WORKING PAPER SERIES
ISSN 2785-1079

Number 25

MAY 2024

URBAN COMPACTNESS AND CARBON EMISSIONS: GLOBAL EVIDENCE OVER THE PERIOD 1975-2020

Giorgio Musto

Marco Percoco

Università Bocconi
GREEN
Centre for Geography, Resources, Environment, Energy and Networks

Urban compactness and carbon emissions: Global evidence over the period 1975-2020

Giorgio Musto* Marco Percoco†

May 7, 2024

Abstract

The share of the population living in urban centres has vastly increased in recent decades, and is predicted to further expand in the future. In this context, research on the environmental impact of different urban environments, in terms of both the form and built-up structure of cities, is particularly important to understand whether smart urban design can help mitigate the nefarious impacts of climate change. This study aims at investigating relevant associations between urban form (and specifically, urban compactness) and carbon dioxide emissions of the residential and on-road transport sectors on a global scale. The study also employs a recently established, internationally comparable definition of "urban centre", which follows population-based criteria to eliminate bias from socio-cultural or administrative factors potentially determining city boundaries. The results show that lower levels of emissions of the residential and transport sectors occur in urban environments taking on more compact shapes especially in Africa and Asia, whereas the impact of urban compactness is found to be limited in Europe and North America.

*GREEN, Università Bocconi
†Department of Social and Political Sciences and GREEN, Università Bocconi

Manuscript

Urban centres have been at the core of human history since the initial agglomeration of population in 3700 BC in ancient Mesopotamia [1], when the needs to rationalise defence efforts and transport costs provided incentives for households to locate in places at the intersection of major trade routes and to build walls to protect themselves [2–4]. The reasons for the growth in size and number have remained substantially unchanged across centuries, although with the advent of the most recent phase of globalisation, cities have become the places of opportunity, where individuals have a better quality of life, can acquire knowledge, educational and health care services and may have better employment opportunities, all factors that have attracted population from non-urban to urban areas in both developing and developed countries [5–7].

As of 2020, 50% of global population live in urban areas, compared to 25% in 1950, and this share is expected to rise to 58% in 2070 [8]. These numbers are not only a synthetic measure of a wider and tumultuous urbanisation trend that has characterised recent decades, but they are also indicative of the spatial shift of consumption and its related external costs. Moving from non-urban areas to cities implies not only moving the location of consumption, but also, as individual income increases, increasing total consumption [9, 10].

This paper is the first attempt at measuring the impact of the dynamics of urban form on carbon emissions for 11,450 cities worldwide over a considerable time period, that is 1975–2020. Several socio-economic mechanisms and spatial patterns in the organisation of production activities make the association between urbanisation and the environment uncertain. On one hand, cities are major contributors to CO_2 emissions as they account for

around 70% of global emissions, mostly from industrial and motorised transport systems that are particularly concentrated around urban centres [11]. Although the concentration of population may promote forms of environmental efficiency, differences in consumption paths between urban and rural areas may lead to a global deterioration in environmental quality [12–14]. In this context, the literature on the environmental impact of urbanization has generally found a positive relationship between the share of the population living in cities and emissions, with some weak evidence in favour of the presence of an environmental Kuznets curve when accounting for income. According to this interpretation, cities at the higher end of the income distribution may end up polluting less thanks to greater technological development, service-based economies and enhanced environmental awareness of their citizens [15–19].

At the same time, however, cities take advantage of economies of scale in the provision of public services and may decrease the quantity of carbon emissions per capita [9, 20]. Figure 1 uses our dataset to estimate the share of the population and the share of total carbon emissions produced in urban centres at the global level and in each continent. The two clearly follow parallel trends and have increased over time at the global level. In our study, however, we only look at emissions directly produced in urban centres, which are less than the total consumption-based carbon footprint of cities. The lower value compared to the urban population share thus indicates not only that efficiency dynamics may be at play, but also that cities out-source a significant share of their emissions.

As cities across the world have increased in number and size in recent decades, the academic debate on the consequences of the spatial organization of urban centres has been flourishing, with a large portion of this research focusing on their internal organisation and spatial dynamics. The urban economic literature has proven theoretically and empirically

that planning for compactness as a measure against urban sprawl reduces external costs as compact cities have more regular shapes that minimize the distance between locations within their boundaries, and present higher population density [21–24]. Research inquiring into the social effects of urban compactness has found positive outcomes, such as faster population growth [25], greater satisfaction thanks to improved access to public transport and liveability [26], greater efficiency [27] and better access to facilities [28]. There exist, however, important downsides, including reduced living space, lack of affordable housing [29], overcrowding and greater exposure to air pollution [27, 30].[1]

There is a multitude of reasons for the significant impact of urban compactness on the environmental performance of cities, including economies of scale. In the United States, research using gridded population, land use and CO_2 emissions data finds that population density is negatively associated to on-road emissions in cities, while the relation is positive for urban sprawl [23]. In Japan, a study using a cross section of 50 cities finds evidence in favour of a negative association between compactness and CO_2 emissions [21]. At the same time, in China, where urbanization has grown considerably in recent decades, Ou *et al.* [22] exploit a panel of cities that expanded rapidly in the period from 1990 to 2010 to find negative associations between different compactness metrics and emissions records, and a positive link between compactness and quality of the urban road infrastructure.[2] A study for EU member states looks at the dynamics of compact cities using country-level data from 2000 to 2012, by computing the weighted average of a compactness metric for urban centres and investigating its effect on carbon dioxide emissions [35]. While a

[1] A review of the academic debate on the consequences of urban compactness versus sprawl confirms that the benefits of compact city shapes are generally found to outweigh those of less dense urban areas, while acknowledging the presence of challenges in both types of built environment [31].

[2] More studies in China and Japan yield similar conclusions [24, 32, 33], while at least one other study [34] finds conflicting results by uncovering a positive link between emissions and urban density while using cross-sectional data on Chinese urban centres in 2013.

beneficial effect emerges of both physical compactness and population density on reducing emissions, the effect of density is found to prevail over physical compactness.

Data on urban centres and CO_2 emissions

Recently, the literature on the environmental impact of cities has also focused on the availability of geo-localised data for studying their carbon footprint. However, obtaining data on emissions at an adequate resolution and quality has historically been complex: while different sources exist, their validity is often undermined by inconsistencies and endogeneity issues. Self-reported emission accounts from urban administrations, for instance, are generally obtained following different methods and rely on administrative definitions of "city" or "municipality", which are themselves influenced by historical, economic and political factors. Recent progress has been made through the effort of organisations such as the Carbon Disclosure Project, which aims at harmonizing the cross-country administrative collection of data on emissions of urban centres, by providing guidelines for reporting and collecting the resulting estimates in yearly databases [36]. These data have already been employed to produce a scientific dataset of urban emissions and ancillary variables [37].

Information from Environmentally Extended Input-Output (EEIO) Tables has been used to trace the carbon footprint of 13,000 cities all around the globe in the Gridded Global Model of City Footprints (GGMCF) [38]. The carbon footprint is a more complex concept with respect to just CO_2 emissions, as it entails an assessment of the emissions of the production processes leading to a certain product, and therefore an additional level of detail. To this end, the EEIO Tables data are used to estimate consumption flows and related carbon emissions across countries, calculating total imported and exported emis-

sions. These are then attributed to regions using subnational CF models, and eventually to cities using data on rural versus urban consumption patterns. In our study, however, we only focus on direct emissions produced in urban centres for two sectors that have the highest potential to be affected by urban form: the residential energy sector (heating and cooling for buildings) and the on-road transport sector.

Finally, a different method to estimate emissions relies on satellite information on CO_2 atmospheric concentration. This can be used to trace emissions via combination with a set of other geographical, climate and socio-economic variables in spatial models. A database relying on this methodology has provided information on CO_2 emissions of 20 cities across the world [39]. More recently, atmospheric data have been used to reach a sample size of 1,236 cities across multiple continents from 2014 to 2020 [19]. The results have also been used to investigate links between population density and the estimated emissions, finding a negative correlation between the two [19]. While promising, this type of data would still restrict our sample and prevent the sectorial differentiation of emissions data, which is an important feature of our analysis.

In general, and apart from these sources, the review of existing literature highlights a lack of global comparable evidence with time varying data on both emissions and urban forms. We attempt to partially bridge this gap with the current study, as we assemble a dataset using time-varying, high resolution spatial data on the degree of urbanization, sectorial carbon emissions, urban form indicators and ancillary variables (temperature, precipitations and GDP) for cities across the world. The database covers the 5-years intervals from 1975 to 2020. Urban boundaries are obtained starting from raster data at a 1 km resolution on the degree of urbanization from the GHS-SMOD dataset, released as part of the European Commission's Global Human Settlement Layer (GHSL) project [40].

We only select entities qualified as "urban centres", i.e. with a population of more than 50,000 inhabitants and a population density above 1,500 inhabitants per km^2, resulting in a database of 11,435 cities in 2020 (our reference year).

We assign to each "urban centre" information on CO_2 emissions drawn from the EDGAR Database [41, 42]. These are obtained for all world countries, using sector-specific indicators of human activity, technology, fuel mix and abatement percentage to estimate total emissions.[3] The country-sectorial results are down-scaled using spatial proxy data on the location of energy and industrial facilities, residential and agricultural areas, etc. to obtain the final resolution of 0.1 degrees [41, 42]. In addition to EDGAR data, we include ancillary information from other sources on the population [43] and built environment of cities [44], their climate [45] and GDP [46], and we compute urban form indicators. A comprehensive, step-by-step explanation of the data construction process is provided in the Methods section and further detailed in the Supplementary Information. We believe our data ensure consistency both in the definition of "urban centre", which is solely based on population criteria and therefore allows us to draw comparisons across continents and income groups, and in the methods used to estimate emissions, which do not resent from the sample bias and methodological heterogeneity of public emission accounts from administrative self-reporting. At the same time, using the newly released GHS-SMOD data on the degree of urbanization across the globe [40] gives us the advantage of obtaining time-varying information on evolving urban boundaries at a relatively high frequency. It allows us to build a fixed effects model and ignore city-level time invariant characteristics, such as factors linked to their geographic location or institutional setting, that may endogenously bias our results.

[3]The final estimates of CO_2 emissions in EDGAR are disaggregated by type of fuel, with a distinction between non-short-cycle-organic and short-cycle-organic fuels. These two categories are aggregated for the purpose of this study.

Changing urban form and the geometry of cities

Previous literature has focused on limited samples of cities, often for specific countries, and has made use of population density as a proxy variable for urban form. In this study, we aim to provide more compelling evidence on the impact of urban compactness on carbon emissions by taking a wider perspective, using comparable data on 11,435 cities across the world and focusing on their form, that is the shape of the built area, under the hypothesis that more compact cities are more efficient and allow for lower energy consumption, especially regarding transportation. In this respect, Figure 2 shows the evolution of the ratios of total urban area and total urban residential built-up surface to urban population from 1975 onwards. The ratios have not varied to a large extent at the global level, but with important regional differences: in Europe and Asia, urban area per citizen has increased, while it has decreased for America, Oceania and especially Africa, whose urban centres have become more dense.

In our empirical analysis, we include total urban area as an indicator of urban sprawl: fixing population by always looking at emissions in per capita terms, a larger area may be expected ex-ante to increase travel distances and be conducive to sprawl. In addition to this, we construct more elaborate geometric urban form indicators including the Compactness Index (CI), the Range Index (RI), and the Sprawl Index (SI). Figure 3 highlights the data and geometric figures involved in the computation of the CI, the RI and the SI for a random city in the sample, Milan in Italy. A more detailed explanation of the formulas and computational methods for each is provided in the Methods section.

The CI was first designed by Li & Yeh [47] in their study on the evolution of land use patterns in the Chinese Pearl River Delta during the period of fast economic growth of the

'80s and '90s, and has often been used to describe the compactness of cities [21, 35, 48]. It is a unit-less measure obtained by comparing the perimeter of the land patch of interest and the perimeter of a circle with equivalent area. The circle represents an "optimally" compact settlement, where the distance between each point is as small as possible for a given area, and the CI represents distance of the urban centre's form from this idealized shape. Higher values of the index correspond to more compact cities. The RI [25, 49] describes compactness in a slightly different way, comparing the diameters of a circle with area equivalent to the urban centre, and that of the smallest circle circumscribing it. Once again, the indicator correlates positively with compactness, but it is more sensitive to irregularities in shape that may strongly increase the diameter of the minimum enclosing circle, and is therefore less stable than other measures.

Finally, the SI was developed in seminal work on the determinants of sprawl [50] and further discussed for its compactness properties and modified in later research [49, 51]. The index draws on built-up classification raster data to measure the extent of urban sprawl in a given territory. For each pixel of residential built-up surface, it is obtained by computing the share of undeveloped cells ("open space") in a given area around the pixel, and then averaging the resulting value over all residential pixels within the territory. Thus, the SI describes the extent to which the surroundings of residential areas are also exploited and occupied by other built-up surface (for residential or non-residential use). Clearly, the surface of an urban centre should not be completely full, i.e. entirely occupied by residential built-up surface, as this would remove amenities such as parks, urban forests, waterways, etc. At the same time, however, the leapfrogging of residential development over large open areas is a symptom of urban sprawl. In urban centres, greater sprawl may reduce energy efficiency resulting from agglomeration, increase travel distances between

locations and therefore have an impact on per capita CO_2 emissions.

Association between urban form and emissions

In our empirical analysis, we estimate panel regressions with city-specific fixed effects and control for time-varying GDP, average yearly temperature and precipitations, clustering errors at the country level. Table 1 presents the coefficients on the urban form indicators for the per capita CO_2 emissions of the residential sector (heating and cooling for buildings), while Table 2 focuses on emissions from on-road transport. For the residential sector, the values of the R^2 are smaller than for the transport sector, for which the indicator is consistently high, indicating a high predictive power of the models employed. This suggests that the dynamics behind the emissions of on-road transport are more easily explained by urban form than those for heating and cooling buildings. Furthermore, GDP per capita fares better at explaining the emissions of the transport sector than the residential, as evidenced by the larger magnitude of its coefficients in Table 2 compared to Table 1.

Regarding urban form indicators, almost all coefficients are significant and with the expected sign: cities with a smaller area and more compact urban forms, proxied by higher values of the compactness and range index, tend to emit less CO_2 per citizen in the residential (heating and cooling for buildings) and the on-road transport sectors. Looking at the compactness index, a 10% decrease in the indicator (such as that experienced, for instance, by the cities of Hamburg in Germany or Bangalore in India in our sample from 2010 to 2015) is associated with an increase of about 2.4% in residential emissions, and 3.4% in on-road transport emissions. For the average city in our sample in 2010, with a population of about 290,000 inhabitants, this translates into a decrease in CO_2 emissions

for heating and cooling buildings of about 5 kg per inhabitant, for a total of 1.4 thousand tonnes. For on-road transport, while the marginal effect is stronger in magnitude, the overall increase in emissions of the average city is entirely comparable, again with about 5 more kg per citizen and 1.4 thousand tonnes for an average urban centre. Still, results are heterogeneous across countries and corresponding income groups, an aspect that we investigate in greater detail later.

For both sectors investigated, the estimated coefficient on the RI is smaller than that on the CI, despite the indexes capturing the same broad notion of compactness. This may be due to the lower stability of the RI, which is altered more strongly by small variations in city shape due to the way it is constructed. Nevertheless, the coefficient on the RI is still negative and statistically significant. For buildings, a relevant channel behind the association of these indicators with emissions may be the sharing of carbon intensive goods linked to the provision of heating and cooling: compact city development is characterised by the construction of buildings, rather than single dwelling units, which consume less energy per person and thus emit less per capita CO_2 [14, 52, 53]. For on-road transport, on the other hand, a larger area is associated with greater average distances within the urban centre, and subsequently higher emissions per person. The CI and RI capture this aspect by looking at geometric differences in urban boundaries that may alter the average distance between two random points within the city, while keeping area constant. Finally, the coefficient on the sprawl index is positive and significant for the residential sector, while it becomes negative, although insignificant, for emissions of the on-road transport sector.

To increase the robustness of our estimates, Extended Data Tables 1 to 4 provide a set of additional specifications for our analysis. As a first step, we re-estimate the

two main regressions after balancing our panel of cities. We only retain urban centres characterised as such for the entire period from 1995 to 2015 (the period for which our regressions including GDP are estimated), thus removing gaps in the data. Extended Data Table 1 shows the results of the fixed effects model estimated for the balanced panel on emissions from the residential and the transport sectors respectively: the coefficients retain the same sign and significance level in all cases. Extended Data Table 2, on the other hand, shows the results of the main regression estimated on total, rather than sectorial, per capita urban carbon emissions. The findings for the residential and transport sector appear to be transferred also to the total emission count: GDP per capita is positively and significantly associated with total emissions, while the compactness indicators appear to be beneficial towards a reduction in total emissions. The effect of the sprawl index is negative, but close to 0 and not significant, which mimics the results obtained for the on-road transport sector.

Extended Data Table 3 includes country-specific linear time trends in the model specification. The levels of carbon emissions in urban centres may indeed also depend on factors changing over the time span of our sample, such as ongoing technological progress or the results of long-term policy agendas aimed at fighting climate change by reducing emissions over time. We model this by including a generic 5-year trend in our main equation and interacting it with country-level dummies to control for specific trends in cities belonging to the same country. The general trend is negative and significant for the residential per capita CO_2 emissions, whose levels have slightly decreased globally over time, while it is not significant for per capita emissions in the on-road transport sector, whose emissions haven't followed a clear pattern and have remained quite constant at the global level, despite differences across continents. As for the coefficients on GDP and

urban form indicators, they remain substantially unaltered.

To reduce endogeneity, Extended Data Table 4 provides the results of transforming our main specification into a dynamic model including a lagged value of the dependent variable and using lagged values of the independent variables as instruments, employing GMM-IV estimations following Arellano and Bond [54]. Such a method has already been adopted in research using similar data on urban centres [55]. The results maintain the direction and significance of the main specification, confirming that the relation between urban form and emissions is not spurious for both sectors.

Heterogeneity across income and continents

Figure 4 plots the marginal effect of area and the urban form indicators, conditional on location in each continent. Extended Data Figure 1 repeats the process, but looking at heterogeneity across income classes. The marginal effects are obtained from a model including interaction terms between urban form and the continent and income dummies (see Methods). They shed light on the geographical differences in the association between urban form and the environmental impact of cities. For area, the compactness and the range index, the effect appears to vary significantly by geographic region: the associations with residential and transport emissions are generally not significant for North America and Europe, whereas they become negative and significant for Asia and South America. In general, and with the exception of the sprawl index, the marginal effects for low and lower middle income countries are more pronounced than for high income countries. This is a promising aspect, as these countries also have higher rates of urbanization and thus greater potential to curb emissions in the future, via careful urban planning.

In Africa, for instance, a 10% increase in the compactness index is associated with a

4% decrease in emissions for heating and cooling buildings, about double the association detected at the global level. For the average urban centre in the continent in 2010, with a population of 230,000, this translates into a decrease of 3 kg per citizen and about 700 tonnes in total in the city. This is a far smaller change compared to the global effect in absolute terms due to the ex-ante lower levels of emissions of African cities, but the higher marginal effect can be expected to play a more important role in the future, as African cities grow bigger and more polluting. In Asia, a 10% increase in the compactness index of the average urban centre, whose population is of about 300,000 inhabitants, is associated with a 2% decrease in carbon emissions from on-road transport, corresponding to around 2 kg per citizen and 600 tonnes in total.

The sprawl index appears to follow a different pattern, as the corresponding coefficients have a stronger magnitude for cities in Europe and North America. As the SI was designed focusing on the US context [50] and captures more complex internal structures in the urban built environment, describing interrelations between residential built-up surface and the surrounding open space, perhaps it is better suited to the analysis of cities in these continents, and especially North America. In the continent, a decrease in about 5% in the index, such as that experienced by San Diego in the United States or Quebec City in Canada between 2010 and 2015, is associated with a 3.4% decrease in residential emissions. For the average North American urban centre, this translates into 24 more kg per citizen, and 9 thousand tonnes in total at the city level.

Before concluding, heterogeneity in our sample of cities is also explored following the distribution of the dependent variable. Extended Data Table 5 shows the coefficients on our urban form indicators in a set of separate regressions (using our main specification) for each quintile in the distribution of either residential CO_2 emissions (panel a) or on-road

transport CO_2 emissions (panel b), in 1990. Despite the sample sizes being approximately the same in each of the regressions, the coefficients on the indicators are generally stronger in magnitude and significance in the lower quintiles, while significance fades away in the upper 5^{th} quintile of the emissions distribution.

Conclusion

The climate crisis is a challenge that governments and policy makers will continue to face, with mounting pressure, in the future. As people choose at an increasing rate to leave rural areas in favour of the opportunities and lifestyle offered by cities, the role of the latter as major emissions hubs is going to increase in parallel. In this context, studying the measures that can be taken to reduce CO_2 emission records of urban settlements through the design and efficiency of the built environment is of great relevance, and can help inform smart city planning. Building compact cities has long been debated as one such measure, and this study exploits a global dataset relying on a consistent definition of "urban centre" and the construction of a set of urban form indicators to shed light on the global associations between compactness and residential and transport CO_2 emissions in cities.

The results of the empirical analysis unveil a negative and significant association between compactness and per capita CO_2 emissions of the residential and transport sector. These findings are obtained while controlling for factors that notoriously affect urban ecosystems and their environmental impact, such as income, but also external climate factors, and estimating a fixed effects model to control for unobserved time-invariant characteristics of cities which may affect their environmental impact. Assessing heterogeneous effects by continent and country income class yields that compactness is found to

be more strongly negatively associated with emissions in countries between the lower and the middle end of the income distribution, suggesting compact city shapes may be more beneficial in these contexts. These results prompt for a more efficient urban planning for both the formal and informal city in Asian and African cities and its effectiveness might bring larger benefits in terms of containment of carbon emissions.

Methods

Data on urban centres

A statistical analysis of the characteristics of cities across the world and their association with CO_2 emissions requires a consistent and internationally comparable definition of "urban centre", as the characterization of a "city" may vary across countries [39]. The issue of international comparability of the information on urban centres was recently brought to the fore by the United Nations Statistical Commission, which endorsed a new, global definition of urban centres aimed at overcoming the lack of consistency in their identification across countries [56].[4] As anticipated, the new definition disregards the influence of social, economic or geographical factors varying at the country level, and defines cities in terms of their overall population and density. It characterizes a "city" as a human settlement with population above 50,000 inhabitants and average population density above 1,500 inhabitants per square km [57].

In line with these developments in official statistics, our dataset is assembled starting from information on settlement classification made available by the European Commission as part of the Global Human Settlement Layer (GHSL) project: the GHS Settlement Model Grid (GHS-SMOD). GHS-SMOD provides spatial information from 1975 to 2020, at 5 years intervals and at a 1 km resolution,[5] on the types of settlements on the Earth's surface, ranging from "rural clusters" to "urban centres" following the Degree of Urbanization definition [40]. To obtain urban boundaries, we isolate contiguous pixels classified

[4]The definition was designed in partnership with other international organisations, including: the European Union (EU), the Organization for Economic Cooperation and Development (OECD), the World Bank, UN-Habitat, the Food and Agriculture Organization (FAO) and the International Labour Organization (ILO).

[5]Each raster pixel covers an area of 1 km^2.

as "urban" in the SMOD data. We obtain information for 6,371 urban centres in 1975, almost doubling to 11,435 in 2020. The boundaries we identify do not coincide with the urban boundaries defined at the administrative level, as the criteria used in the SMOD database to classify pixels solely rely on total population and population density. To form our panel dataset, we use the most recent time period available, 2020, as reference year and assign the same panel unit ID to cities based on whether the polygons representing their boundaries in different years were overlapping with the 2020 boundaries.[6] The Supplementary Information displays graphically the raster polygonization process, mapping the original GHS-SMOD data, their transformation into the "urban centres" boundaries and the evolving boundaries over time.

Urban form indicators

Having obtained spatial information on urban boundaries, we can construct measures of urban form and structure, to determine the degree of compactness of cities. First, we easily compute total urban area for each year of data availability. In addition, we consider a set of compactness indicators: the compactness index (CI), the range index (RI), and the sprawl index (SI). The CI of a land patch is computed as the ratio between the perimeter of a circle having the same area as the land patch of interest and the area

[6]By taking 2020 as reference year, we include in the sample all cities that met the Degree of Urbanization threshold that year. Gaps are allowed, e.g. there may be urban centres that met the threshold in 1975 and 1980, fell short in the following decades and met again the threshold in 2020. In exceptional cases where two or more separate urban centres merged into a single city before 2020, we only include in the sample the city with the largest overlapping surface with the 2020 urban centre. This time flexibility in the entities considered to be "urban centres" is made possible by the use of a definition, the Degree of Urbanization, that is entirely population-based and, as anticipated, does not take into account administrative or other country-level criteria.

of the land patch itself [47, 51], according to the following equation:

$$\mathrm{CI}_i = \frac{P_i}{p_i} = \frac{2\sqrt{\pi S_i}}{p_i}$$

where p_i and P_i represent, respectively, the perimeters of urban centre i and of the equivalent-area circle, while S_i is the area of the urban centre. The index takes values between 0 and 1, where the maximum value of one would be reached in the extreme case where the city's shape is a perfect circle. The RI is similar, but it takes the ratio of the diameter of a circle having the same area as the urban centre of interest, and the diameter of the smallest circle circumscribing it. The equation:

$$RI_i = \frac{D_i}{d_i} = \frac{2}{d_i}\sqrt{\frac{S_i}{\pi}}$$

describes the index, where D_i and d_i represent, respectively, the diameters of the equivalent area circle and of the minimum enclosing circle, and S_i is once again the urban centre's surface.

To compute the SI, we draw on raster data on residential and non-residential built-up surface at a 100 m resolution from the GHS-BUILT dataset [44], part of the GHSL project. We determine a pixel to be "developed" if 20% of its 10,000 m^2 surface is built, and "residential" if 20% of its 10,000 m^2 surface is residential. For each pixel, we focus on the 810,000 m^2 square centred around it (i.e. containing 80 additional pixels each with a size of 100 m) and compute the share of pixels classified as "open space" in this area, excluding water surfaces from the count.[7] The process is automatically repeated for each "residential" pixel, and then the average within the urban boundaries is taken to obtain

[7]To include water surfaces, we use raster data from the GHS-LAND dataset at a 100 m resolution [58], and qualify a cell as "water surface" if at least 20% of its surface is covered by water.

the sprawl index.

CO_2 emissions and ancillary variables

To complete the database, we include information on our main variable of interest, carbon dioxide emissions. We complement this with data on population, GDP, average temperature and precipitations. This wealth of information is drawn from a variety of high-resolution raster data sources, listed in Supplementary Information Table 1. To merge the data together, the urban centres boundaries from the GHS-SMOD dataset are first re-projected to the same Coordinate Reference System (CRS) as the raster data. Then, raster pixels falling within the urban boundaries are assigned to the corresponding urban centre, while those falling only partially within urban boundaries are weighted based on the share of surface overlapping the urban centre. Then, a weighted sum is computed for CO_2 emissions, population and GDP, and a weighted average for temperature and precipitations. The process of merging different spatial datasets is further illustrated in the Supplementary Information.

Data on CO_2 emissions is drawn from the European Commission's Emissions Database for Global Atmospheric Research (EDGAR), providing yearly estimates of human-induced CO2 emissions (in tonnes) from 1970 to 2022 at a 0.1 degrees resolution, approximately 11.1 km at the equator [41, 42]. The estimates are available at the sectoral level: in our analysis, we include emissions from the residential sector (heating and cooling for buildings) and the on-road transport sector, as they have the highest potential to be affected by alterations in urban form and structure of the built environment. Information on gridded GDP over the territory of the urban centre is drawn from an economic atlas [46] using national statistics from the World Bank and the CIA's World Factbook, down-scaled at

a 5 arcmin resolution (approximately 9 km at the equator) via triangulation with administrative sub-national data on GDP per capita [59] and population data from the HYDE database.

Data on variables such as temperature and precipitations are also important, as climate patterns may be relevant drivers of emission levels. Low temperatures, for instance, may require increased heating in buildings and therefore increased energy consumption and emissions. For this reason, we draw on information on temperature (in °C) and precipitations (in mm) from the CRU TS v4.06 dataset, where it is available at a resolution of 0.5 degrees (approximately 56 km at the equator) as an yearly historical series from 1901 to 2021 [45]. The data are estimated via interpolation of real measurement spatial information from weather station observations across the world. Information on population at a resolution of 1 km is merged using the GHS-POP dataset, which relies on spatial census data, down-scaled using processed satellite information on the distribution, density and classification of built-up surface [43]. The total urban population is then used to transform carbon emissions and GDP in per capita terms.

Finally, cities are assigned to their country, continent and World Bank income group using information on administrative boundaries from the Global Administrative Boundaries dataset (GADM) [60]. In this case, the merged data are spatial polygons representing the boundaries of administrative entities across the world at different levels (municipalities, provinces, regions, countries and continents). We attach to each urban centre the information of the municipality overlapping its surface, and the corresponding country and continent. In cases where a city is intersecting multiple municipalities, we assign the name of the municipality (and corresponding information on country and continent) with the largest overlapping surface. We use this additional information to cluster errors at

the country level in our regressions, to implement country specific time trends, for our analysis of heterogeneity by income and continent and for descriptive purposes.

Description of the data

To provide an overview of the urban centres in the database, Extended Data Table 6 displays descriptive statistics on the average values of key variable in cities of different continents, whereas Extended Data Figure 2 plots them as points on a world map, on a colour scale based on the corresponding quintile in the distribution of CO_2 emissions in the residential sector. As expected, cities in developed regions such as North America and Europe have higher values of per capita emissions compared to their counterparts in Africa and most of Asia. Regarding the geographical distribution of urban centres, a majority of the world's urban population lives in Africa and Asia (and within the Asian continent, as evidenced by Extended Data Figure 2, in China and India, two countries that have experienced a booming urbanization rate in recent decades). Most cities in the panel are located either in lower middle income countries (40%) or in upper middle income countries (37%), and only a minority is located in low income (9%) or high income countries (14%). Urban form indicators are quite evenly distributed, even though some structural differences exist across continents (we look into this further below).

Extended Data Figure 3 zooms in on total area and the urban form indicators, the main independent variables in our analysis. Specifically, it looks at changes in a simple average of the indicators over time and by continent. In terms of total area, there has been a slight convergence in the size of the average city across continents, as mean area has slightly increased for Asia, Africa and at the global level, while it has decreased in other regions. Nevertheless, urban centres in North America and Oceania remain far bigger

on average than their counterparts in other continents. The CI and RI are persistent, and do not vary by a large amount over the time period considered (1975 to 2020). Their average values are quite similar (with the RI being slightly higher) reflecting the fact that they capture the same concept, urban form, from slightly different angles. In terms of geographical differences, urban centres in North America and Oceania appear to have smaller values of both the CI and RI, indicating less compact cities. African and European cities are instead more compact on average, as indicated by both the RI and CI. The CI and RI of Asian cities have both dropped from 1975 to 2020, suggesting that a deterioration in the compactness of these urban centres occurred in the time period considered. Finally, the sprawl index has decreased at the global level over time, and is lowest for Europe, South and North America.

Supplementary Information Figure 4 plots time trends in average population, the share of residential built-up surface, GDP and CO_2 emissions per capita over the period of data availability (which varies depending on the variable) for urban centres in each continent. All cities have experienced a rise in the share of residential built-up surface, with the world average rising from 15% in 1975 to almost 20% in 2020. This suggests that as urbanization increased across the world, the greater need for residential buildings in already existing urban centres led to an increase in the surface devoted to residential use within their borders. It also reflects the decrease in the sprawl index across continents already displayed in Extended Data Figure 3. In terms of emissions and GDP per capita, results are in line with expectations: urban centres in wealthier regions tend to emit more CO_2 on average, although they fare better now than they did in the past decades. The opposite happens in Asia, where urban emissions are lower but have increased over time, whereas Africa remains the continent with the smallest contribution to global CO_2

emissions per capita. Finally, urban GDP per capita of European, North American and Oceanian cities is above the world average, whereas Asian, African and South American cities are below.

Statistical information

In order to assess the impact of urban form on the emissions of CO_2, we estimate regressions with city-specific fixed effects that allow us to control for time-invariant factors, such as the geographic or institutional characteristics that may influence the emission levels of the city over time. The baseline specification is:

$$\ln \mathrm{CO2}_{it}^{S} = \alpha_i + \beta \ln X_{it}^{F} + \gamma \ln \mathrm{GDP}_{it} + \theta' E_{it} + \mu_t + \epsilon_{it}$$

where $\ln \mathrm{CO2}_{it}^{S}$ is the natural logarithm of per capita CO_2 emissions of sector S (where S can be either the residential or the on-road transport sector) in city i at time t. $\ln X_{it}^{F}$ is the logarithm of urban form indicator F, corresponding to the area or to one of the three urban form indicators computed ($F \in \{\mathrm{Area}, \mathrm{CI}, \mathrm{RI}, \mathrm{SI}\}$), and β is the corresponding coefficient. The remaining controls include $\ln \mathrm{GDP}_{it}$, i.e. log GDP per capita, with corresponding coefficient γ, and E_{it}, the vector of log environmental controls (namely, temperature and precipitations) with corresponding vector of coefficients θ. Finally, μ_t represents year specific fixed effects and ϵ_{it} is the error term, clustered at the country level. It should be noted that the use of natural logarithms for both the dependent and the independent variables of the regression allows for the interpretation of estimated coefficients as elasticities. The coefficient β, for instance, should be interpreted as the percentage change in per capita emissions associated with a 1% change in the urban form indicator used.

To further explore the links between urban form and emissions, while exploiting the heterogeneity in the data, we also include interaction terms between compactness and the continent and income class of the country where the city is located, to assess whether the association detected is altered in groups of cities sharing more homogeneous characteristics. The regression equation takes the form:

$$\ln \mathrm{CO2}_{it}^{S} = \alpha_i + \beta \ln X_{ict} + \sum_{G \in \mathcal{G}} \beta_G * \ln X_{it} * 1(c \in G) + \gamma \ln \mathrm{GDP}_{it} + \theta' E_{it} + \mu_t + \epsilon_{it}$$

where all variables are the same as in the previous fixed effects model, and $\ln \mathrm{CO2}_{it}^{S}$ is the natural logarithm of per capita CO_2 emissions of sector S in city i, and time t. This time, the urban form indicator $\ln X_{it}$ is interacted with a set of indicator functions taking value 1 if country c where the urban centre is located belongs to group G. The group can refer either to one of the World Bank income classes (with $G \in \mathcal{G} = \{\mathrm{LIC}, \mathrm{LMIC}, \mathrm{UMIC}, \mathrm{HIC}\}$), or to one of the 6 continents already employed for the computation of descriptive statistics ($G \in \mathcal{G} = \{\text{Africa}, \text{Asia}, \text{Europe}, \text{North America}, \text{Oceania}, \text{South America}\}$). Omitting from the regression one of the continents or income groups, the coefficient β on the urban form indicator $\ln X_{it}$ represents the effect of urban form in the omitted reference group, whereas the β_G coefficients convey the additional effect of urban form in the remaining groups. From such a model, we can estimate the marginal effect of each of the urban form indicators on per capita emissions, conditional on the city being located in a specific income class or continent, an approach allowing us to investigate heterogeneous effects.

Data availability

Both the code used to perform the spatial join of the data (in the R Markdown language) and the final database are to be made publicly available. We are in the process of final

cleaning of the code.

References

1. Reba, M., Reitsma, F. & Seto, K. C. Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000. *Scientific data* **3,** 1–16 (2016).

2. Bairoch, P. *De Jéricho à Mexico: villes et économie dans l'histoire* (1985).

3. Duranton, G. Labor specialization, transport costs, and city size. *Journal of Regional Science* **38,** 553–573 (1998).

4. Duranton, G. & Puga, D. The growth of cities. *Handbook of economic growth* **2,** 781–853 (2014).

5. Williamson, J. G. Migration and urbanization. *Handbook of development economics* **1,** 425–465 (1988).

6. Brueckner, J. K. & Lall, S. V. Cities in Developing Countries: Fueled by Rural–Urban Migration, Lacking in Tenure Security, and Short of Affordable Housing. *Handbook of regional and urban economics* **5,** 1399–1455 (2015).

7. Selod, H. & Shilpi, F. Rural-urban migration in developing countries: Lessons from the literature. *Regional Science and Urban Economics* **91,** 103713 (2021).

8. UN Human Settlement Programme. *World Cities Report 2022: Envisaging the Future of Cities* (United Nations Research Institute for Social Development, 2022).

9. Luqman, M., Rayner, P. J. & Gurney, K. R. On the impact of urbanisation on CO_2 emissions. *npj Urban Sustainability* **3,** 6 (2023).

10. Seto, K. C., Güneralp, B. & Hutyra, L. R. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences* **109,** 16083–16088 (2012).

11. Dasgupta, S., Lall, S. & Wheeler, D. *Cutting global carbon emissions: where do cities stand?* (World Bank: Washington, DC, USA, 2022).

12. Jorgenson, A. K. Consumption and Environmental Degradation: A Cross-National Analysis of the Ecological Footprint. *Social Problems* **50,** 374–394 (2003).

13. Li, Y., Zhao, R., Liu, T. & Zhao, J. Does urbanization lead to more direct and indirect household carbon dioxide emissions? Evidence from China during 1996–2012. *Journal of Cleaner Production* **102,** 103–114 (2015).

14. Fremstad, A., Underwood, A. & Zahran, S. The environmental impact of sharing: household and urban economies in CO2 emissions. *Ecological economics* **145,** 137–147 (2018).

15. Poumanyvong, P. & Kaneko, S. Does urbanization lead to less energy use and lower CO2 emissions? A cross-country analysis. *Ecological economics* **70,** 434–444 (2010).

16. Martínez-Zarzoso, I. & Maruotti, A. The impact of urbanization on CO2 emissions: evidence from developing countries. *Ecological economics* **70,** 1344–1353 (2011).

17. Sadorsky, P. The effect of urbanization on CO2 emissions in emerging economies. *Energy economics* **41,** 147–153 (2014).

18. McGee, J. A. & York, R. Asymmetric relationship of urbanization and CO2 emissions in less developed countries. *PLoS One* **13,** e0208388 (2018).

19. Dasgupta, S., Lall, S. & Wheeler, D. *Urban CO_2 emissions: A global analysis with new satellite data* (World Bank, Washington, DC, USA, 2021).

20. Gurney, K. R. *et al.* Greenhouse gas emissions from global cities under SSP/RCP scenarios, 1990 to 2100. *Global Environmental Change* **73,** 102478 (2022).

21. Makido, Y., Dhakal, S. & Yamagata, Y. Relationship between urban form and CO2 emissions: Evidence from fifty Japanese cities. *Urban Climate* **2,** 55–67 (2012).

22. Ou, J., Liu, X., Li, X. & Chen, Y. Quantifying the relationship between urban forms and carbon emissions using panel data analysis. *Landscape ecology* **28,** 1889–1907 (2013).

23. Gudipudi, R., Fluschnik, T., Ros, A. G. C., Walther, C. & Kropp, J. P. City density and CO2 efficiency. *Energy Policy* **91,** 352–361 (2016).

24. Meng, X. & Han, J. Roads, economy, population density, and CO2: A city-scaled causality analysis. *Resources, Conservation and Recycling* **128,** 508–515 (2018).

25. Harari, M. Cities in bad shape: Urban geometry in India. *American Economic Review* **110,** 2377–2421 (2020).

26. Mouratidis, K. Is compact city livable? The impact of compact versus sprawled neighbourhoods on neighbourhood satisfaction. *Urban studies* **55,** 2408–2430 (2018).

27. Yao, Y., Pan, H., Cui, X. & Wang, Z. Do compact cities have higher efficiencies of agglomeration economies? A dynamic panel model with compactness indicators. *Land Use Policy* **115,** 106005 (2022).

28. Kamble, T. & Bahadure, S. Investigating application of compact urban form in central Indian cities. *Land use policy* **109,** 105694 (2021).

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

30. Carozzi, F. & Roth, S. Dirty density: Air quality and the density of American cities. *Journal of Environmental Economics and Management* **118,** 102767 (2023).

31. Ewing, R. & Hamidi, S. Compactness versus sprawl: A review of recent evidence from the United States. *Journal of planning literature* **30,** 413–432 (2015).

32. Fang, C., Wang, S. & Li, G. Changing urban forms and carbon dioxide emissions in China: A case study of 30 provincial capital cities. *Applied energy* **158,** 519–531 (2015).

33. Lv, K., Sun, F. & Wang, L. Spatial compactness and carbon emission: Nighttime light satellite-based exposure assessment. *Regional Science Policy & Practice,* 1–17 (2023).

34. Miao, L. Examining the impact factors of urban residential energy consumption and CO2 emissions in China–Evidence from city-level data. *Ecological indicators* **73,** 29–37 (2017).

35. Xu, C., Haase, D., Su, M. & Yang, Z. The impact of urban compactness on energy-related greenhouse gas emissions across EU member states: Population density vs physical compactness. *Applied Energy* **254,** 113671 (2019).

36. Carbon Disclosure Project. *Why disclose? As a city* Accessed February 2024. https://www.cdp.net/en/cities-discloser. 2024.

37. Nangini, C. *et al.* A global dataset of CO_2 emissions and ancillary data related to emissions for 343 cities. *Scientific data* **6,** 1–29 (2019).

38. Moran, D. *et al.* Carbon footprints of 13 000 cities. *Environmental Research Letters* **13,** 064041 (2018).

39. Wu, D., Lin, J. C., Oda, T. & Kort, E. A. Space-based quantification of per capita CO2 emissions from cities. *Environmental Research Letters* **15,** 035004 (2020).

40. Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M. & Maffenini, L. *GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A* 2019.

41. 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).

42. Janssens-Maenhout, G. *et al. EDGAR v4. 3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012* 2019.

43. Schiavina, M., Freire, S. & MacManus, K. *GHS-POP R2022A—GHS Population Grid Multitemporal (1975–2030)* 2023.

44. Pesaresi, M. & Politis, P. *GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975–2030)* 2023.

45. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data* **7,** 109 (2020).

46. Kummu, M., Taka, M. & Guillaume, J. H. *Gridded global datasets for gross domestic product and Human Development Index over 1990–2015* 2018.

47. Li, X. & Yeh, A. G.-O. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and GIS. *Landscape and Urban planning* **69,** 335–354 (2004).

48. Ye, H. *et al.* A sustainable urban form: The challenges of compactness from the viewpoint of energy consumption and carbon emission. *Energy and Buildings* **93,** 90–98 (2015).

49. 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).

50. 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).

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

52. Borck, R. Will skyscrapers save the planet? Building height limits and urban greenhouse gas emissions. *Regional Science and Urban Economics* **58,** 13–25 (2016).

53. Resch, E., Bohne, R. A., Kvamsdal, T. & Lohne, J. Impact of urban density and building height on energy use in cities. *Energy Procedia* **96,** 800–814 (2016).

54. Arellano, M. & Bond, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies* **58,** 277–297 (1991).

55. Castells-Quintana, D., Dienesch, E. & Krause, M. Air pollution in an urban world: A global view on density, cities and emissions. *Ecological Economics* **189,** 107153 (2021).

56. UN Statistical Commission. *A recommendation on the method to delineate cities, urban and rural areas for international statistical comparisons* Available at: https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3j-Recommendation-E.pdf. 2020.

57. 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).

58. Pesaresi, M. & Politis, P. GHS-LAND R2022A-Land Fraction as Derived from Sentinel2 Image Composite (2018) and OSM Data. *European Commission, Joint Research Centre (JRC)* (2022).

59. Gennaioli, N., La Porta, R., Lopez-de-Silanes, F. & Shleifer, A. Human capital and regional development. *The Quarterly journal of economics* **128,** 105–164 (2013).

60. University of California, Berkeley. *Global administrative areas (boundaries), Version 4.1* Available at: https://gadm.org/data.html. Accessed 20/12/2022. 2022.

Acknowledgements

Financial support from Fondazione Invernizzi and PNRR Musa Project is gratefully acknowledged.

Author contributions

Marco Percoco has contributed to the conceptualisation of the research, to the design of the empirical analysis and to the final writing, whereas Giorgio Musto has contributed to the empirical analysis and to the final writing.

Competing interests

Authors declare no competing interests

Materials & Correspondence

Correspondence should be addressed to Marco Percoco.

Tables

Table 1: Associations between urban form indicators and residential CO_2 emissions

	(1)	(2)	(3)	(4)
GDP p.c.	0.145***	0.171***	0.175***	0.176***
	(0.047)	(0.050)	(0.050)	(0.049)
Area	0.188***			
	(0.030)			
CI		-0.244***		
		(0.035)		
RI			-0.123***	
			(0.023)	
SI				0.197**
				(0.071)
Constant	-4.271***	-3.991***	-3.963***	-3.789***
	(0.422)	(0.425)	(0.426)	(0.424)
Env. controls	Yes	Yes	Yes	Yes
Observations	53,134	53,134	53,134	53,098
R^2	0.099	0.081	0.079	0.082

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Table 2: Associations between urban form indicators and on-road transport CO_2 emissions

	(1)	(2)	(3)	(4)
GDP p.c.	0.525***	0.564***	0.570***	0.571***
	(0.155)	(0.158)	(0.157)	(0.156)
Area	0.223***			
	(0.062)			
CI		-0.338***		
		(0.077)		
RI			-0.211***	
			(0.050)	
SI				-0.120
				(0.247)
Constant	-7.483***	-7.165***	-7.139***	-7.126***
	(1.153)	(1.279)	(1.300)	(1.219)
Env. controls	Yes	Yes	Yes	Yes
Observations	20,000	20,000	20,000	19,998
R^2	0.755	0.747	0.746	0.745

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Figures

Figure 1: Changes in the share of urban population and urban CO_2 emissions over time

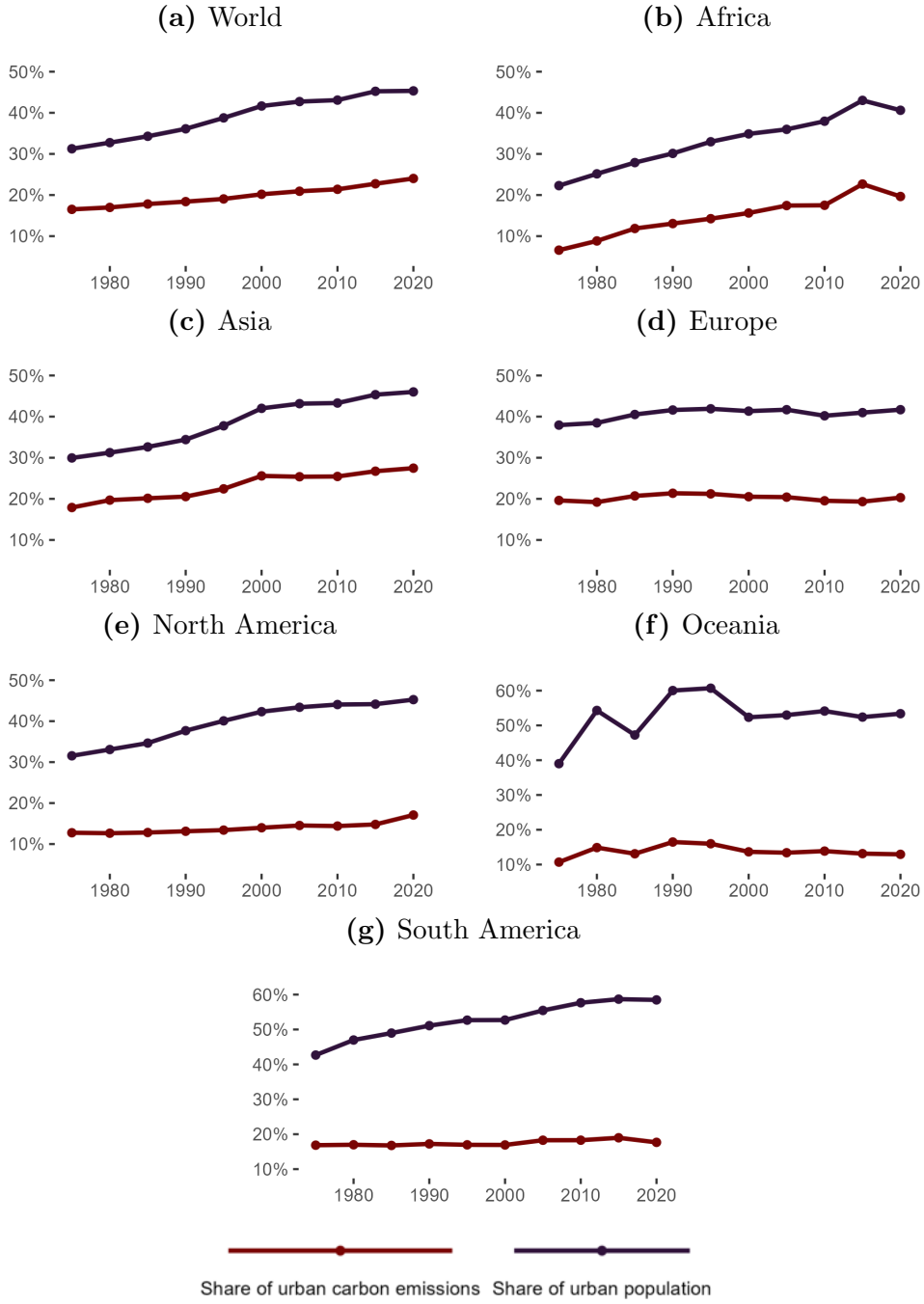
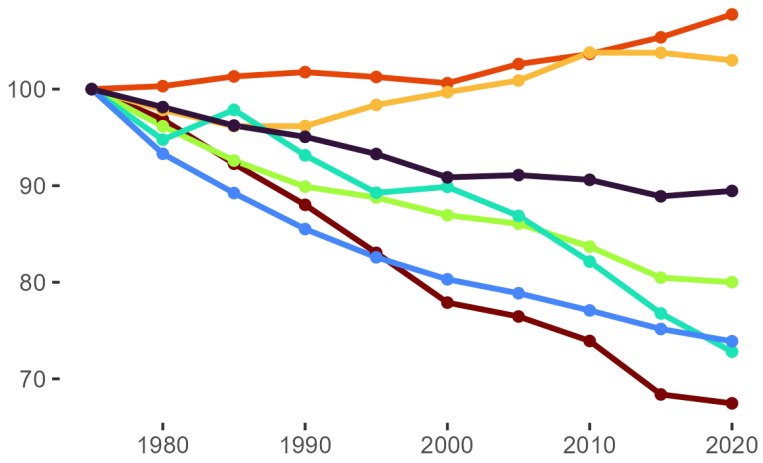
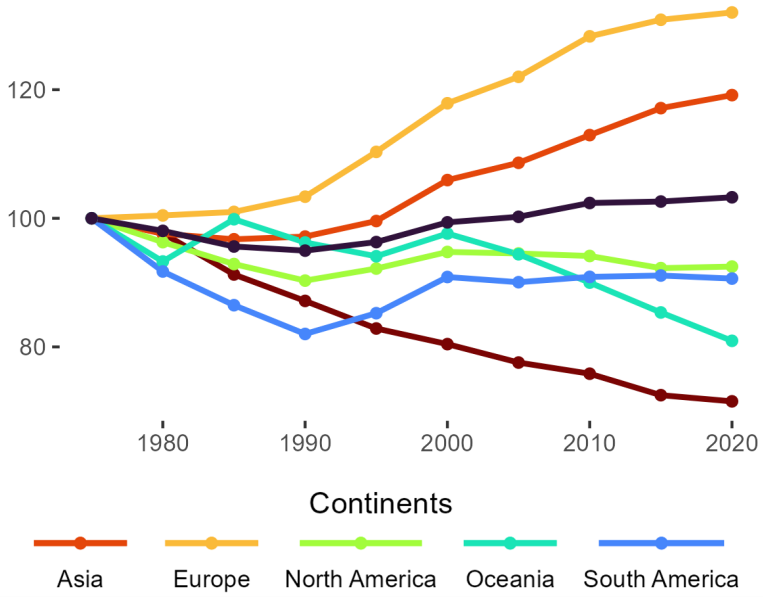


Figure 2: Time trends in urban land use and urban population

(a) Total urban area over urban population



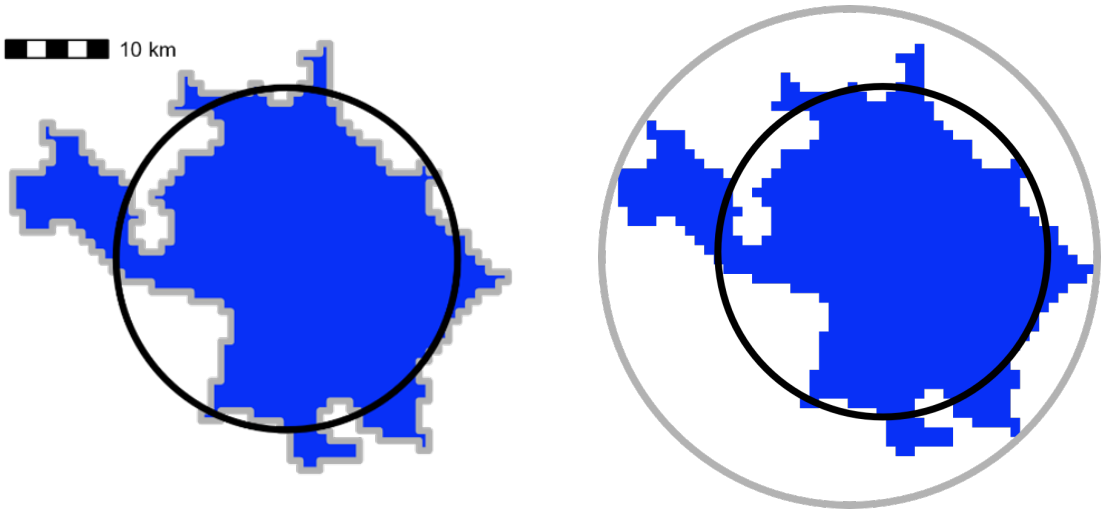
(b) Residential built-up surface over urban population



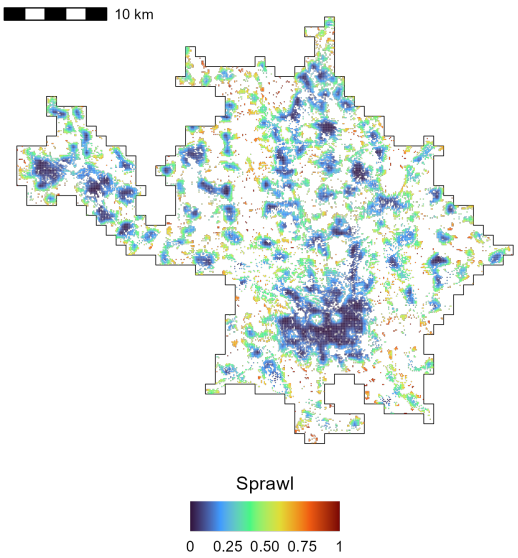
Note: the ratios are normalised to 100 in 1975, the first year of data availability, and should be interpreted in relation to that reference year.

Figure 3: Urban form indicators for a sample urban centre

(a) Compactness Index (CI) **(b)** Range Index (RI)

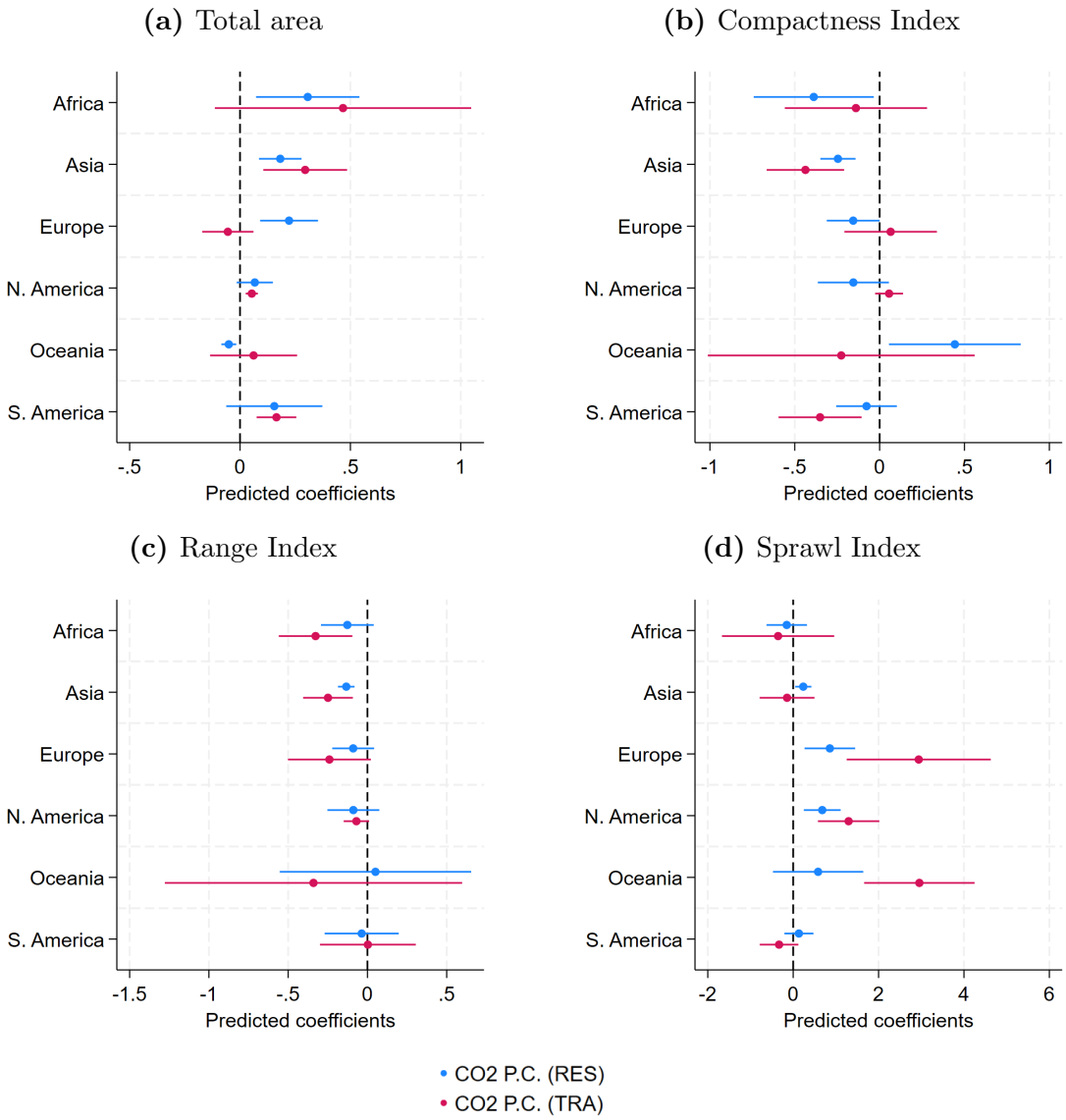


(c) Sprawl Index (SI)



Note: the maps refer to a random city in the sample, Milan in Italy, for the year 2020.

Figure 4: Heterogeneous effects of urban form indicators across continents



Note: the bands represent 95% confidence intervals around the coefficients.

Urban compactness and carbon emissions: Global evidence over the period 1975-2020

Giorgio Musto* Marco Percoco†

May 7, 2024

*GREEN, Università Bocconi
†Department of Social and Political Sciences and GREEN, Università Bocconi

Extended Data

Table 1: Associations between urban form indicators and CO_2 emissions (balanced panel)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP p.c.	0.138***	0.168***	0.171***	0.172***	0.528***	0.568***	0.573***	0.574***
	(0.047)	(0.049)	(0.049)	(0.048)	(0.150)	(0.155)	(0.155)	(0.152)
Area	0.223***				0.282***			
	(0.035)				(0.045)			
CI		-0.269***				-0.341***		
		(0.039)				(0.077)		
RI			-0.146***				-0.164***	
			(0.022)				(0.042)	
SI				0.179**				-0.207
				(0.072)				(0.248)
Constant	-4.700***	-4.328***	-4.294***	-4.123***	-7.823***	-7.285***	-7.219***	-7.286***
	(0.346)	(0.339)	(0.342)	(0.348)	(1.194)	(1.311)	(1.328)	(1.245)
Dependent var.	RES	RES	RES	RES	TRA	TRA	TRA	TRA
Env. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,402	41,402	41,402	41,396	15,052	15,052	15,052	15,051
R^2	0.112	0.089	0.086	0.088	0.772	0.762	0.760	0.760

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Table 2: Associations between urban form indicators and total CO_2 emissions

	(1)	(2)	(3)	(4)
GDP p.c.	0.400***	0.438***	0.442***	0.442***
	(0.054)	(0.053)	(0.052)	(0.052)
Area	0.253***			
	(0.030)			
CI		-0.245***		
		(0.038)		
RI			-0.072**	
			(0.028)	
SI				-0.010
				(0.104)
Constant	-4.687***	-4.287***	-4.240***	-4.220***
	(0.343)	(0.342)	(0.342)	(0.336)
Env. controls	Yes	Yes	Yes	Yes
Observations	53,685	53,685	53,685	53,648
R^2	0.441	0.424	0.423	0.423

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Table 3: Associations between urban form indicators and CO_2 emissions, with country-specific time trends

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP p.c.	0.263***	0.284***	0.287***	0.282***	0.187*	0.233*	0.240*	0.239*
	(0.043)	(0.046)	(0.046)	(0.047)	(0.104)	(0.123)	(0.125)	(0.124)
Global trend	-0.019***	-0.018***	-0.018***	-0.015***	0.010*	0.010	0.009	0.010
	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
Area	0.116***				0.225***			
	(0.025)				(0.036)			
CI		-0.160***				-0.335***		
		(0.030)				(0.063)		
RI			-0.092***				-0.176**	
			(0.018)				(0.072)	
SI				0.324***				0.186
				(0.043)				(0.127)
Constant	19.914***	19.361***	19.312***	13.538***	-90.496***	-90.011***	-89.828***	-92.580***
	(4.208)	(4.422)	(4.443)	(4.694)	(11.198)	(13.101)	(13.386)	(11.956)
Dependent var.	RES	RES	RES	RES	TRA	TRA	TRA	TRA
Env. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53,134	53,134	53,134	53,098	20,000	20,000	20,000	19,998
R^2	0.343	0.336	0.335	0.345	0.845	0.837	0.836	0.836

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Table 4: Associations between urban form indicators and CO_2 emissions, one-lag Arellano-Bond estimator

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
GDP p.c.	0.121***	0.153***	0.156***	0.176***	0.363***	0.421***	0.428***	0.433***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.014)	(0.014)	(0.014)	(0.014)
L1.CO2	-0.068***	-0.022	-0.019	0.294***	0.621***	0.619***	0.620***	0.612***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.014)	(0.015)	(0.015)	(0.016)
Area	0.170***				0.285***			
	(0.006)				(0.016)			
CI		-0.174***				-0.369***		
		(0.019)				(0.052)		
RI			-0.112***				-0.184***	
			(0.020)				(0.055)	
SI				0.535***				-0.044
				(0.014)				(0.047)
Constant	-3.842***	-3.465***	-3.445***	-2.200***	-5.747***	-5.381***	-5.342***	-5.358***
	(0.087)	(0.086)	(0.086)	(0.087)	(0.219)	(0.222)	(0.223)	(0.227)
Dependent var.	RES	RES	RES	RES	TRA	TRA	TRA	TRA
Env. controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38,276	38,276	38,276	38,262	7,910	7,910	7,910	7,910

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms.

Table 5: Regressions per quintile of the per capita residential emissions distribution

	Quintiles: 1	2	3	4	5
	Dependent var.: CO_2 emissions p.c. (residential)				
Area	0.308**	0.404***	0.213**	0.151***	0.045
	(0.132)	(0.083)	(0.081)	(0.036)	(0.036)
CI	-0.400***	-0.375***	-0.219***	-0.162***	-0.013
	(0.088)	(0.077)	(0.057)	(0.038)	(0.048)
RI	-0.268***	-0.152**	-0.110*	-0.081*	0.041
	(0.074)	(0.068)	(0.060)	(0.046)	(0.061)
SI	0.221*	0.335**	0.325***	0.177*	0.201
	(0.120)	(0.136)	(0.105)	(0.092)	(0.149)
	Dependent var.: CO_2 emissions p.c. (on-road transport)				
Area	0.682***	0.226***	0.337**	0.210**	0.083***
	(0.031)	(0.009)	(0.062)	(0.094)	(0.005)
CI	-0.566**	-0.226***	-0.200*	-0.143	-0.091
	(0.103)	(0.009)	(0.059)	(0.335)	(0.060)
RI	-0.370**	-0.024	-0.054	0.185	-0.083
	(0.072)	(0.014)	(0.050)	(0.249)	(0.090)
SI	0.876***	0.551***	0.470**	-0.298**	-0.176***
	(0.055)	(0.037)	(0.093)	(0.106)	(0.044)

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All variables are taken in natural log terms. The quintiles in the distribution of the dependent variable are calculated for the first year of data availability in the regressions, i.e. 1990 for residential emissions and 2000 for on-road transport emissions. Environmental controls (temperature and precipitations) and GDP per capita are included.

Table 6: Summary statistics by continent

Var.	Africa	Asia	Europe	North America	Oceania	South America	World
N. cities	2,133	6,677	1,125	703	79	718	11,435
LIC	38%	4%	0%	0%	0%	0%	9%
LMIC	57%	48%	8%	7%	13%	2%	40%
UMIC	5%	43%	24%	39%	29%	91%	37%
HIC	0%	5%	69%	54%	58%	7%	14%
Population	224,146	282,846	267,019	387,931	333,892	353,932	282,386
Area	37.04	43.52	61.99	144.34	175.79	66.19	52.62
CI	0.69	0.65	0.64	0.60	0.55	0.65	0.65
RI	0.68	0.64	0.66	0.63	0.56	0.66	0.65
SI	0.42	0.48	0.39	0.36	0.36	0.34	0.44
CO_2 p. c.	0.61	2	4.41	4.98	4.73	1.40	2.32
CO_2 p. c. (RES)	0.11	0.21	0.83	1.23	0.45	0.13	0.34
CO_2 p. c. (TRA)	0.03	0.15	0.38	0.93	0.70	0.17	0.25
GDP p.c.	2,372	6,292	18,022	26,561	25,219	8,494	8,938
Precipitations	955	1,234	715	1,088	1,563	1,296	1,125
Temperature	23.60	19.97	10.10	17.66	19.12	21.39	19.27

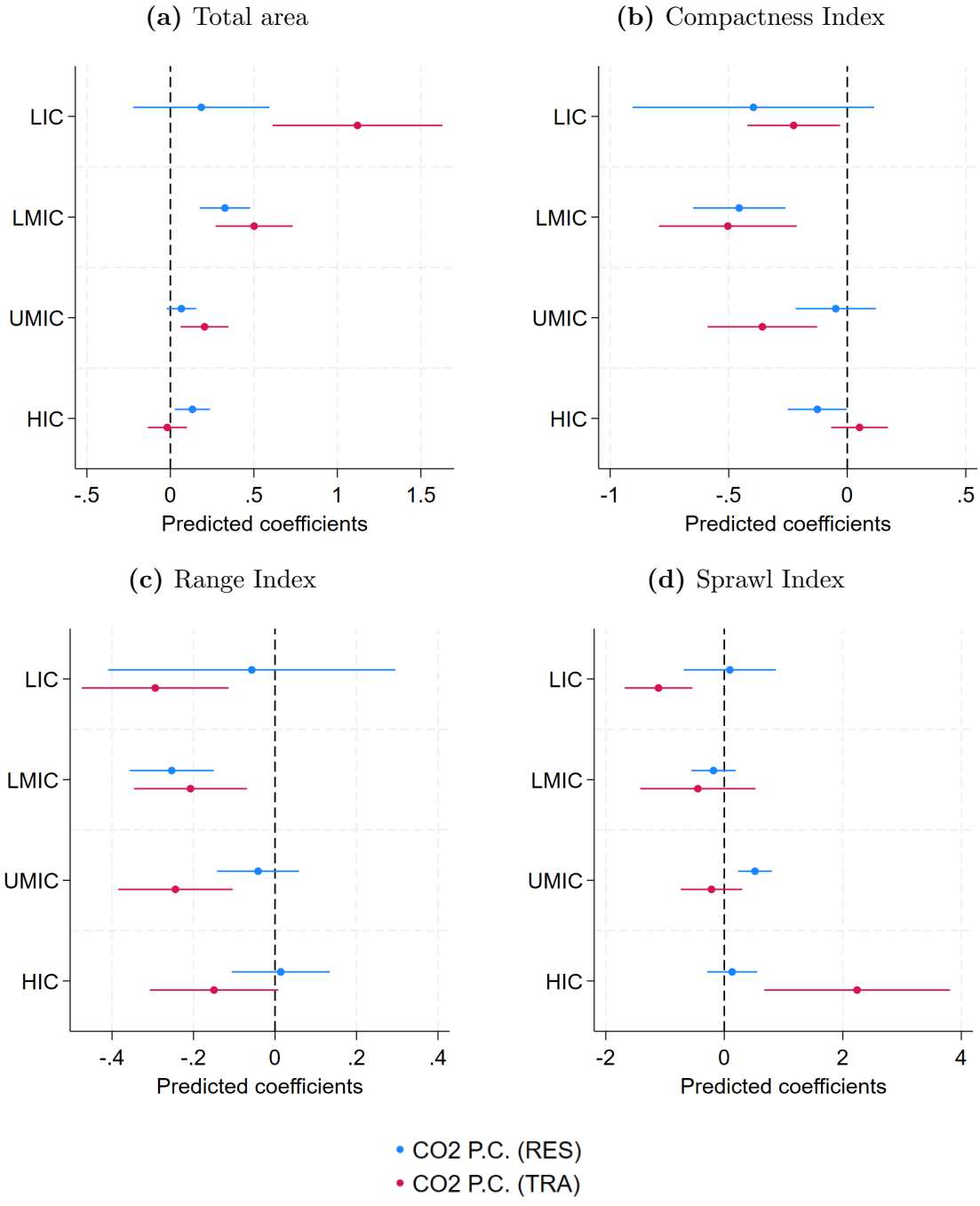
Note: CI = Compactness Index; RI = Range Index; SI = Sprawl Index.

Table 7: Correlation matrix for main regressors

	GDP p.c.	Area	CI	RI	SI
GDP p.c.	1.000				
Area	0.332***	1.000			
CI	-0.323***	-0.413***	1.000		
RI	-0.137***	-0.158***	0.752***	1.000	
SI	-0.224***	-0.163***	0.104***	-0.037***	1.000
Observations	83,201				

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Heterogeneous effects of urban form indicators across income classes



Note: the bands represent 95% confidence intervals around the coefficients.

Figure 2: World map of urban centres and CO_2 emissions per capita (residential sector) in 2020

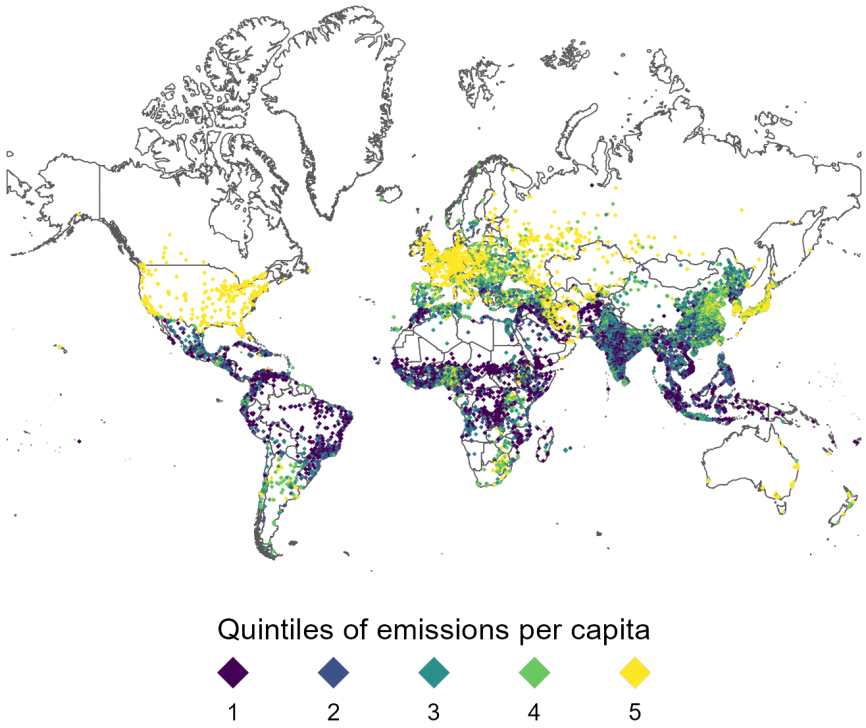
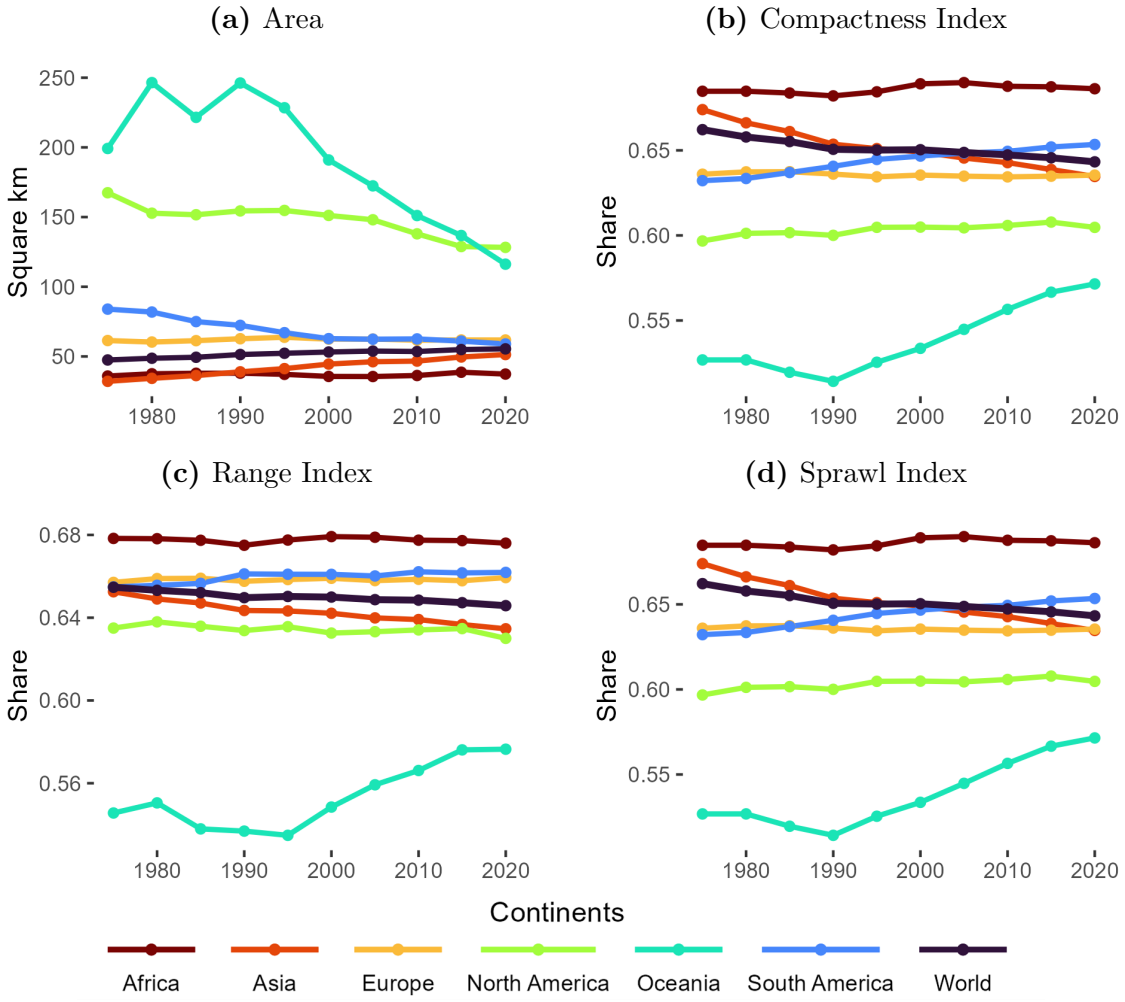


Figure 3: Changes in the share of urban population and urban CO_2 emissions over time



Supplementary information

The analysis of urban form and CO_2 emissions of the residential (heating and cooling for buildings) and on-road transport sectors is based on different spatial data sources, joined together using tools available in the R programming language. The starting point for these data consists of the urban boundaries identified following the Degree of Urbanization definition (where cities are defined based on a population threshold of at least 50,000 inhabitants and population density of at least 1,500 inhabitants per square km). To obtain the boundaries, we use the GHS-SMOD raster dataset, which codes 1 square km pixels based on whether they belong to rural or urban clusters, and assigns a specific code to pixels falling within the "Degree of Urbanization" population and density thresholds [1]. We polygonise continuous raster pixels classified as "urban centres" and obtain spatial vector data on the boundaries of each of these entities. Supplementary Information Figure 1 maps the polygonization process for a sample region, the greater Paris area in France, for our reference year, 2020 (the GHS-SMOD data cover the entire world at 5 years intervals between 1975 and 2020, and we repeat the same process each year). Supplementary Information Figure 2 shows the changes in resulting boundaries across time periods, from 1975 to 2020, for two random cities in the sample.

To assign cities to corresponding information on CO_2 emissions, population, GDP and climate variables (temperatures and precipitations) each year, we proceed with a spatial join of the different data sources available and listed in Supplementary Information Table 1. For the merge, the polygons data on urban centres obtained starting from the GHS-SMOD raster data represents the master database, to which the raster information on other variables is joined. Before implementing the merge, the raster and the polygons data

need to be projected spatially to the same Coordinate Reference System (CRS): we choose to re-project the polygons data to the same CRS as the raster data, which generally use the WGS84 projection, so we repeat this process for each spatial join. The only exceptions are information on population and built-up surface, which use the same CRS as our data on urban centres, and whose merge requires no re-projection. Supplementary Information Figure 3 shows the spatial juxtaposition of the urban boundaries and the raster data on on-road transport CO_2 emissions (taken as an example) in 2020, again for the Paris greater region. We match each polygon to the overlapping raster pixels, and we compute the share of raster cell surface that intersects the urban centre (the "coverage fraction"). This share will be equal to 1 for cells entirely inside the urban centre: in Supplementary Information Figure 3, for instance, the raster pixel highlighted with an arrow has a value of emissions of 37,513 tonnes. Being entirely contained in the Paris urban centre, it is assigned weight 1 in the computation. Finally, we sum together all the emissions values within the polygon boundaries, multiplied by the weights corresponding to the coverage fraction values. We thus come up with estimates of values of the raster variables (e.g. emissions, GDP, etc.) within the territory of the urban centre. The assumption being made to aggregate the raster data at the urban level is that the values of raster variables are uniformly distributed across the surface occupied by the raster pixel. This assumption is more believable for climate variables such as temperature and precipitations, while it becomes more problematic for CO_2 or GDP, whose values are clearly more concentrated where human activity is stronger. Nevertheless, the method allows for consistency in the way we estimate key variables across cities and provides us with an overview of urban socio-economic variables. Furthermore, for data at a higher resolution the bias due to the merge is reduced, even though estimates may be less accurate precisely because of the more pronounced spatial down-scaling. For information on population and built-up

surface, the match with the urban boundaries is most reliable as the sources come from the same wider GHSL data project and share the same CRS and high resolution of 1 km.

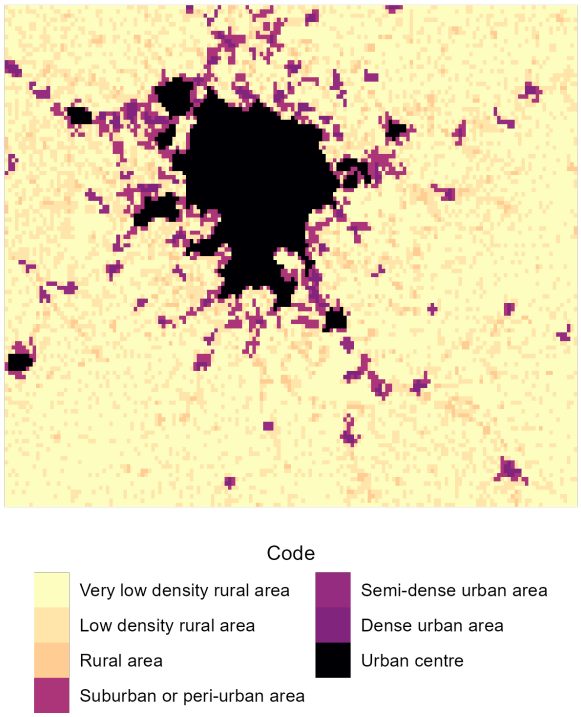
Time trends for CO_2 emissions and other ancillary variables obtained from the merge are displayed in Supplementary Information Figure 4, illustrating how urban population, the share of built-up surface, per capita GDP and CO_2 emissions evolved over the time span of our sample. Finally, we also assign to each urban centre information on the corresponding municipality, region and country using the Global Administrative Boundaries (GADM) database [2]. In this case, the spatial join is performed between two spatial vector datasets with polygon geometry, and the boundaries of polygons in the two datasets do not necessarily coincide. A city defined using the Degree of Urbanization criterion may in fact overlap the boundaries of different administrative-level municipalities. To merge the data, we assign to each urban centre the administrative unit with the largest overlapping surface, and the corresponding region, country, continent and country income class.

Table 1: List of data sources

Data sources	**Resolution**	**Years**	**Type of spatial join**	**Main variables**
GHS–SMOD 2023 [1]	1 km	1975–2020 (5 years intervals)	Master data	Urban boundaries
GHS–POP 2023 [3]	1 km	1975–2020 (5 years intervals)	Sum within urban boundaries	Population
GHS–BUILT 2023 [4]	1 km	1975–2020 (5 years intervals)	Sum within urban boundaries	Built-up surface: total, residential and non-residential (m^2)
EDGAR v7.1 [5, 6]	0.1 degrees (approx. 11.1x11.1 km at the equator)	1970–2021 (yearly)	Weighted sum within urban boundaries	CO_2 emissions for heating and cooling of buildings, from on-road transport and total (tonnes)
CRU TS v4.07 [7]	0.5 degrees (approx. 56x56 km at the equator)	1901–2021 (yearly)	Weighted sum within urban boundaries	Temperature (°C) and precipitations (mm)
Gridded data on GDP [8]	5 arcmin (approx. 9x9 km at the equator)	1990–2015 (yearly)	Weighted sum within urban boundaries	GDP (PPP, 2011 USD)
GADM Dataset [2]	Geo-referenced boundaries of adminis-trative units	2022	The name of the administrative unit with largest overlapping surface was joined	Names of admin. Units (country name, province name, municipality name)

Figure 1: Polygonization of GHS-SMOD raster data for urban centres

(a) GHS-SMOD raster data



(b) Polygonized raster data

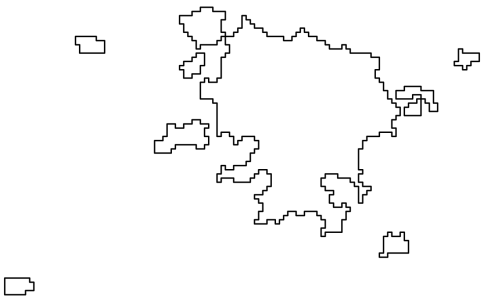
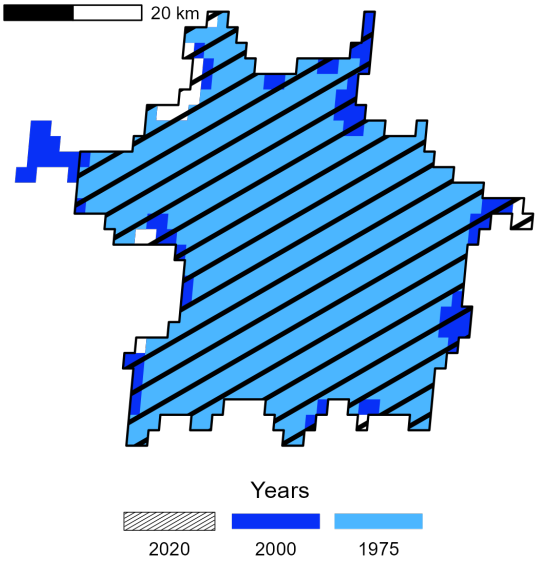


Figure 2: Changes in urban boundaries at selected time periods for two cities in the sample

(a) Berlin, Germany



(b) Havana, Cuba

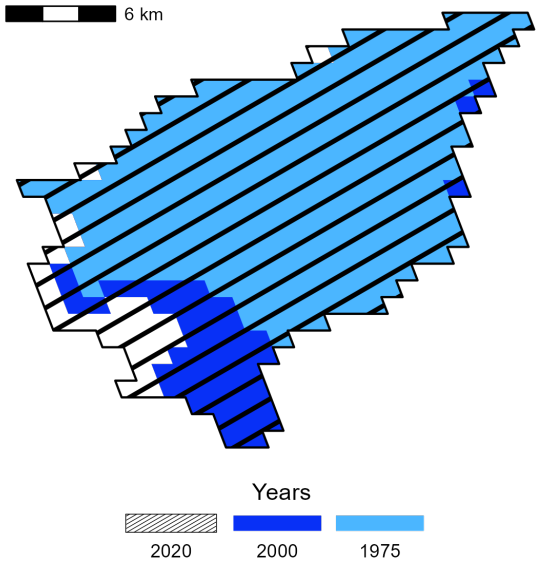
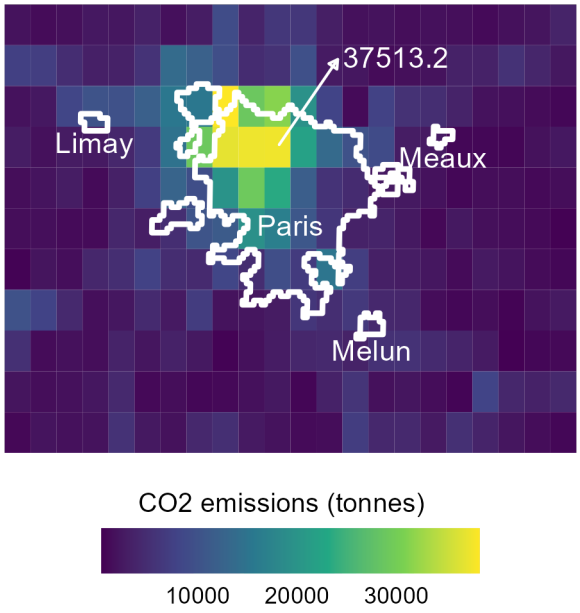


Figure 3: Spatial juxtaposition of the boundaries of urban centres and on-road transport CO_2 emissions data in 2020 for the Paris region



References

1. Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M. & Maffenini, L. *GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A* 2019.

2. University of California, Berkeley. *Global administrative areas (boundaries), Version 4.1* Available at: https://gadm.org/data.html. Accessed 20/12/2022. 2022.

3. Schiavina, M., Freire, S. & MacManus, K. *GHS-POP R2022A—GHS Population Grid Multitemporal (1975–2030)* 2023.

4. Pesaresi, M. & Politis, P. *GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel2 composite and Landsat, multitemporal (1975–2030)* 2023.

5. 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).

6. Janssens-Maenhout, G. *et al. EDGAR v4. 3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012* 2019.

7. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data* **7,** 109 (2020).

8. Kummu, M., Taka, M. & Guillaume, J. H. *Gridded global datasets for gross domestic product and Human Development Index over 1990–2015* 2018.

GREEN
Centre for Geography, Resources, Environment, Energy and Networks
via Röntgen, 1
20136 Milano - Italia

www.green.unibocconi.eu