	0.0184***
					(0.00374)	(0.000348)	(0.00181)
Log of Total Population				0.365***	0.0372***	0.0205***	0.138***
				(0.00981)	(0.00428)	(0.000398)	(0.00207)
Constant	10.44***	9.259***	0.168***	5.319***	9.630***	-0.0809***	-0.400***
	(0.374)	(0.166)	(0.0169)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.274	0.664	0.551	0.611	0.683	0.651	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S11-6. Impacts of geographic barriers on major socioeconomic outcome with random radiuses 3

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
(Random Radius 3) nonconvexity	-2.104***	1.787***	-0.164***	-0.733***	1.633***	-0.128***	0.0998**
	(0.201)	(0.0893)	(0.00854)	(0.217)	(0.0896)	(0.00834)	(0.0435)
Dummy of coastal city	0.375***	-0.0716***	0.00119	-0.0513*	-0.0632***	-0.00704***	0.0138**
	(0.0301)	(0.0134)	(0.00136)	(0.0309)	(0.0128)	(0.00119)	(0.00619)
Dummy of Capital City	2.978***	0.0229	0.0454***	-0.239**	-0.00444	-0.0207***	-0.159***
	(0.0943)	(0.0419)	(0.00426)	(0.0993)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00535	-0.0246*	0.00124	0.0708**	-0.0284**	0.00130	0.0384***
	(0.0283)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0416*	-0.0395***	-0.000729	0.0898***	-0.0322***	-0.00190*	-0.00382
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.00100)	(0.00521)
Log of Average Temperature	0.194*	0.386***	-0.00296	-0.499***	0.352***	-0.00678	0.117***
	(0.105)	(0.0467)	(0.00472)	(0.108)	(0.0445)	(0.00415)	(0.0216)
Log Average Elevation	0.00334	0.0367***	0.000861*	-0.0322***	0.0314***	0.000951**	0.0330***
	(0.0105)	(0.00467)	(0.000473)	(0.0108)	(0.00445)	(0.000414)	(0.00216)
Log of GDPPC					-0.0889***	0.00466***	0.0183***
					(0.00373)	(0.000348)	(0.00181)
Log of Total Population				0.366***	0.0376***	0.0205***	0.138***
				(0.00981)	(0.00428)	(0.000398)	(0.00207)
Constant	10.38***	9.312***	0.164***	5.289***	9.668***	-0.0833***	-0.392***
	(0.373)	(0.166)	(0.0168)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,048	13,047	13,109	12,464	12,463	12,464	12,463
R-squared	0.274	0.664	0.552	0.611	0.684	0.652	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table S11-7. Impacts of geographic barriers on major socioeconomic outcome with random radius 4

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
(Random Radius 4) nonconvexity	-2.345***	1.776***	-0.172***	-0.946***	1.675***	-0.131***	0.119***
	(0.200)	(0.0890)	(0.00850)	(0.216)	(0.0893)	(0.00831)	(0.0434)
Dummy of coastal city	0.381***	-0.0691***	0.00127	-0.0445	-0.0628***	-0.00704***	0.0132**
	(0.0300)	(0.0134)	(0.00135)	(0.0308)	(0.0127)	(0.00118)	(0.00618)
Dummy of Capital City	2.978***	0.0214	0.0455***	-0.236**	-0.00908	-0.0204***	-0.160***
	(0.0942)	(0.0419)	(0.00425)	(0.0992)	(0.0410)	(0.00382)	(0.0199)
City not in Major River Watershed	0.00242	-0.0228*	0.00105	0.0696**	-0.0267**	0.00118	0.0386***
	(0.0282)	(0.0126)	(0.00127)	(0.0291)	(0.0120)	(0.00112)	(0.00584)
Log of Average Precipitation	0.0428*	-0.0394***	-0.000612	0.0910***	-0.0322***	-0.00187*	-0.00387
	(0.0253)	(0.0112)	(0.00114)	(0.0260)	(0.0107)	(0.001000)	(0.00521)
Log of Average Temperature	0.184*	0.389***	-0.00356	-0.503***	0.351***	-0.00673	0.118***
	(0.105)	(0.0467)	(0.00472)	(0.108)	(0.0445)	(0.00414)	(0.0216)
Log Average Elevation	0.00330	0.0372***	0.000844*	-0.0319***	0.0317***	0.000939**	0.0330***
	(0.0105)	(0.00467)	(0.000472)	(0.0108)	(0.00444)	(0.000413)	(0.00216)
Log of GDPPC					-0.0882***	0.00461***	0.0184***
					(0.00373)	(0.000347)	(0.00181)
Log of Total Population				0.364***	0.0387***	0.0204***	0.138***
				(0.00982)	(0.00428)	(0.000398)	(0.00208)
Constant	10.41***	9.298***	0.166***	5.314***	9.649***	-0.0819***	-0.395***
	(0.373)	(0.166)	(0.0168)	(0.394)	(0.164)	(0.0153)	(0.0796)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,047	13,046	13,108	12,463	12,462	12,463	12,462
R-squared	0.276	0.664	0.554	0.611	0.684	0.652	0.692

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.6 Case studies

As discussed in our paper, Barcelona and Atlanta had similar populations in 1990 but have since diverged significantly in urban development status, likely due to differing geographic conditions, as illustrated in Figure S22. Barcelona has a higher proportion of geographic barriers (0.286) compared to Atlanta (0.003), as well as greater nonconvexity (0.031 vs 0.002), indicating more constrained geographic conditions. This directly impacts city connectivity, with Barcelona having a detour index of 0.72—double that of Atlanta at 0.37. These conditions lead to a denser population distribution in Barcelona, evidenced by a large number of population pixels depicted in deep red, which are rare in Atlanta. Consequently, the average population density in the core built-up area of Barcelona is nearly four times that of Atlanta (34,699 per km² vs 8,210 per km²).

Similarly, a comparison between Oslo and Helsinki shows that Oslo faces more challenging geographic conditions, with a higher share of barriers (0.28) and a greater nonconvexity index (0.11). These factors result in lower connectivity and higher population density in built-up areas. These details are also depicted in Figure S22-1.

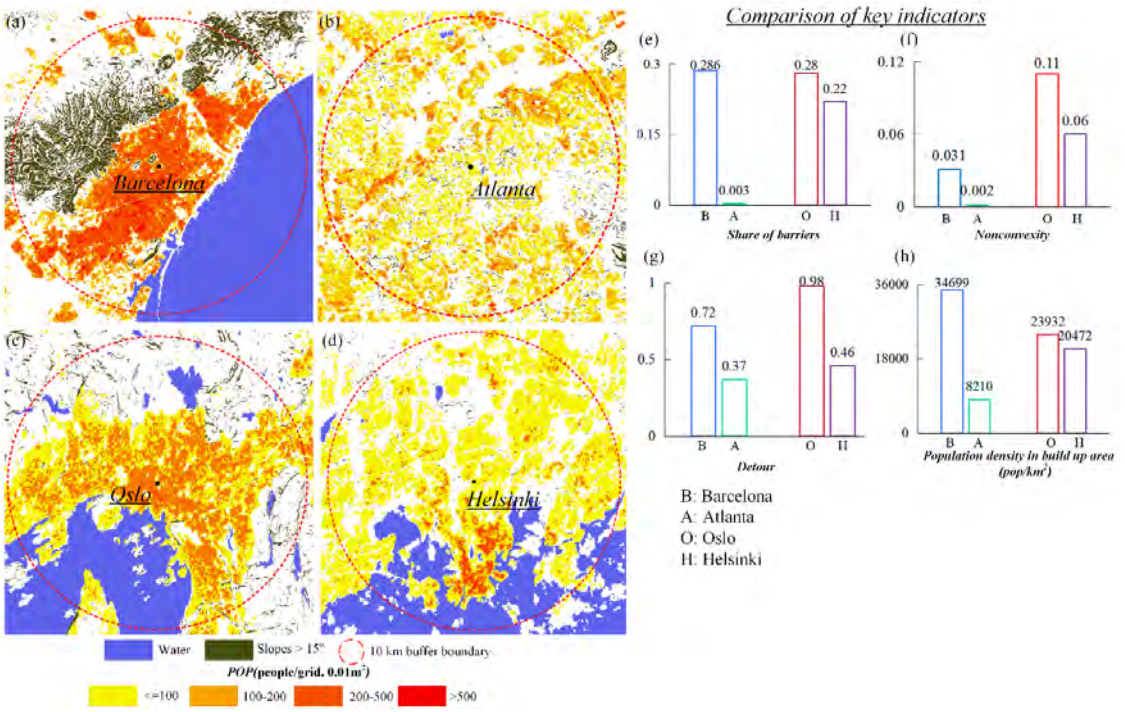

Figure S22-1. Comparisons between four cities with different geographic conditions: Barcelona, Atlanta, Oslo and Helsinki. Distribution of urban barriers and population density of the four cities in pixel level (100m resolution) **(a)-(d)**. The comparison of several key indicators in city-level **(e)-(h)**.

Note: The data used in the figure is derived from UCDB dataset.

Our results provide insights into the socioeconomic statuses of cities influenced by geographic factors, as exemplified by Lugano, Switzerland. In Figure S22-2, we compare key indicators between Lugano and the Swiss national average. Lugano exhibits more challenging geographic conditions, with a barrier share nearly three times higher (0.645 vs. 0.224) and a significantly higher

nonconvexity (0.264 vs. 0.082) than the national average. Consequently, Lugano has a larger detour index, indicating reduced connectivity in and around the city, consistent with our findings presented in Table 1 of the main paper. Socioeconomically, Lugano shows a higher population density and greater building height, consistent with the idea that a challenging geography contributes to higher urban density. Additionally, the difficult geographic conditions complicate residential construction (impervious surface), resulting in a lower proportion of built-up area (0.157) compared to the national average (0.161). These adverse conditions also correlate with a lower GDP per capita in Lugano ($34,589) than the Swiss average ($45,001). These observations align with the analyses in Table 2 of our main paper.

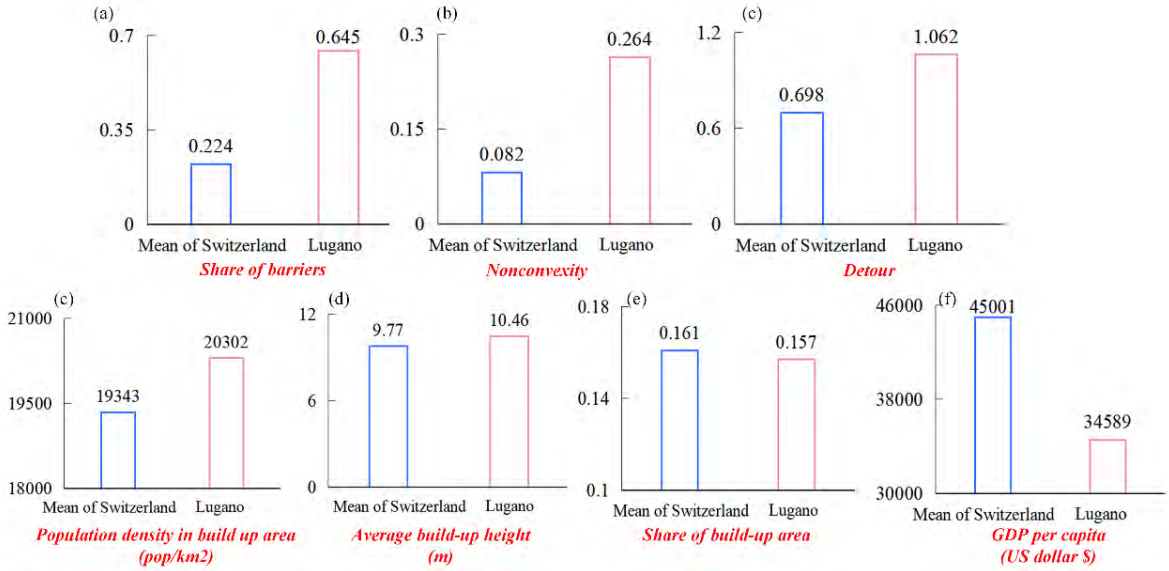

Figure S22-2. The comparison of key indicators in Lugano with the average level in Switzerland.

Note: The data used in the figure is derived from UCDB dataset.

4.7 Measurement error: robustness test using instrumental variables.

Here, we perform some extra regressions to show that our results are robust to the presence of random errors. Start by considering the detour regressions in Table 1. While measurement error is inevitable, note that we are averaging across all roads within each city buffer, so that random errors in each route cancel out. We are also **using the detour index on the left-hand side, so that measurement error cannot bias the estimates**. Consider the linear projection that summarizes the first-order average relationship between an outcome of interest in city i (Y_i) and several explanatory variables (X_i).

$$Y_i = \beta X_i + \varepsilon_i \tag{8}$$

For instance, Y_i, can refer to the "true" average detour between points, and X_i to geographic nonconvexity. Now, the issue is that Y_i is measured with error so that we observe: $Y_i^* = Y_i - u_i$, where u_i is a measurement error. Therefore, the statistical association between variables with the data at hand now becomes:

$$Y_i^* = \beta X_i + \varepsilon_i + u_i \tag{9}$$

Generally, the estimates of β are unbiased and consistent even with considerable measurement error u_i, but less efficient in small samples. Of course, this does not negate that each observation of Y_i^* may be a bit noisy: in this case, we would not advocate for using each Y_i^* as the "absolute truth" for urban dynamics in city *i*. Nonetheless statistical inference *across cities* is very accurate in large samples. Note that any measurement error on the main dependent variable, the nonconvexity index, will generally introduce conservative quantitative estimates—forcing researchers not to overstate their case.

Nonetheless, a better solution to address measurement error on the right-hand side is to use instrumental variables (Sargan, 1954). In order to allay any concerns, we use this technique in the results in Appendix Table S12. Remember that we now have 5 alternative measurements for nonconvexity based on 5 random radiuses around the UCDB city centers. Each radius may randomly encompass slightly different barriers and nonconvexity calculations. The idea is to use these alternative calculations as instrumental variables, to make sure that the "randomness" in the measurement involved in choosing a specific idiosyncratic distance bias the results. In Table S12, the measurement-error corrected results from the instrumental variables approach are—as expected—slightly larger in magnitude, further vindicating the results in the paper.

Table S12. Robustness test using instrumental variable (UCDB).

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-2.759***	2.202***	-0.206***	-0.985***	2.098***	-0.163***	0.159***
(instrumented with random radius indexes)	(0.224)	(0.0996)	(0.00971)	(0.244)	(0.101)	(0.00946)	(0.0490)
Dummy for coastal city	0.464***	-0.139***	0.00761***	-0.0190	-0.129***	-0.00193	0.00794
	(0.0322)	(0.0143)	(0.00145)	(0.0333)	(0.0138)	(0.00129)	(0.00667)
Dummy of Capital City	2.980***	0.0202	0.0456***	-0.234**	-0.0141	-0.0200***	-0.160***
	(0.0945)	(0.0420)	(0.00428)	(0.0992)	(0.0411)	(0.00384)	(0.0199)
City not in Major River Watershed	0.0116	-0.0300**	0.00169	0.0735**	-0.0344***	0.00178	0.0380***
	(0.0283)	(0.0126)	(0.00128)	(0.0291)	(0.0121)	(0.00113)	(0.00584)
Log of Average Precipitations	0.0504**	-0.0460***	1.81e-05	0.0935***	-0.0397***	-0.00130	-0.00449
	(0.0254)	(0.0113)	(0.00115)	(0.0260)	(0.0108)	(0.00101)	(0.00522)
Log of Average Temperature	0.103	0.456***	-0.00993**	-0.530***	0.418***	-0.0119***	0.123***
	(0.106)	(0.0470)	(0.00478)	(0.108)	(0.0449)	(0.00419)	(0.0217)
Log Average Elevation	0.0156	0.0272***	0.00177***	-0.0280***	0.0228***	0.00163***	0.0323***
	(0.0106)	(0.00472)	(0.000480)	(0.0108)	(0.00449)	(0.000419)	(0.00217)
Log of GDPPC					-0.0866***	0.00448***	0.0185***
					(0.00375)	(0.000350)	(0.00181)
Log of Population				0.364***	0.0394***	0.0203***	0.138***
				(0.00983)	(0.00429)	(0.000401)	(0.00208)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13,047	13,046	13,108	12,463	12,462	12,463	12,462
R-squared	0.087	0.037	0.026	0.113	0.076	0.250	0.330

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

4.8 Robustness test for large cities

Although more than 95% of the cities can be fully contained within a circle with a 10km radius, there are 299 large cities that exceed the 10 km buffer. To assess the robustness of the findings, we again used geographic maps and re-calculated our nonconvexity index at 40 km radiuses from the center for all UCDB cities for which the area exceeded 314 square kilometers. We then created a dummy for the area of the city being above said 314 square kilometers. This takes value 1 for 299 cities. Finally, we interacted that dummy by the 40km-radius-nonconvexity. This is, effectively, a spline that "quicks in" once a city reaches an area of 314 km. The specification below can be understood as examining the *additional* impact of nonconvexities beyond 10km in large cities:

$$Y_i = \beta_1 NC_{10,i} + \beta_2 NC_{40,i} \cdot D_{314.sq.km} + X_i\Phi + \varepsilon_i \quad (10)$$

Here $NC_{10,i}$ is the geographic nonconvexity at 10km (our main measure), $NC_{40,i}$ is nonconvexity in a wider 40km radius, and $D_{314.sq.km}$ is a dummy that "quicks in" only if a city reaches 314 sq.km. Other variables (obviously including $D_{314.sq.km}$ for an interacted model) are captured by X.

The results for the relevant parameters ($\beta_1$and β_1) are shown in Table S13 for the UCDB specification. The main findings are as follows:

(1) Estimates on the main coefficient (convexity measured at 10 km) are identical.

(2) The spline that applies to cities that break through 314 sq. kilometers suggests additional marginal impacts of geography as measured in larger surrounding environments. For large cities, nonconvexity at 40 kilometers radiuses has an impact above and beyond that captured by the 10 km measurement: the signs of the marginal (spline) effects are always in the same direction as the main ones.

(3) Therefore, our baseline specifications at 10km offer a conservative estimate of the impact of geographic barriers.

Table S13. Robustness test for larger cities (UCDB).

VARIABLES	(1) Log of Population	(2) Log of density	(3) Share of build-up	(4) Log of GDPPC	(5) Log of density	(6) Share of build-up	(7) Log of average height
Average Local Non-convexity, 10km	-1.636***	1.572***	-0.132***	-0.855***	1.435***	-0.0943***	0.195***
	(0.163)	(0.0775)	(0.00762)	(0.186)	(0.0773)	(0.00714)	(0.0370)
Nonconvexity at 40 km Radius x Area Above 314 sq.km	-0.438	0.425*	-0.0437*	0.167	0.441**	-0.0334*	0.291***
	(0.484)	(0.230)	(0.0234)	(0.518)	(0.215)	(0.0199)	(0.103)
Dummy of Large area	2.878***	0.0331	0.0244***	-0.859***	-0.0468	-0.0438***	-0.298***
	(0.0852)	(0.0404)	(0.00411)	(0.0960)	(0.0400)	(0.00369)	(0.0192)
Dummy of Coastal City	0.389***	-0.103***	0.00322**	-0.0361	-0.0908***	-0.00649***	0.00167
	(0.0291)	(0.0138)	(0.00140)	(0.0318)	(0.0132)	(0.00122)	(0.00632)
Dummy of Capital City	1.614***	-0.0214	0.0365***	0.0428	-0.00983	-0.00491	-0.0727***
	(0.0940)	(0.0446)	(0.00454)	(0.102)	(0.0423)	(0.00391)	(0.0203)
City not in Major River Watershed	0.0140	-0.0288**	0.00156	0.0710**	-0.0331***	0.00154	0.0375***
	(0.0265)	(0.0126)	(0.00127)	(0.0290)	(0.0120)	(0.00111)	(0.00576)
Log of Average Precipitations	0.0432*	-0.0408***	-0.000635	0.0913***	-0.0341***	-0.00191*	-0.00497
	(0.0237)	(0.0112)	(0.00114)	(0.0259)	(0.0107)	(0.000992)	(0.00515)
Log of Average Temperature	0.0878	0.420***	-0.00623	-0.513***	0.383***	-0.00796*	0.127***
	(0.0986)	(0.0468)	(0.00474)	(0.107)	(0.0446)	(0.00412)	(0.0214)
Log Average Elevation	0.00843	0.0312***	0.00126***	-0.0286***	0.0265***	0.00123***	0.0321***
	(0.00988)	(0.00469)	(0.000476)	(0.0107)	(0.00446)	(0.000412)	(0.00214)
Log of GDPPC					-0.0880***	0.00408***	0.0156***
					(0.00375)	(0.000347)	(0.00180)
Log of Population				0.409***	0.0372***	0.0233***	0.154***
				(0.0106)	(0.00465)	(0.000430)	(0.00223)
Country Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Soil Type, Climate, and Biome Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Observations	13, 048	13, 047	13, 109	12, 464	12, 463	12, 464	12, 463
R-squared	0.363	0.665	0.551	0.615	0.684	0.658	0.7

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

References

ANGEL, S., BLEI, A. M., CIVCO, D. L. & PARENT, J. 2016. *Atlas of urban expansion*, Lincoln Institute of Land Policy Cambridge, MA, USA.

BANK, W. 2016. World development indicators 2016. *In:* BANK, T. W. (ed.).

CAO, W., DONG, L., WU, L. & LIU, Y. J. R. S. O. E. 2020. Quantifying urban areas with multi-source data based on percolation theory. 241**,** 111730.

DANIELSON, J. J. & GESCH, D. B. 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010).

DIJKSTRA, L., FLORCZYK, A. J., FREIRE, S., KEMPER, T., MELCHIORRI, M., PESARESI, M. & SCHIAVINA, M. 2021. Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics,* 125**,** 103312.

DIJKSTRA, L. & POELMAN, H. 2014. A harmonised definition of cities and rural areas: the new degree of urbanisation, European Commission. *Regional Policy Working Papers.*

FLORCZYK, A. J., CORBANE, C., SCHIAVINA, M., PESARESI, M., MAFFENINI, L., MELCHIORRI, M., POLITIS, P., SABO, F., FREIRE, S. & EHRLICH, D. 2019. GHS Urban Centre Database 2015, multitemporal and multidimensional attributes, R2019A. *European Commission, Joint Research Centre (JRC)*.

FREIRE, S., MACMANUS, K., PESARESI, M., DOXSEY-WHITFIELD, E. & MILLS, J. J. P. 2016. Development of new open and free multi-temporal global population grids at 250 m resolution. 250.

GADM. 2022. *GADM maps and data* [Online]. Available: https://gadm.org/ [Accessed January 22, 2024].

GONG, P., LI, X., WANG, J., BAI, Y., CHEN, B., HU, T., LIU, X., XU, B., YANG, J. & ZHANG, W. 2020. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment,* 236**,** 111510.

GRDC 2020. Major River Basins of the World/Global Runoff Data Centre. Federal Institute of Hydrology Koblenz, Germany.

HARRIS, I., OSBORN, T. J., JONES, P. & LISTER, D. 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data,* 7**,** 109.

HOUSE, F. 2022. *Freedom in the World 2022* [Online]. Available: https://freedomhouse.org/report/freedomworld [Accessed January 15, 2024].

HUANG, X., HUANG, J., WEN, D., LI, J. J. I. J. O. A. E. O. & GEOINFORMATION 2021. An updated MODIS global urban extent product (MGUP) from 2001 to 2018 based on an automated mapping approach. 95**,** 102255.

JARVIS, A., REUTER, H. I., NELSON, A. & GUEVARA, E. J. A. F. T. C.-C. S. M. D. 2008. Hole-filled SRTM for the globe Version 4. 15**,** 5.

JONES, P. D., LISTER, D. H., OSBORN, T. J., HARPHAM, C., SALMON, M. & MORICE, C. P. J. J. O. G. R. A. 2012. Hemispheric and large‐scale land‐surface air temperature variations: An extensive revision and an update to 2010. 117.

KUMMU, M., TAKA, M. & GUILLAUME, J. H. 2018. Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. *Scientific data,* 5**,** 1-15.

LAMARCHE, C., SANTORO, M., BONTEMPS, S., D'ANDRIMONT, R., RADOUX, J., GIUSTARINI, L., BROCKMANN, C., WEVERS, J., DEFOURNY, P. & ARINO, O. 2017.

Compilation and validation of SAR and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community. *Remote Sensing,* 9**,** 36.

LI, X., GONG, P., ZHOU, Y., WANG, J., BAI, Y., CHEN, B., HU, T., XIAO, Y., XU, B. & YANG, J. J. E. R. L. 2020. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. 15**,** 094044.

LIU, F., WANG, S., XU, Y., YING, Q., YANG, F. & QIN, Y. J. P. O. 2020. Accuracy assessment of Global Human Settlement Layer (GHSL) built-up products over China. 15**,** e0233164.

MUTTI, P. R., DUBREUIL, V., BEZERRA, B. G., ARVOR, D., DE OLIVEIRA, C. P. & SANTOS E SILVA, C. M. J. A. 2020. Assessment of gridded CRU TS data for long-term climatic water balance monitoring over the São Francisco Watershed, Brazil. 11**,** 1207.

PANDEY, B., BRELSFORD, C. & SETO, K. C. 2022. Infrastructure inequality is a characteristic of urbanization. *Proceedings of the National Academy of Sciences,* 119**,** e2119890119.

PEKEL, J.-F., COTTAM, A., GORELICK, N. & BELWARD, A. S. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature,* 540**,** 418-422.

PESARESI, M., EHRLICH, D., FERRI, S., FLORCZYK, A. J., FREIRE, S., HALKIA, M., JULEA, A., KEMPER, T., SOILLE, P. & SYRRIS, V. 2016. *Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014*, Publications Office of the European Union Luxembourg.

PESARESI, M., MAFFENINI, L., FREIRE, S., POLITIS, P. & SCHIAVINA, M. 2023. GHS-SDATA R2023A - GHS supporting data. *In:* EUROPEAN COMMISSION, J. R. C. J. (ed.).

RODRIGUEZ, E., MORRIS, C. S., BELZ, J. E. J. P. E. & SENSING, R. 2006. A global assessment of the SRTM performance. 72**,** 249-260.

XU, Z., JIAO, L., LAN, T., ZHOU, Z., CUI, H., LI, C., XU, G. & LIU, Y. 2021. Mapping hierarchical urban boundaries for global urban settlements. *International Journal of Applied Earth Observation and Geoinformation,* 103**,** 102480.

YAMAZAKI, D., TRIGG, M. A. & IKESHIMA, D. J. R. S. O. E. 2015. Development of a global~ 90 m water body map using multi-temporal Landsat images. 171**,** 337-351.