Gallé, Johannes

Working Paper

City Shape and Air Pollution

Ruhr Economic Papers, No. 1012

Provided in Cooperation with:
RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

Suggested Citation: Gallé, Johannes (2023) : City Shape and Air Pollution, Ruhr Economic Papers, No. 1012, ISBN 978-3-96973-178-9, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, https://doi.org/10.4419/96973178

This Version is available at:
https://hdl.handle.net/10419/270957

RUHR
ECONOMIC PAPERS

Johannes Gallé

City Shape and Air Pollution

RUB

#1012

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Ludger Linnemann
Technische Universität Dortmund, Department of Business and Economics
Economics – Applied Economics
Phone: +49 (0) 231/7 55-3102, e-mail: : Ludger.Linnemann@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt, Prof. Dr. Ansgar Wübker
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #1012

Responsible Editor: Thomas Bauer

ISSN 1864-4872 (online) – ISBN 978-3-96973-178-9

Ruhr Economic Papers #1012

Johannes Gallé

City Shape and Air Pollution

RUB

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/96973178
ISSN 1864-4872 (online)
ISBN 978-3-96973-178-9

Johannes Gallé[1]

City Shape and Air Pollution

Abstract

Air pollution has become an increasing health threat for the local population in many cities around the world. Using high resolution remote sensing data on nightlights and fine particulate matter ($PM_{2.5}$) for the years 1998-2013, I study the contemporary nexus between city shape and air pollution in India. I find that the compactness of a city has statistically significant and negative effects on local air quality. The results are more pronounced in larger cities and robust with respect to different compactness measures. While geographic dispersion allows for more fresh air corridors, differences in commuting patterns could serve as an additional explanation. People in less compact cities are more likely to use public transport and thereby reducing the overall road traffic within cities translating into less pollution. However, the statistically significant effects do not translate into substantial changes in the relative risk of $PM_{2.5}$-induced diseases.

JEL-Codes: R10, R41, Q53

Keywords: Urbanization; air pollution; commuting; India

April 2023

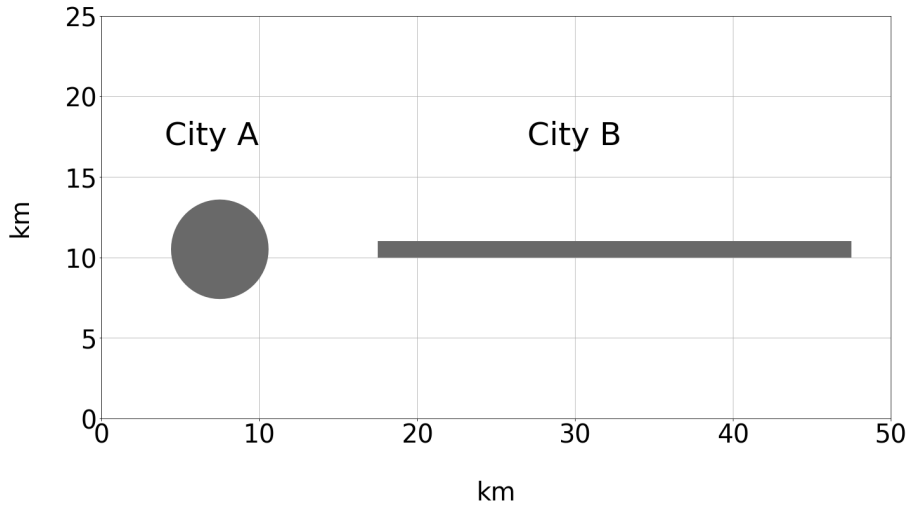
1 Johannes Gallé, RUB. *– I am grateful for valuable feedback from colleagues at the RTG 2484 "Regional Disparities and Economic Policiy" and the Ruhr University Bochum. Furthermore, I thank participants at the 34th ERSA Summer School 2021 and at the 60th ERSA Congress 2021 for helpful comments. Finally, I gratefully acknowledge financial support from the German Research Foundation (DFG).* *– All correspondence to: Johannes Gallé, RUB, Universitätsstr. 150, 44801 Bochum, Germany, e-mail: johannes.galle@ruhr-uni-bochum.de*

1 Introduction

Agglomeration economies come with many benefits (Fujita and Thisse, 1996; Combes et al., 2008; Redding and Rossi-Hansberg, 2017). Yet, they also create negative local externalities such as air pollution, which poses a severe health risk to residents in cities around the world. In 2019, the World Health Organization (WHO) estimated that globally 4.2 million premature deaths were attributable to ambient outdoor air pollution (WHO, 2022). Especially urban agglomerations in developing countries, such as in India, face high levels of air pollution. According to the World Air report 2019, 21 out of the 30^{st} most polluted cities worldwide were located in India (IQAir, 2019).

Some recent literature has studied the relationship between urban population density and air pollution (Ahlfeldt and Pietrostefani, 2019; Borck and Schrauth, 2021; Piracha and Chaudhary, 2022; Borck and Schrauth, 2022; Carozzi and Roth, 2023). However, while population density is informative about how many people live within the total of a given area, density measures abstract from the actual shape of that area. Therefore, standard density measures may miss crucial aspects of urban properties such as the compactness of a city (Duranton and Puga, 2020; Harari, 2020). Although the terms "compactness" and "density" are closely related and often used interchangeably in the literature (Ahfeldt and Pietrostefani, 2017), they follow two different concepts. While density is the simple ratio between mass (e.g. population) and volume (urban area), compactness relies on the idea that for a given area a circle represents the most compact form (Angel et al., 2010). Put differently, compactness indices measure the deviation of the shape of a given area from the shape of a circle with the same area. Figure 1 illustrates the difference between "density" and "compactness" based on two stylized cities.

Figure 1: Compactness and the shape of cities



Notes: Figure 1 illustrates the compactness of urban footprints. Both cities, A and B, have the identical area of $30km^2$. Given that a circle constitutes the most compact shape for a given area, city A is considered to be more compact than city B. Own illustration.

The size of both cities, A and B, equals $30km^2$. Further, assuming that both cities have the same population, the population density of both cities would be identical. Yet, the shape of both cities clearly differs. While city A's shape constitutes a perfect circle, city B is shaped like a sustained rectangle.[1] Despite of having the same population density, city A is considered to be more compact than city B.

This paper asks the question, whether the compactness of urban shapes impacts local air quality. Beside the fact that the dispersion of less compact cities allows for more fresh air corridors (Borrego et al., 2006), the shape of a city determines the average commuting distances within cities. This could affect overall road traffic, which is considered to be a substantial contributor to air pollution in urban areas (Badami, 2005; Pant and Harrison, 2013; Vohra et al., 2021). Assuming that air pollution is a positive increasing function of total road traffic within a city, it is a priori not clear whether the effect of city compactness on total road traffic is positive or negative. On the one hand, less compact cities are associated with on average larger commuting distances. On the other hand, longer distances within cities could discourage residents from undertaking many trips reducing the overall amount of within city trips taken by residents leading to less aggregate traffic and congestion. Furthermore, city shape may impact the modal choices of commuters and thereby affecting the total amount of within city traffic.

In order to answer the research question, I identify urban footprints based on remote sensing nightlight data (NOAA, 1992-). The data on urban footprints is further complemented with remote sensing data on outdoor air pollution measured by ground-level fine particulate matter ($PM_{2.5}$) (Van Donkelaar et al., 2016) and with population information from the Population Census 2001 (Meiyappan et al., 2018). By restricting the sample to cities with 50.000 or more inhabitants as of 2001, the final sample comprises a panel of 443 Indian cities for the years 1998-2013 containing annual information on compactness measures as well as on mean, maximum and minimum $PM_{2.5}$ exposure. In the baseline specification, I employ the disconnection index to measure city compactness. The index is defined as the average distance between any pair of points in a city serving as a proxy for the average commuting distance. I extend previous work by Harari (2020), who studies the impact of city shape on economic outcomes such as population growth and wages. A causal relationship between city shape and air pollution is estimated by following the instrumental variable approach of Harari (2020), which is based on a combination of historic urban footprints, historic population growth and geographical obstacles.

I find that city compactness has a statistically significant and negative effect on air quality. Conditioning on total urban area, a 1km increase in the disconnection index reduces the average $PM_{2.5}$ exposure by 0.6%. The effect is robust to alternative compactness measures as well as using maximum or minimum $PM_{2.5}$ exposure as alternative dependant variables. The effect size increases with city size. In cities with more than 500.000

[1] Although the shapes of the cities are highly stylized, they are relevant in contemporary urban planning projects. Saudi Arabia is currently constructing a 170 kilometres line-stretched and only 200m wide city aiming to host about 9mio people (CNN, 2022).

inhabitants (as of 2001) a 1km increase in the disconnection index decreases the average $PM_{2.5}$ exposure by 1.1%. Albeit data limitations, I study commuting as a potential mechanism leading to the negative effect of compactness on air quality. Although the actual length of commutes taken by residents is larger in less compact cities, residents in less compact cites are more likely to use public transport for their commute such as trains or busses. In compact cities, residents are more likely to commute by individual vehicles such as scooters. Hence, one potential channel could be that due to longer commutes in less compact cities, residents rely on public transport leading to fewer overall traffic and congestion in cities, resulting in less air pollution. Lastly, I asses the statistically significant compactness effects with respect to their health implications. I find that the compactness-induced changes in $PM_{2.5}$ exposure translate into very small changes in relative risks of diseases that are frequently associated with $PM_{2.5}$ exposure (e.g. lung cancer or chronic obstructive pulmonary diseases). Evaluated at the sample mean of $PM_{2.5}$, a one standard deviation increase in the disconnection index reduces the relative risk of lung cancer by 1.1% (3% for large cities). This result is primarily driven by the already high levels of air pollution in Indian cities and the non-linear relationship between relative health risk and $PM_{2.5}$ exposure.

By focusing on alternative city characteristics such as compactness, the findings contribute to a growing body of literature that examines the relationship between urbanization and air pollution. The papers most closely related to this one are Borck and Schrauth (2021) and Carozzi and Roth (2023), who find a positive impact of population density on air pollution in Germany and the US respectively. Extending the framework of Harari (2020), this paper is -to the best of my knowledge- the first paper establishing a causal link between city shape and air pollution. Further, the geographical focus of this paper is on India, which hosts the second largest urban population of almost 500mio people (UN, 2018) and characterized by extremely high levels of urban air pollution (IQAir, 2019).

The study further relates more broadly to the literature studying the causal effects of ambient air pollution on human health (Lepeule et al., 2012; Greenstone and Hanna, 2014; Chowdhury and Dey, 2016; Ghude et al., 2016; Fowlie et al., 2019) and other outcomes such as labor productivity or education (Graff Zivin and Neidell, 2012; Ebenstein et al., 2016; Balakrishnan and Tsaneva, 2021; Aguilar-Gomez et al., 2022).

The remainder of the paper is structured as follows. Section 2 describes the identification strategy followed by the data section 3. Section 4 presents the results and section 5 concludes.

2 Empirical approach

To grasp a first idea on the empirical relation between city shape and $PM_{2.5}$ concentration, I employ a standard two-way fixed effects model, estimated as follows:

$$log(PM2.5_{it}) = \beta_1 shape_{it} + \beta_2 ln(area_{it}) + \alpha_i + \gamma_t + \epsilon_{it}, \quad (1)$$

where the dependent variable $PM2.5_{it}$ is the log of the average $PM_{2.5}$ concentration measured in micrograms per cubic meter ($\frac{\mu g}{m^3}$) in city i in year t. The variable $shape_{it}$ is the respective shape property, that measures the compactness (in km) of city i in year t, where a higher value indicates a less compact city shape. Since compactness measures are mechanically correlated with the total area of a city, I control for the log of the total area of a city. The model is complemented by city fixed effects α_i to control for time-invariant unobservable city heterogeneity. Year fixed effects γ_t control for common annual shocks including measurement differences in the remote sensing data due to annual differences in satellite composition. The standard errors are clustered at the city-level.

The main challenge of identifying a causal effect of city shape on $PM_{2.5}$ concentration is that urban footprints, defining the shape of a city ($shape_{it}$), are endogenous. The shape of a city is an outcome determined by exogenous factors such as geography as well as endogenous factors such as urban planning policies, institutional capacity as well as individual location decisions by firms and residents (Harari, 2020). Cities with a better local institutional capacity might implement urban planning policies to grow more compact, which is considered be be economically more efficient. Furthermore, cities with a higher institutional quality might have better policy enforcement on local air quality regulations leading to lower $PM_{2.5}$ concentration. In order to isolate the exogenous part that determines city shape, I adopt a slightly modified version of the instrumental variable approach proposed by Harari (2020). The basic idea is to instrument the actual shape of the urban footprint of city i in year t, with the shape of each city's potential footprint over time, that is determined by exogenous factors such as geography and uncorrelated with the error term. The time varying potential urban footprints are constructed by combining historic urban footprints obtained from georeferenced maps (U.S. Army Map Service, 1955) with historic population information (Mitra, 1980) and detailed information on geographical obstacles for urban development such as mountain ranges or water bodies (METI and NASA, 2019a,b). Details on the construction of the potential footprints are given in the next section 3. The obtained potential footprints are used to calculate the shape properties measuring the potential compactness of city i in year t.

This results in the following first-stage regression:

$$shape_{it} = \delta_1 \widetilde{shape_{it}} + \delta_2 ln(area_{it}) + \alpha_i + \gamma_t + \epsilon_{it}, \quad (2)$$

where the actual shape of city i in year t ($shape_{it}$) is regressed on its potential shape ($\widetilde{shape_{it}}$), while controlling for total area, city and year fixed effects.

The second-stage is written as follows:

$$log(PM2.5_{it}) = \delta_1 \widehat{shape_{it}} + \delta_2 ln(area_{it}) + \alpha_i + \gamma_t + \epsilon_{it}, \quad (3)$$

where the variable definitions correspond to Eq. 1, except that the actual shape ($shape_{it}$) is replaced by $\widehat{shape_{it}}$, which corresponds to the predicted shape obtained from the first-stage Eq. 2.

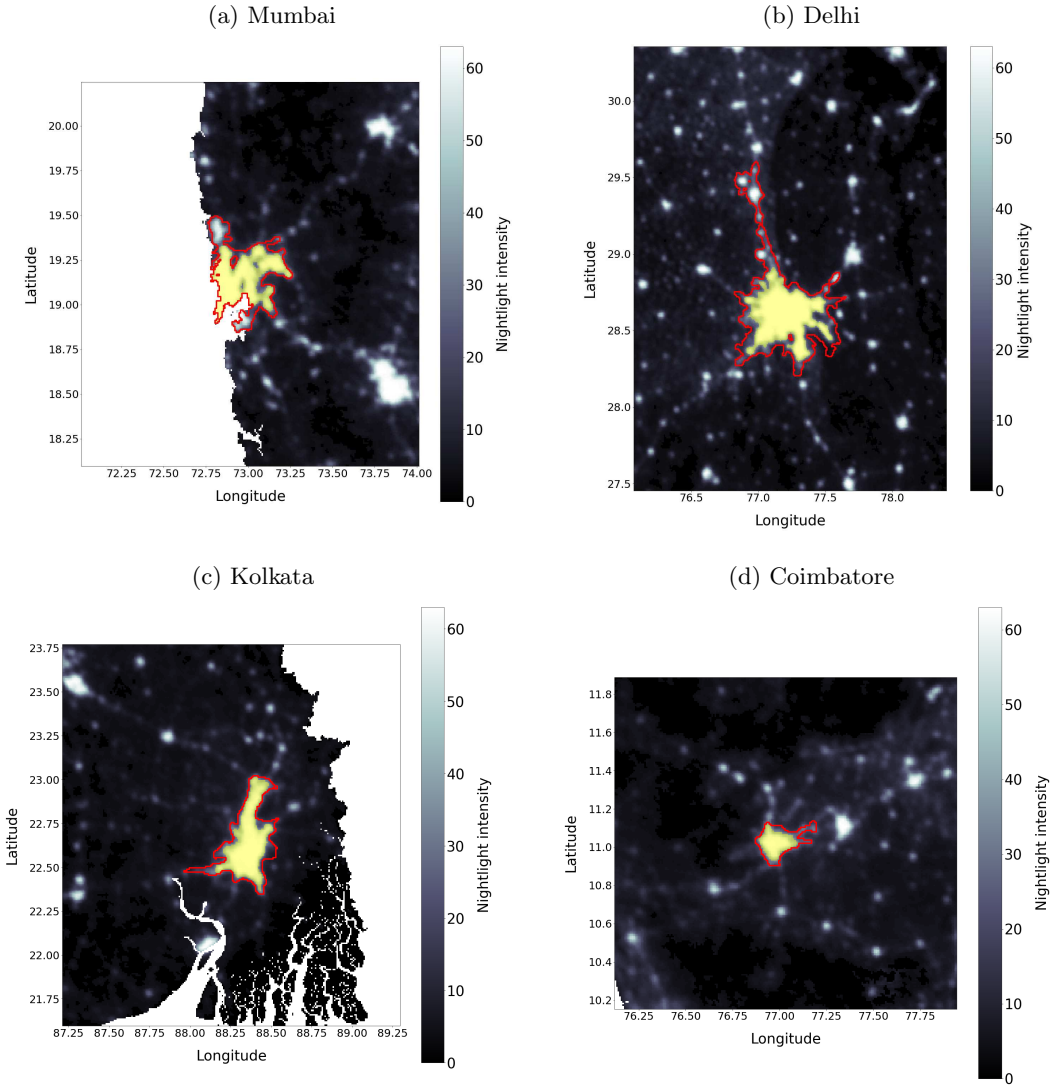
3 Data

Urban footprints. To delineate and track urban footprints of Indian cities over time, I use the Version 4 DMSP-OLS Nighttime Lights Time Series from NOAA's Earth Observation Group (NOAA, 1992-), which is a method in line with previous literature (Henderson, 2003; Small et al., 2005; Balk et al., 2006; Dingel et al., 2019; Harari, 2020). While commuting-based definitions of urban areas are often not available for developing countries, such as in India, Dingel et al. (2019) show that nightlight-based urban areas serve as good alternative proxies. Further, the use of nightlights allows to track the growth of urban footprints over time, which is of specific relevance in developing countries, where urbanization happens at a much faster pace than in developed countries nowadays. The DMSP-OLS nighttime lights data comes at a spatial resolution of 30 arc seconds, which roughly corresponds to 1km x 1km at the equator. The scale of nightlight intensity ranges from 0 (no light) to 63 (top-coded brightest value). Following the approach of Harari (2020), I consider all pixels as urban with a nightlight intensity ≥ 35. The urban footprints are constructed by dissolving contiguous pixels, that meet the urban nightlight threshold of 35, to city specific polygons. I apply this procedure to the nightlight data for the years 1998-2013. Thereby, I can track the urban footprints of every Indian city annually over time. In order to verify that the nightlight-based footprints refer to actual cities, I spatially merge the retrieved footprints for the year 2001 with a shapefile of all Indian municipalities containing population information from the Population Census 2001 (Meiyappan et al., 2018). I restrict the sample to urban footprints that have a population total of at least 50.000 as of 2001.

Figure 2 exemplary shows the obtained urban footprints of Mumbai, Delhi, Kolkata and Coimbatore for the years 2001 and 2013. In the background, the raw nightlight data for the year 2001 is plotted, where pixels outside of India are masked. The yellow polygons correspond to the urban footprints of each city as of 2001. With a total area of 2,632km^2, Delhi is the largest of the four cities followed by Mumbai (1,242km^2), Kolkata (1,224km^2) and Coimbatore (267km^2). The red line demarcates the urban footprints of the same four cities in 2013. It is evident that all four cities grew in size. Delhi grew the most by more than doubling its size (increase by 103%). Yet, the growth is not only driven by the growth of the 2001 footprint, but also by the fact that smaller cities, that were not connected by contiguous and sufficiently bright pixels (≥ 35) to the footprint of Delhi in 2001, were

absorbed by growing together with the footprint of Delhi. Between 2001 and 2013 Delhi absorbed a total of six cities (Muradnagar, Modinagar, Ganaur, Samalkha and Panipat), which were still separate cities in 2001. In case of two separate cities (as of 2001) growing together, I assign the ID of the larger city, assuming that the smaller city was absorbed by the larger one. Figure A1 in Appendix A1 plots the location of the full sample of 443 cities for the years 1998, 2001 and 2013.

Figure 2: Urban footprints



Notes: Figure 2 plots the urban footprints of Mumbai, Delhi, Kolkata and Coimabtore for the years 2001 (yellow polygon) and 2013 (red contour). Urban footprints are derived by dissolving contiguous nightlight pixels with a value ≥ 35. In the background the raw Version 4 DMSP-OLS Nighttime Lights Time Series (NOAA, 1992-) for the year 2001 is plotted, where pixels outside of India are masked.

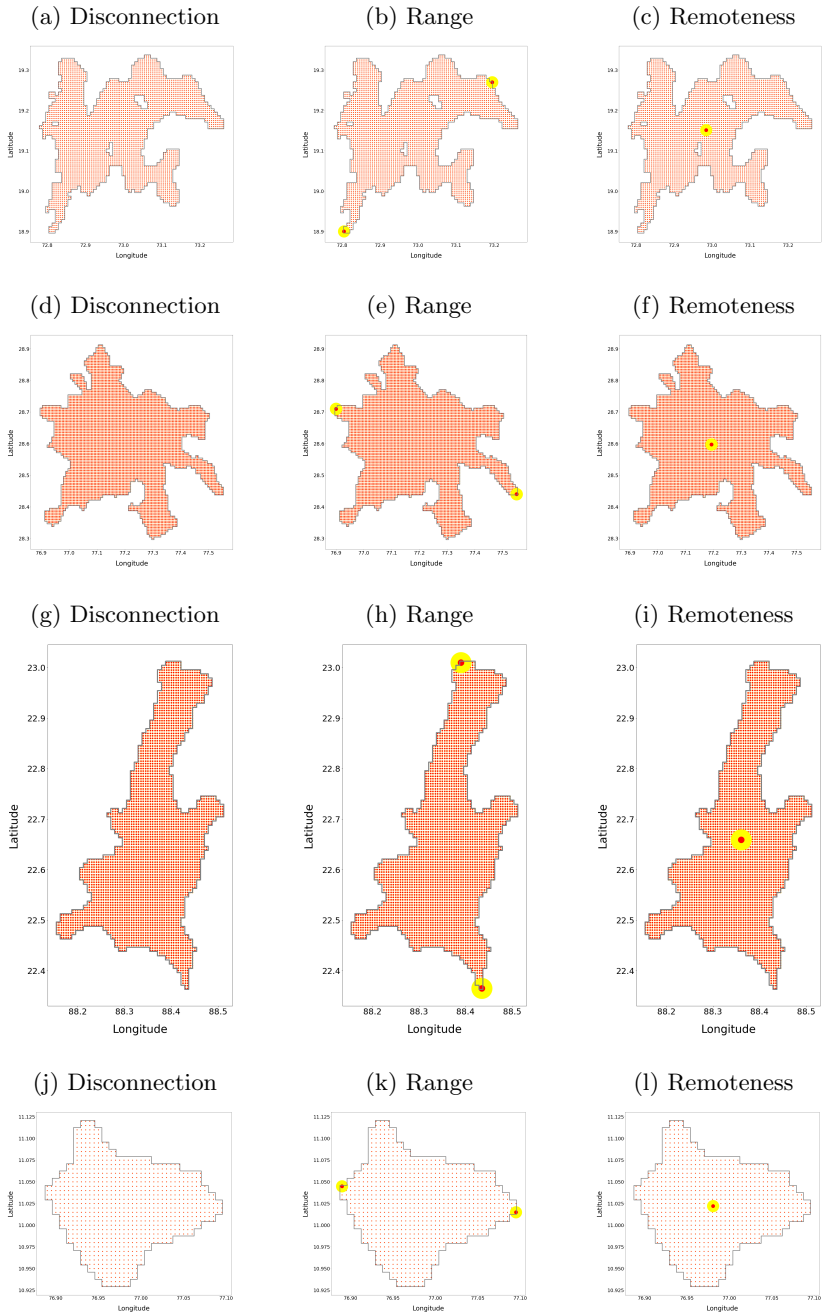
Shape properties. After having obtained the urban footprints of all Indian cities for the years 1998-2013, I derive the shape properties of each city-year observation. Following (Harari, 2020), I borrow three compactness indices from the urban planning literature

(Angel et al., 2010). The disconnection index, which is my preferred index and used in all base specifications, measures the average Euclidean distance between all points within a city. Hence, the disconnection index can be interpreted as a proxy for the average commuting distance within a city. The range index is defined as the maximum Euclidean distance within a city illustrating the longest possible commuting distance. The last index I employ is called remoteness index and calculated by the average Euclidean distance to the geographic centroid of a city. In contrast to Angel et al. (2010), I implement the construction of the compactness indices independently in an automatized workflow in python. This allows for a higher degree of flexibility in choosing uniform spacing of the underlying grid of interior city points and does not rely on the resampling of random grid points for calculating the indices.

Figure 3 illustrates the construction of the three compactness indices for the 2001 urban footprints of Mumbai, Delhi, Kolkata and Coimbatore. The red dots represent the grid of equally spaced points. I set the distance between the points to 0.005° (latitude/longitude) corresponding to approximately 555m at the equator. Taking Mumbai as an example, the grid of Mumbai consists of a total of 3,788 points (Delhi = 7,462, Kolkata = 3,660, Coimbatore = 855), where the disconnection index is simply the average distance between all pairs of points and equals 20.9km (Panel (a)). The range index - defined by the longest possible distance between all pairs of points - is indicated by the yellow-encircled red dots in Panel (b) and amounts to 59.8km for Mumbai. In Panel (c), the remoteness index is defined by the average distance of all 3,788 points to the geographic centroid (yellow-encircled red dot) and corresponds to 17.6km. In order to obtain time-varying shape properties of each city, I compute the city-level compactness indices for every urban footprint for the years between 1998-2013.

Air pollution data. I further complement the data on urban footprints with annual remote sensing data on fine particulate matter ($PM_{2.5}$), that is a collective term for inhalable floating particles with a diameter of 2.5 micrometers or less. A permanent exposure to high levels of $PM_{2.5}$ is frequently associated with adverse health effects and found to be a cause for premature deaths (Chowdhury and Dey, 2016; Ghude et al., 2016; Vohra et al., 2021). The data is obtained from the Global Annual $PM_{2.5}$ Grids from MODIS, MISR and SeaWiFS Aerosol Optical Depth (AOD) with GWR, Version 1 and consists of annual concentrations (measured in micrograms per cubic meter) of ground-level $PM_{2.5}$ (Van Donkelaar et al., 2016). The resolution of the data is 0.01° (latitude/longitude) and annually available for the years 1998-2016. I spatially merge the $PM_{2.5}$ raster data with the urban footprints and calculate the minimum, maximum and average ground-level $PM_{2.5}$ concentration for every city for the years 1998-2013.

Figure 3: Compactness indices



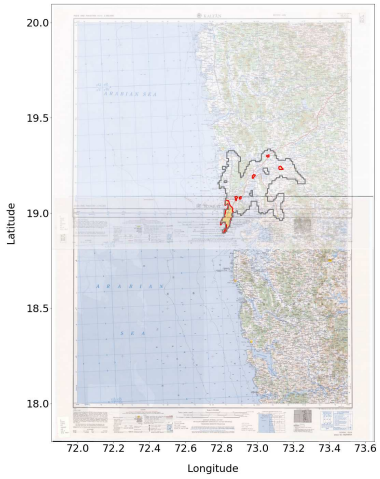
Notes: Figure 3 plots the urban footprints of Mumbai (a-c), Delhi (d-f), Kolkata (g-i) and Coimabtore (j-l) for the year 2001 together with the respective grids for calculating the disconnection, range and remoteness index. The spacing of the interior city grid is set to 0.005°, which corresponds to approx. 555m at the equator. The disconnection index is defined by the average distance between all pairs of points. The range index is defined as the maximum distance between any pair of points illustrated by the yellow-encircled red dots in panel (b), (e), (h) and (k). The remoteness index is defined as the average distance to the geographical centroid of a city. The geographical centroid is illustrated by the yellow-encircled red dot in panel (c), (f), (i) and (l).

Instrument construction. As highlighted in the previous section, I instrument the actual shape of a city with its potential shape. Following Harari (2020), I construct time varying potential urban footprints by combining historic urban footprints obtained from

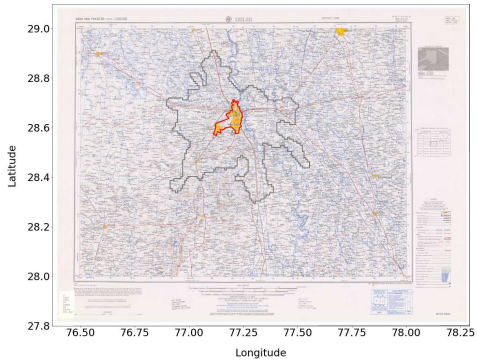
georeferenced maps (U.S. Army Map Service, 1955) with historic population information (Mitra, 1980) and detailed information on geographical obstacles for urban sprawl such as mountain ranges or water bodies (METI and NASA, 2019a,b). Note, that the identification of the effects of city shape on air pollution relies on within city variation in shape over time. The main idea of the instrument is to isolate the changes in city shape over time, that are driven by historically predicted urban sprawl constrained by geographical obstacles (Harari, 2020).

Figure 4: Historic urban footprints

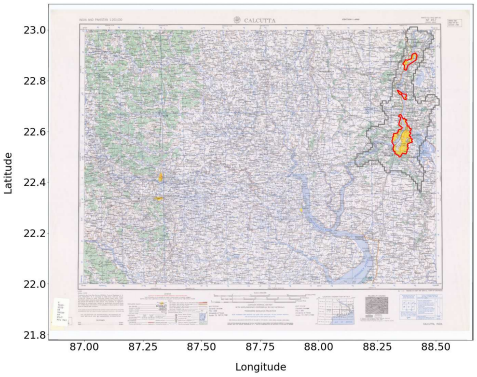
(a) Mumbai



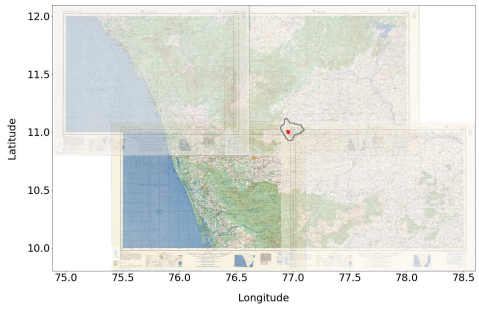
(b) Delhi



(c) Kolkata



(d) Coimbatore



Notes: Figure 4 illustrates the historic urban footprints of the cities of Mumbai, Delhi, Kolkata and Coimbatore. The yellow polygons correspond to the urban footprints in 1950 and the gray area indicates the nightlight-based urban footprints as of 2001. In the background corresponding historic maps of India, used to retrieve the historic urban footprints, are shown (U.S. Army Map Service, 1955).

In order to mechanically predict urban sprawl based on historic data, I georeference historic maps of India provided by the U.S. Army Map service, indicating the urban

footprints of Indian cities in 1950 (U.S. Army Map Service, 1955). I manually retrieve the historic urban footprints, that fall within the limits of the nightlight-based urban footprints as of 2001. Figure 4 illustrates the historic urban footprints for the cities of Mumbai, Delhi, Kolkata and Coimbatore. While Delhi and Coimbatore consist of one historic footprint, the contemporary urban footprints of Mumbai and Kolkata have multiple historic origins. Taking Mumbai as an example, the historic footprints are Mumbai, Kurla, Ghatkopar, Thane, Kaylan and Bhiwandi. Next, I match the historic urban footprints with population information of nine waves of the population census for the years 1871-1951 (Mitra, 1980). I predict future urban sprawl of the historic urban footprints mainly following the approach by Harari (2020) with slight deviations, since I allow for multiple historic footprints and do not need to instrument for historic footprints themselves. The exact steps for predicting future urban sprawl are as follows:

(i) I extrapolate log-linearily the 1871-1951 population of the historic footprints c belonging to city i, obtaining $\widehat{pop_{ict}}$, where $t \in 1998, ..., 2013$.

(ii) I estimate the following model:

$$log(area_{ict}) = \beta_1 log(\widehat{pop_{ict}}) + \beta_2 log(\frac{pop_{ic,t=1951}}{area_{ic,t=1950}}) + \alpha_i + \gamma_t + \epsilon_{ict}, \tag{4}$$

where $area_{ict}$ is the nightlight-based area of of the historic footprint c belonging to city i in year t. Given that some cities consist of multiple historic footprints, I assign the contemporary area of historic footprints according to their share of the total historic area of a city. $\frac{pop_{ic,1951}}{area_{ic,1950}}$ is the population density of the historic footprint, α_i and γ_t city and year fixed effects.

(iii) From the obtained coefficients in Eq. 4, I predict the area of the historic urban footprint c belonging to city i in year t ($\widehat{area_{ict}}$).

(iv) I compute the predicted radius $\widehat{r_{ict}}$, where

$$\widehat{r_{ict}} = \sqrt{\frac{\widehat{area_{ict}}}{\pi}} \tag{5}$$

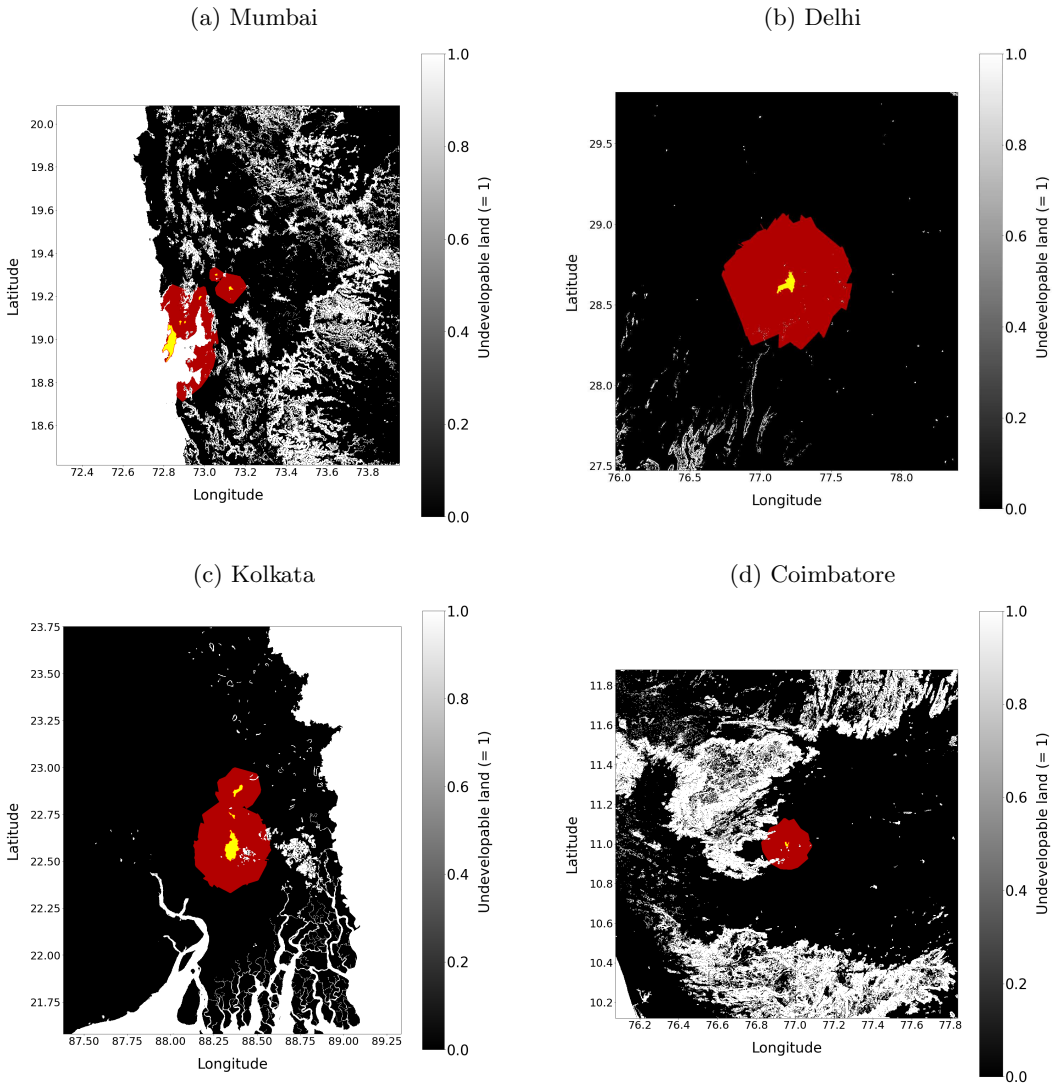
The predicted radius $\widehat{r_{ict}}$ is used to compute the buffered area of the historic footprint c in year t if it would have grown at the same rate as in 1871-1951, keeping population density constant and geographically expanding at same rates in all directions (Harari, 2020).

The second source of variation in the instrument is derived from classifying the land surface of India into potentially developable and undevelopable land based on geographical obstacles (Harari, 2020). I obtain the ASTER Global Digital Elevation Model V003 for India (METI and NASA, 2019a). For computational reasons, I aggregate the resolution of the raster data ending up with a resolution of approximately 120m at the equator.[2]

[2]The original resolution is 1 arc second, which equals approximately 30m at the equator.

I compute the slope at each pixel and define all pixels with a slope steeper than 15% as undevelopable (Saiz, 2010). I combine the slope data with the ASTER Global Water Bodies Database V001 (METI and NASA, 2019b), where I define pixels containing water bodies such as lakes as undevelopable. Figure A2 in Appendix A1 shows the classification for all India, where pixels with a slope greater than 15% or containing a water body are classified as undevelopable.

Figure 5: Potential urban footprints



Notes: Figure 5 illustrates the derived potential urban footprints of Mumbai, Delhi, Kolkata and Coimbatore for the year 2013, indicated by the red area. The historic footprints of the cities are plotted in yellow. The background shows the raster classification into developable (=0) and undevelopable (=1) land, derived from METI and NASA (2019a,b). Pixels outside of India are masked.

Lastly, I combine the buffered area of the historic footprint c in year t with the raster of (un)developable land in India by masking undevelopable parts of the predicted footprints. This yields the potential urban footprints of every city in my sample for the years 1998-

2013. Figure 5 illustrates the derived potential footprints for the cities of Mumbai, Delhi, Kolkata and Coimbatore for the year 2013, indicated by the red area. White pixels in the background represent undevelopable land pixels. As shown in panel (a), the potential footprint of Mumbai is not only constrained by the Arabian Sea in the west, but also by various smaller mountain ranges in the inland. In contrast, Delhi and Kolkata do not face any major geographical constraints apart from various smaller water bodies. The potential footptrint of Coimbatore is mainly constraint by the foothills of the Western Ghats, when expanding westwards. Analogously to the actual urban footprints, I compute the disconnections, range and remoteness index for the potential urban footprints. This yields the instrument $\widetilde{shape}_{it}$ used in the first stage (Eq. 2). Figure A3 in Apendix A1 illustrates the computation of the three compactness indices for the potential urban footprints.

Table 1: Descriptive statistics

	(1) *Mean*	(2) *SD*	(3) *Min*	(4) *Median*	(5) *Max*	(6) *N*
Air Pollution ($PM_{2.5}$)						
- City mean ($\frac{\mu g}{m^3}$)	35.46	18.47	8.291	29.62	123.3	6,414
- City min ($\frac{\mu g}{m^3}$)	34.78	18.18	7.100	29.20	123.2	6,414
- City max ($\frac{\mu g}{m^3}$)	36.11	18.71	8.400	30.10	123.5	6,414
City shape properties						
- Area (km^2)	77.26	229.0	0.873	25.99	5,336	6,414
- Disconnection (km)	3.466	3.297	0.278	2.519	47.91	6,414
- Potential Disconnection (km)	3.793	3.111	0.494	2.888	41.93	6,414
- Range (km)	9.101	10.10	0.556	6.016	159.4	6,414
- Potential Range (km)	9.138	8.146	1.112	6.766	102.3	6,414
- Remoteness (km)	2.619	2.488	0.278	1.903	44.48	6,414
- Potential Remoteness (km)	2.801	2.281	0.656	2.134	30.84	6,414

Notes: The final sample comprises 443 Indian cities with at least 50.000 inhabitants in 2001. $PM_{2.5}$ concentration is measured in micrograms per cubic meter ($\frac{\mu g}{m^3}$). City shape properties are measured in km and sq. km respectively. Potential shape indices are based the potential footprints of a city. Unit of observation is city-year. The information presented in Table 1 is based on city-year observations for the years 1998-2013.

Descriptive statistics. The final sample comprises 443 cities covering the time period 1998-2013. Recall that the panel is unbalanced since the footprints of smaller cities might be absorbed by the urban sprawl of larger neighbouring cities in the course of time. Table 1 provides details on the descriptive statistics of the final sample. The average annual $PM_{2.5}$ concentration is $35.46 \frac{\mu g}{m^3}$ surpassing the recommended guidelines of the WHO by

a factor of more than three.[3] The city with the best air quality is Nagercoil, which is a coastal town in Tamil Nadu, having an average $PM_{2.5}$ concentration of $12.6\frac{\mu g}{m^3}$. With an average $PM_{2.5}$ concentration of $97.3\frac{\mu g}{m^3}$, Muradnagar a satellite town of Delhi, is the city with the worst air quality. Turning to the shape properties of the urban footprints, the average area of the urban footprints is $77.3km^2$ with an average commuting distance (disconnection) within cities of $3.5km$. The average maximum commuting distance (range) amounts to $9.1km$ and the average commuting distance to the geographic center $2.6km$ respectively.

4 Results

4.1 Baseline results

Table 2 provides the estimation results of the baseline specifications including both the pooled OLS and the two-stage least squares of the instrumental variable approach. In columns (1) - (3), the disconnection index is used as explanatory variable for measuring city shape. Recall that a greater value indicates less compactness. When estimating the pooled OLS in column (3), I find a negative and weakly significant effect of city compactness on $PM_{2.5}$ concentration. A $1km$ increase in the disconnection index increases average $PM_{2.5}$ concentration by 0.41%. The IV estimation in column (2) provides in absolute terms a slightly larger effect indicating a potential upward bias in the OLS coefficient. A one standard deviation (sd) increase in in the disconnection index increases average $PM_{2.5}$ concentration by 1.91% (3.297 $\times$ (-0.58%)). The negative and significant effect of city compactness is robust to using the range index as well as the remoteness index as alternative compactness measures. Throughout all specifications, there is a positive association between total area of a city and $PM_{2.5}$ concentration. Figure A1 and A2 in Appendix A2 report the results when using maximum and minimum $PM_{2.5}$ concentrations instead of average $PM_{2.5}$ concentration as the dependent variable.

[3]The WHO guidelines recommended an average annual $PM_{2.5}$ concentration of $10\frac{\mu g}{m^3}$ at the time of the study period, which were further reduced to $5\frac{\mu g}{m^3}$ in 2021.

Table 2: Baseline results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Shape index:		Disconnection			Range			Remoteness	
		IV	OLS		IV	OLS		IV	OLS
	First stage	Second stage		First stage	Second stage		First stage	Second stage	
Dependent variable:	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)
Shape (km)		-0.0058**	-0.0041*		-0.0018**	-0.0011*		-0.0076**	-0.0061**
		(0.0026)	(0.0024)		(0.0007)	(0.0006)		(0.0031)	(0.0028)
log(area (sq.km))	0.9993***	0.0071*	0.0059*	2.8995***	0.0063*	0.0049	0.7536***	0.0070*	0.0061*
	(0.1087)	(0.0038)	(0.0035)	(0.3481)	(0.0036)	(0.0034)	(0.0840)	(0.0037)	(0.0034)
Potential shape (km)	1.0258***			1.4143***			1.1035***		
	(0.0936)			(0.1184)			(0.1062)		
Observations	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414
F-test stat	11.55			10.52			8.66		
City fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. 1, 2 and 3 with log of $PM_{2.5}$ concentration as the dependent variable measured in micrograms per cubic meter ($\frac{\mu g}{m^3}$). Column (1), (4) and (7) report the results for the first stage. Column (2), (5) and (8) report results for the second stage. Column (3), (6) and (9) report results for the pooled OLS. Standard errors are clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1.

In order to study heterogeneous effects of city shape on air pollution, I split the sample by city size. Table 3 reports the baseline results separately for cities with a population below and above 500.000 (as of 2001). As denoted in column (2) of panel (A), the effect of city shape on $PM_{2.5}$ concentration grows with the size of cities. Further, note that the IV estimate is twice as large as the OLS estimate in column (3). A $1km$ increase in the disconnection index reduces $PM_{2.5}$ concentration by 1.1% (6.2% for a $1sd$ increase of $5.5km$). Again, the effects are robust to using alternative compactness indices. However, the estimates of the range and the remoteness index should be interpreted with caution, given their low F-statistic indicating a weak instrument. Turning to the sample of small cities in panel (B), I do not find any significant effect of city shape on $PM_{2.5}$ concentration. Hence, the significant relationship between city shape and $PM_{2.5}$ concentration of the baseline results in Table 2 seems to be primarily driven by larger cities.

Table 3: Results by city size

Shape index:	(1)	(2) Disconnection	(3)	(4)	(5) Range	(6)	(7)	(8) Remoteness	(9)
		IV	OLS		IV	OLS		IV	OLS
	First stage	Second stage		First stage	Second stage		First stage	Second stage	
Dependent variable:	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)
Panel (A): Large cities ≥ 500.000									
Shape (km)		-0.0112***	-0.0064**		-0.0034***	-0.0014		-0.0134***	-0.0082***
		(0.0031)	(0.0031)		(0.0009)	(0.0009)		(0.0038)	(0.0026)
log(area (sq.km))	2.6677***	0.0053	-0.0051	8.0410***	0.0022	-0.0103	2.1319***	0.0040	-0.0049
	(0.6933)	(0.0231)	(0.0216)	(2.0254)	(0.0225)	(0.0203)	(0.5734)	(0.0231)	(0.0213)
Potential shape (km)	1.0604***			1.5025***			1.2052***		
	(0.0902)			(0.0921)			(0.1072)		
Observations	947	947	947	947	947	947	947	947	947
F-test stat	11.55			3.13			2.65		
Panel (B): Small cities < 500.000									
Shape (km)		-0.0042	-0.0026		-0.0013	-0.0008		-0.0061	-0.0043
		(0.0047)	(0.0035)		(0.0015)	(0.0010)		(0.0062)	(0.0048)
log(area (sq.km))	0.9023***	0.0076	0.0063*	2.5845***	0.0070	0.0061*	0.6696***	0.0078*	0.0068*
	(0.1222)	(0.0046)	(0.0038)	(0.3942)	(0.0044)	(0.0036)	(0.0939)	(0.0045)	(0.0038)
Potential shape (km)	1.0020***			1.3389***			1.0715***		
	(0.2622)			(0.3303)			(0.2710)		
Observations	5,467	5,467	5,467	5,467	5,467	5,467	5,467	5,467	5,467
F-test stat	11.55			6.97			6.81		
City fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. 1, 2 and 3 with log of $PM_{2.5}$ concentration as the dependent variable measured in micrograms per cubic meter ($\frac{\mu g}{m3}$). Panel (A) reports the results for cities with at least 500.000 inhabitants and panel (B) for cities having below 500.000 inhabitant. Column (1), (4) and (7) report the results for the first stage. Column (2), (5) and (8) report results for the second stage. Column (3), (6) and (9) report results for the pooled OLS. Standard errors are clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Commuting and city shape

Given that road traffic is a substantial contributor to air pollution (Badami, 2005; Pant and Harrison, 2013; Vohra et al., 2021), local air quality can be regarded as a function of total road traffic within a city. Hence, a potential channel for the negative impact of compactness could be that less compact cities have fewer overall road traffic. While data on actual commuting patterns at the city-level is unfortunately not available, the Population Census 2011 provides some coarse information on commuting patterns at the district level differentiating between urban and rural population. The census contains aggregate information on how many people use a specific mode of transport to work as well as a binned classification count of actual commuting distances (e.g. 0-1km, 2-5km, 6-10km,...). I spatially match the urban footprints of 2011 with their respective hosting district and evaluate if city compactness has any effect on actual commuting patterns including modal choices. I assign a city to all districts, that intersect with the urban footprint of the respective city. Hence, a city can be assigned to multiple districts simultaneously. Further, I restrict the district information to the urban population only. Note, that while the main specification relies on within city variation in compactness over time, the following analysis uses cross-sectional variation in city shape within states for identification. The detailed estimation equation is given in Eq. A1 in Appendix A2.

Table 4 provides results on the impact of city compactness (measured by the disconnection index) on actual commuting distance and the initial decision to commute. The

dependent variables refer to the share of people commuting more than $10km$ and the share of people that commute to work at all (commuting distance $> 0km$). As indicated by column (1), a $1km$ increase in the disconnection index increases the share of people commuting above $10km$ by $0.69pp$ (percentage points). Turning to the share of people commuting independent of the distance, I do not find any significant effect of city compactness on the initial decision to commute (column (3)). In column (5) - (12), I study heterogeneous mobility effects by gender. Yet, I do not find any substantial difference in commuting responses by males and females with respect to city compactness.

Table 4: Commuting distance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Total				Male				Female			
	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS
Dependent variable:	> 10km		Commuting share		> 10km		Commuting share		> 10km		Commuting share	
Shape (km)	0.0069***	0.0064***	0.0002	0.0002	0.0070***	0.0065***	0.0003	0.0003	0.0063***	0.0059***	-0.0004	-0.0008
	(0.0009)	(0.0011)	(0.0017)	(0.0016)	(0.0009)	(0.0011)	(0.0016)	(0.0015)	(0.0013)	(0.0014)	(0.0026)	(0.0022)
log(area (sq.km))	-0.0054**	-0.0041*	0.0063	0.0064	-0.0050**	-0.0037*	0.0062	0.0062	-0.0078**	-0.0067**	0.0098	0.0108
	(0.0023)	(0.0021)	(0.0055)	(0.0051)	(0.0022)	(0.0021)	(0.0051)	(0.0047)	(0.0033)	(0.0030)	(0.0079)	(0.0069)
Observations	515	515	515	515	515	515	515	515	515	515	515	515
F-test stat	6.38		6.38		6.38		6.38		6.38		6.38	
State fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. A1. The dependant variables are the share of people commuting more than 10km and 0km respectively, where the shares range from 0 to 1. Column (1) -(4) report the results for all commutes. Column (5) -(12) report the results separate for male and female commuters. In all specifications, the disconnection index is used for measuring city shape. Standard errors are clustered at the state level. Unit of observation is city-district for the year 2011, where cities are assigned to every district that intersects with the respective urban footprint of the city. Commuting information is obtained from the Population census 2011. *** p<0.01, ** p<0.05, * p<0.1.

While it seems that a less compact city shape leads to longer commutes, it may also lead to different modal choices. Table 5 reports estimates on the effect of city shape on the modal choice for commuting. The dependent variable is calculated as the share of people using a specific mode of transport for commuting to work. I do not find any significant effect of city shape on the share of people commuting by foot (column (1)). However, there is a significant negative impact of city shape on the share of people using bicycle or scooters for commuting. For instance, a $1km$ increase in the disconnection index decreases the share of people commuting with a scooter by $0.73pp$ (column (5)). Contrary, I find significant and positive effects for commutes by car/van, bus and train. Hence, in less compact cities, the likelihood to commute with multi person vehicles increases while it decreases for individual transport such as scooters. Note, that the estimates on the actual commuting patterns are only suggestive and do not provide information on the overall road traffic within a city. Yet, differences in modal choices could be a potential mechanism adding to the explanation of the negative impact of compactness on air quality. Beside the fact that dispersed city shapes allow for more fresh air corridors (Borrego et al., 2006), the increased use of public transport in less compact cities might reduce overall road traffic leading to less air pollution.

Table 5: Transport mode

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS
Dependent variable:	Foot		Bicycle		Scooter		Rickshaw		Car/Van		Bus		Train	
Shape (km)	0.0029	0.0046***	-0.0078***	-0.0087***	-0.0073***	-0.0075***	-0.0004	-0.0006	0.0029**	0.0024**	0.0041**	0.0039**	0.0057**	0.0058**
	(0.0019)	(0.0015)	(0.0016)	(0.0017)	(0.0012)	(0.0015)	(0.0005)	(0.0005)	(0.0011)	(0.0011)	(0.0017)	(0.0014)	(0.0024)	(0.0025)
log(area (sq.km))	-0.0272***	-0.0316***	0.0154**	0.0177***	0.0239***	0.0244***	0.0042**	0.0047***	-0.0031	-0.0021	-0.0048	-0.0043	-0.0081*	-0.0086**
	(0.0071)	(0.0064)	(0.0059)	(0.0057)	(0.0038)	(0.0040)	(0.0017)	(0.0015)	(0.0024)	(0.0024)	(0.0050)	(0.0043)	(0.0039)	(0.0035)
Observations	515	515	515	515	515	515	515	515	515	515	515	515	515	515
F-test stat	6.38		6.38		6.38		6.38		6.38		6.38		6.38	
State fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. A1. The dependant variables are the share of people commuting with a specific mode of transport, where the shares range from 0 to 1. In all specifications, the disconnection index is used for measuring city shape. Standard errors are clustered at the state level. Unit of observation is city-district for the year 2011, where cities are assigned to every district that intersects with the respective urban footprint of the city. Commuting information is obtained from the Population Census 2011. *** p<0.01, ** p<0.05, * p<0.1.

4.3 Health implications

The results in the previous section reveal a significant and negative effect of the compactness of a city on local air quality measured by average $PM_{2.5}$ concentration. However, these effects are only of specific policy relevance if they translate into significant health effects. A permanent exposure of high levels of $PM_{2.5}$ concentration increases the risk of adverse health effects including premature death (Pope et al., 2004; Brauer et al., 2012; Vohra et al., 2021). In order to asses the health implications of the obtained estimates, I estimate the relative risk for the entire range in $PM_{2.5}$ concentration of my sample for four different diseases that are frequently associated with a permanent $PM_{2.5}$ exposure. The four diseases are chronic obstructive pulmonary disease (COPD), ischemic heart disease (IHD), stroke and lung cancer. I apply the relative risk functions of Chowdhury and Dey (2016), who provide empirically adjusted parameters for India.
The relative risk function is written as follows:

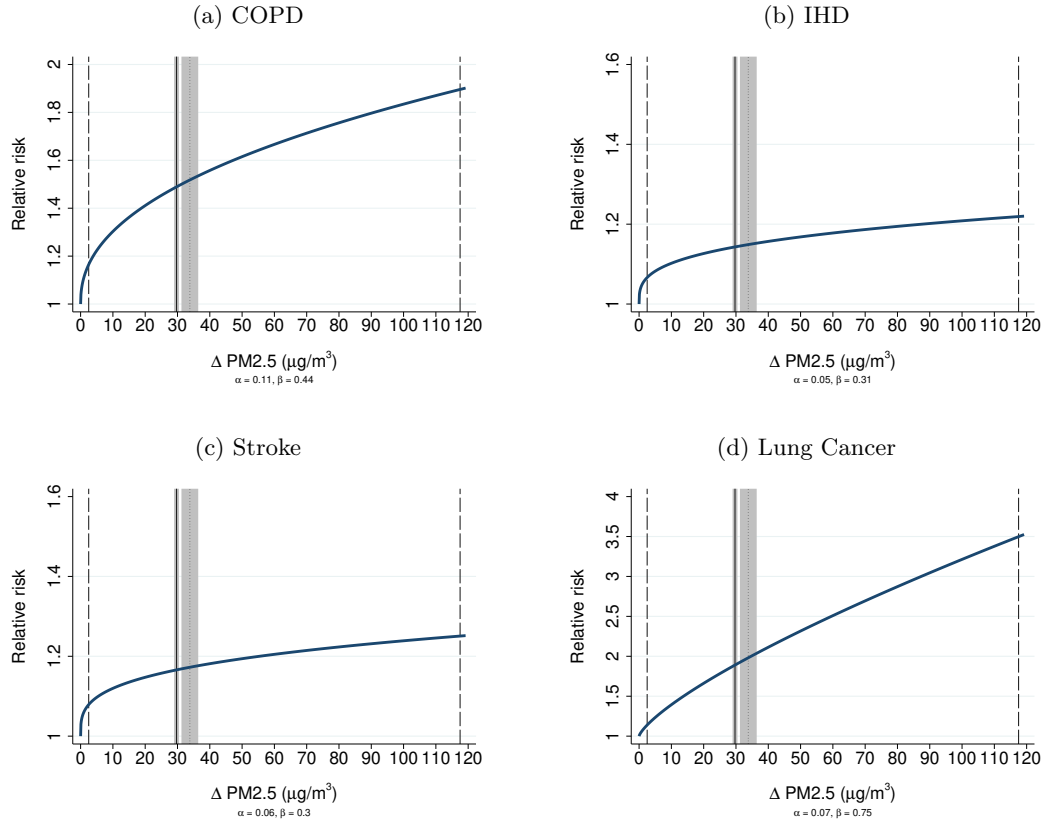
$$RR_i = 1 + \alpha_i(\Delta PM_{2.5})^{\beta_i}, \tag{6}$$

where RR refers to the relative risk of disease i. α_i and β_i are disease specific parameters and $\Delta PM_{2.5}$ is the deviation in $PM_{2.5}$ concentration from the counterfactual distribution of $5.8\frac{\mu g}{m^3}$. Hence, the relative risk indicates the probability of falling sick of disease i relative to the counterfactual distribution, which is kept at non-hazardous levels of $PM_{2.5}$ concentration (Chowdhury and Dey, 2016; WHO, 2021). Figure 6 reports the relative risk of the four diseases for the entire sample range of $PM_{2.5}$ concentration. In all panels, the left dashed line indicates the sample minimum in $\Delta PM_{2.5}$ of $2.5\frac{\mu g}{m^3}$ ($8.3\frac{\mu g}{m^3} - 5.8\frac{\mu g}{m^3}$). The dashed line on the right reports the sample maximum ($117.5\frac{\mu g}{m^3}$). The solid line indicates the unconditional sample mean ($29.7\frac{\mu g}{m^3}$) for all cities and the dotted line for large cities ($33.8\frac{\mu g}{m^3}$). Lastly, the grey-shaded areas indicate the change in $PM_{2.5}$ concentration induced by a one standard deviation change in city compactness (disconnection index) evaluated at the respective mean.

Using panel (d) as an example for illustration, the relative risk of lung cancer at the sample mean equals 1.89. This means that if $PM_{2.5}$ concentration increases from $5.8\frac{\mu g}{m^3}$ to $35.5\frac{\mu g}{m^3}$, the probability of lung cancer increases by 89%. A one standard deviation increase

in the disconnection index would decrease $PM_{2.5}$ concentration by 1.91% translating into an absolute reduction of $0.68\frac{\mu g}{m^3}$ at the sample mean of $35.5\frac{\mu g}{m^3}$. Hence, a one standard deviation increase in the disconnection index would decrease the relative risk of lung cancer by one percentage point from 1.89 to 1.88 (0.5% in relative terms). Evaluated at the sample mean of large cities, a one standard deviation increase in the disconnection index would decrease the relative risk of lung cancer by three percentage points (from 1.98 to 1.95). Due to the overall high levels of $PM_{2.5}$ concentration in Indian cities as well as the concavity of the risk functions, the grey-shaded areas propose only minimal changes in relative risks for all four diseases.

Figure 6: Relative risk by disease



Notes: Figure 6 plots the relative risk functions of chronic obstructive pulmonary disease (COPD), ischemic heart disease (IHD), stroke and lung cancer with respect to deviations in $PM_{2.5}$ concentration from the counterfactual distribution of $5.8\frac{\mu g}{m^3}$. Relative risk functions are based on Chowdhury and Dey (2016). Dashed lines indicate the sample minimum and maximum. The solid line indicates the unconditional sample mean for all cities and the dotted line for large cities. Grey-shaded areas illustrate changes in $PM_{2.5}$ concentration induced by a one standard deviation change in within city commuting distance (disconnection index).

5 Conclusion

Air pollution poses a significant health risk for the local population in many cities around the world. This paper studies the impact of city shape on air pollution and thereby extends

the framework of previous work by Harari (2020). By focusing on the compactness of urban footprints, it contributes to the literature by incorporating alternative agglomeration properties than density (Borck and Schrauth, 2021; Carozzi and Roth, 2023). I find a negative and statistically significant effect of city compactness on air quality. A $1km$ increase in the disconnection index decreases average $PM_{2.5}$ concentration by 0.58%. The effect is further increasing in city size. Albeit only suggestive in nature, commuting could serve as a partial mechanism in explaining these differences. Due to longer commutes in less compact cities, residents rely on public transport leading to fewer overall road traffic and congestion in cities resulting in less air pollution. However, I find that these effects barely translate into substantial changes in the relative risks assessment of diseases that are frequently associated with a permanent exposure to high levels of $PM_{2.5}$. This is due to the concavity of the relative risk functions as well as the already high levels of $PM_{2.5}$ concentration in Indian cities.

While the presented work primarily uses remote sensing data allowing the annual coverage of a wide range of Indian cities and thereby formulating general patterns, future work could focus on specific case studies to dive deeper into the mechanisms between agglomeration and air pollution. Further robustness checks could include the use of simulated commuting data from google maps (Akbar et al., 2018) or stationary air pollution data (Greenstone and Hanna, 2014).

References

Aguilar-Gomez, Sandra, Holt Dwyer, Joshua Graff Zivin, and Matthew Neidell, "This is air: The 'nonhealth' effects of air pollution," *Annual Review of Resource Economics*, 2022, *14*, 403–425.

Ahfeldt, Gabriel M and Elisabetta Pietrostefani, "The compact city in empirical research: A quantitative literature review," 2017.

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

Akbar, Prottoy A, Victor Couture, Gilles Duranton, and Adam Storeygard, "Mobility and congestion in urban India," Technical Report, National Bureau of Economic Research 2018.

Angel, Shlomo, Jason Parent, and Daniel L Civco, "Ten compactness properties of circles: measuring shape in geography," *The Canadian Geographer/Le Géographe Canadien*, 2010, *54* (4), 441–461.

Badami, Madhav G, "Transport and urban air pollution in India," *Environmental Management*, 2005, *36*, 195–204.

Balakrishnan, Uttara and Magda Tsaneva, "Air pollution and academic performance: Evidence from India," *World Development*, 2021, *146*.

Balk, Deborah L, Uwe Deichmann, Greg Yetman, Francesca Pozzi, Simon I Hay, and Andy Nelson, "Determining global population distribution: methods, applications and data," *Advances in Parasitology*, 2006, *62*, 119–156.

Borck, Rainald and Philipp Schrauth, "Population density and urban air quality," *Regional Science and Urban Economics*, 2021, *86*.

_ and _ , "Urban pollution: A global perspective," 2022.

Borrego, Carlos, Helena Martins, Oxana Tchepel, L Salmim, Alexandra Monteiro, and Ana Isabel Miranda, "How urban structure can affect city sustainability from an air quality perspective," *Environmental Modelling & Software*, 2006, *21* (4), 461–467.

Brauer, Michael, Markus Amann, Rick T Burnett, Aaron Cohen, Frank Dentener, Majid Ezzati, Sarah B Henderson, Michal Krzyzanowski, Randall V Martin, Rita Van Dingenen et al., "Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution," *Environmental Science & Technology*, 2012, *46* (2), 652–660.

Carozzi, Felipe and Sefi Roth, "Dirty density: Air quality and the density of American cities," *Journal of Environmental Economics and Management*, 2023, *118*.

Chowdhury, Sourangsu and Sagnik Dey, "Cause-specific premature death from ambient PM2. 5 exposure in India: Estimate adjusted for baseline mortality," *Environment International*, 2016, *91*, 283–290.

CNN, "Future or fantasy? Designs unveiled for one-building city stretching 106 miles in Saudi Arabia," 2022. URL: https://edition.cnn.com/style/article/saudi-arabia-the-line-city-scli-intl/index.html (accessed 10/03/2023).

Combes, Pierre-Philippe, Thierry Mayer, and Jacques-François Thisse, "Economic geography," in "Economic Geography," Princeton University Press, 2008.

Dingel, Jonathan I, Antonio Miscio, and Donald R Davis, "Cities, lights, and skills in developing economies," *Journal of Urban Economics*, 2019.

Donkelaar, Aaron Van, Randall V Martin, Michael Brauer, N Christina Hsu, Ralph A Kahn, Robert C Levy, Alexei Lyapustin, Andrew M Sayer, and David M Winker, "Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors," *Environmental Science & Technology*, 2016, *50* (7), 3762–3772.

Duranton, Gilles and Diego Puga, "The economics of urban density," *Journal of Economic Perspectives*, 2020, *34* (3), 3–26.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth, "The long-run economic consequences of high-stakes examinations: Evidence from transitory variation in pollution," *American Economic Journal: Applied Economics*, 2016, *8* (4), 36–65.

Fowlie, Meredith, Edward Rubin, and Reed Walker, "Bringing satellite-based air quality estimates down to earth," in "AEA Papers and Proceedings," Vol. 109 2019, pp. 283–88.

Fujita, Masahisa and Jacques-François Thisse, "Economics of agglomeration," *Journal of the Japanese and International Economies*, 1996, *10* (4), 339–378.

Ghude, Sachin D, DM Chate, C Jena, G Beig, R Kumar, MC Barth, GG Pfister, S Fadnavis, and Prakash Pithani, "Premature mortality in India due to PM2. 5 and ozone exposure," *Geophysical Research Letters*, 2016, *43* (9), 4650–4658.

Greenstone, Michael and Rema Hanna, "Environmental regulations, air and water pollution, and infant mortality in India," *American Economic Review*, 2014, *104* (10), 3038–72.

Harari, Mariaflavia, "Cities in bad shape: Urban geometry in India," *American Economic Review*, 2020, *110* (8), 2377–2421.

Henderson, Vernon, "The urbanization process and economic growth: The so-what question," *Journal of Economic Growth*, 2003, *8*, 47–71.

IQAir, "World Air Report," 2019. URL: https://www.iqair.com/world-most-polluted-cities (accessed 10/03/2023).

Lepeule, Johanna, Francine Laden, Douglas Dockery, and Joel Schwartz, "Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard Six Cities study from 1974 to 2009," *Environmental Health Perspectives*, 2012, *120* (7), 965–970.

Meiyappan, P., P. S. Roy, A. Soliman, T. Li, P. Mondal, S. Wang, and A.K. Jain, "India Village-Level Geospatial Socio-Economic Data Set: 1991, 2001," Technical Report, NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY 2018.

METI and NASA, "ASTER Global Digital Elevation Model V003," 2019. Data retrieved from NASA Earthdata, URL: https://cmr.earthdata.nasa.gov/search/concepts/C1575731655-LPDAAC_ECS.html (accessed 10/03/2023).

_ **and** _, "ASTER Global Water Bodies Database V001," 2019. Data retrieved from NASA Earthdata, URL: https://cmr.earthdata.nasa.gov/search/concepts/C1575734433-LPDAAC_ECS.html (accessed 10/03/2023).

Mitra, Ashok, "Population and Area of Cities, Towns, and Urban Agglomerations, 1872-1971," 1980. Bombay: Allied Publishers.

NOAA, "Version 4 DMSP-OLS Nighttime Lights Time Series," 1992-. Data retrieved from National Oceanic and Atmospheric Administration, URL: https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html (accessed 10/03/2023),.

Pant, Pallavi and Roy M Harrison, "Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review," *Atmospheric Environment*, 2013, *77*, 78–97.

Piracha, Awais and Muhammad Tariq Chaudhary, "Urban air pollution, urban heat island and human health: a review of the literature," *Sustainability*, 2022, *14* (15).

Pope, C Arden, Richard T Burnett, George D Thurston, Michael J Thun, Eugenia E Calle, Daniel Krewski, and John J Godleski, "Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease," *Circulation*, 2004, *109* (1), 71–77.

Redding, Stephen J and Esteban Rossi-Hansberg, "Quantitative spatial economics," *Annual Review of Economics*, 2017, *9*, 21–58.

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

Small, Christopher, Francesca Pozzi, and Christopher D Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Remote Sensing of Environment*, 2005, *96* (3-4), 277–291.

UN, "World Urbanization Prospects 2018," 2018. Data retrieved from UN, URL: https://population.un.org/wup/ (accessed 10/03/2023).

U.S. Army Map Service, "India and Pakistan 1:250,000 Series U502," 1955. Data retrieved from University of Texas Libraries, The University of Texas at Austin, URL: https://maps.lib.utexas.edu/maps/ams/india/ (accessed 10/03/2023).

Vohra, Karn, Alina Vodonos, Joel Schwartz, Eloise A Marais, Melissa P Sulprizio, and Loretta J Mickley, "Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem," *Environmental Research*, 2021, *195*.

WHO, "WHO global air quality guidelines," Technical Report, World Health Organization 2021.

_ , "Ambient (outdoor) air pollution," 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (accessed 10/03/2023).

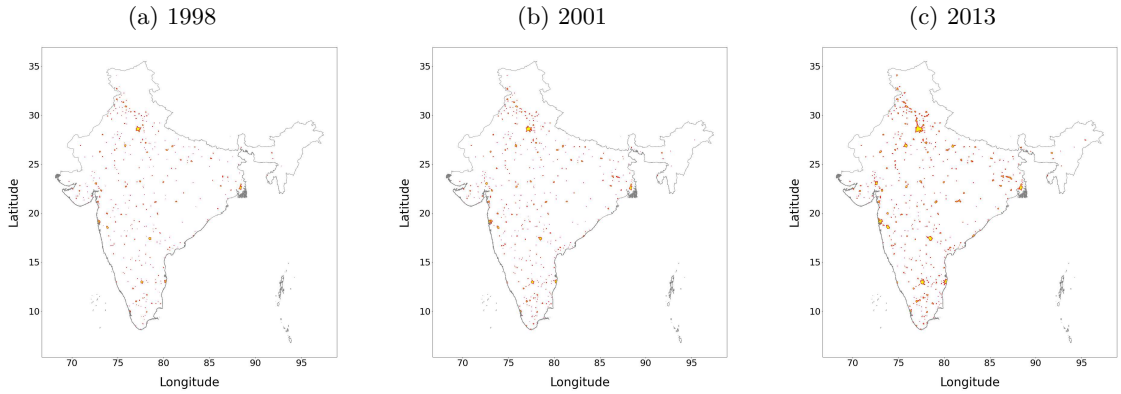
Zivin, Joshua Graff and Matthew Neidell, "The impact of pollution on worker productivity," *American Economic Review*, 2012, *102* (7), 3652–73.

A Appendix

A1 Data

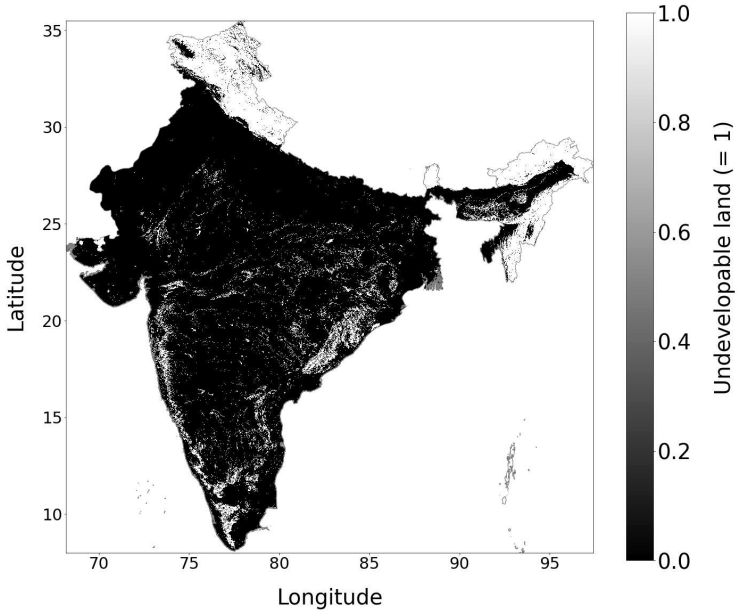
This section complements Section 3 in the main paper. Figure A1 plots the urban footprints of the final sample for the years 1998, 2001 and 2013. Figure A2 plots the the (un)developabale land in India. Figure A3 plots the construction of the three compactness indices for the potential footprints of Mumbai, Delhi, Kolkata and Coimbatore in 2001.

Figure A1: City sample



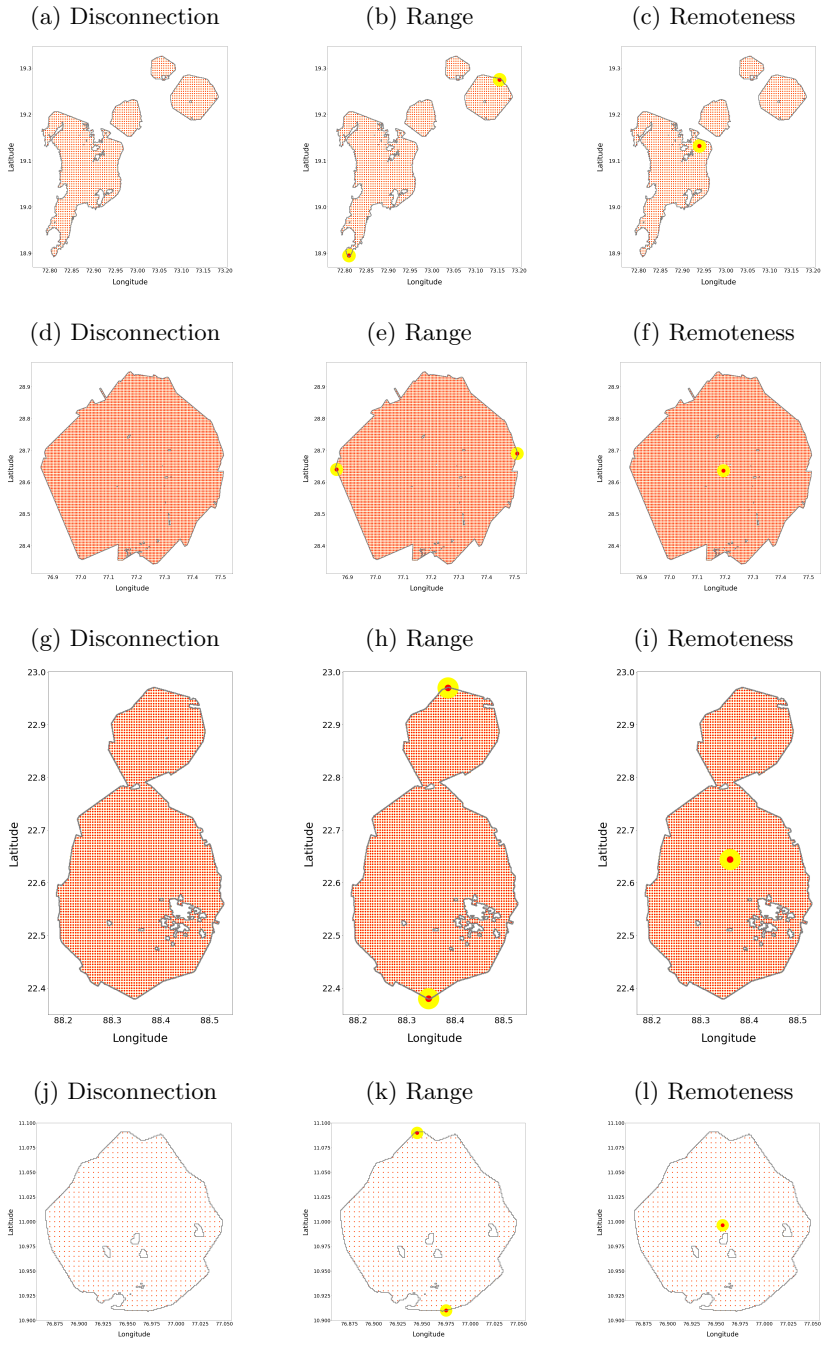
Notes: Figure A1 plots the urban footprints of the final sample for the years 1998, 2001 and 2013. Urban footprints are derived from dissolving contiguous nightlights pixels with a value ≥ 35. Nightlights are obtained from Version 4 DMSP-OLS Nighttime Lights Time Series (NOAA, 1992-).

Figure A2: Developable land in India



Notes: Figure A2 shows the raster classification into developable (=0) and undevelopable (=1) land, derived from METI and NASA (2019a,b). The resoultion has been coarsend by a factor of 4 for plotting.

Figure A3: Compactness indices (IV)



Notes: Figure A3 plots the potential urban footprint of Mumbai (a-c), Delhi (d-f), Kolkata (g-i) and Coimabtore (j-l) for the years 2001 together with the respective grids for calculating the disconnection, range and remoteness index. The spacing of the interior city grid is set to 0.005°, which corresponds to approx. 555m at the equator. The disconnection index is defined by the average distance between all pairs of points. The range index is defined as the maximum distance between any pair of points illustrated by the yellow-encircled red dots in panel (b), (e), (h) and (k). The remoteness index is defined as the average distance to the geographical centroid of a city. The geographical centroid is illustrated by the yellow-encircled red dot in panel (c), (f), (i) and (l).

A2 Results

This section complements Section 4 in the main paper. Figure A1 and A2 report the baseline results, when using the log of the maximum/minimum $PM_{2.5}$ concentration as the dependant variable. Eq. A1 provides the estimation equation to identify the effect of city shape on commuting.

Table A1: Main results (max)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Shape index:		Disconnection			Range			Remoteness	
		IV	OLS		IV	OLS		IV	OLS
	First stage	Second stage		First stage	Second stage		First stage	Second stage	
Dependent variable:	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)
Shape (km)		-0.0045*	-0.0025		-0.0014**	-0.0006		-0.0059*	-0.0038
		(0.0025)	(0.0025)		(0.0007)	(0.0007)		(0.0031)	(0.0029)
log(area (sq.km))	0.9993***	0.0140***	0.0125***	2.8995***	0.0134***	0.0119***	0.7536***	0.0139***	0.0128***
	(0.1087)	(0.0040)	(0.0037)	(0.3481)	(0.0038)	(0.0036)	(0.0840)	(0.0039)	(0.0036)
Potential shape (km)	1.0258***			1.4143***			1.1035***		
	(0.0936)			(0.1184)			(0.1062)		
Observations	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414
F-test stat	11.55			10.52			8.66		
City fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. 1, 2 and 3 with log of $PM_{2.5}$ concentration as the dependent variable measured in micrograms per cubic meter ($\frac{\mu g}{m^3}$). Column (1), (4) and (7) report the results for the first stage. Column (2), (5) and (8) report results for the second stage. Column (3), (6) and (9) report results for the pooled OLS. Standard errors are clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Main results (min)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Shape index:		Disconnection			Range			Remoteness	
		IV	OLS		IV	OLS		IV	OLS
	First stage	Second stage		First stage	Second stage		First stage	Second stage	
Dependent variable:	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)	Shape	log(PM 2.5)	log(PM 2.5)
Shape (km)		-0.0089**	-0.0056*		-0.0026***	-0.0015*		-0.0116***	-0.0082**
		(0.0038)	(0.0034)		(0.0010)	(0.0009)		(0.0488)	(0.0044)
log(area (sq.km))	0.9993***	0.0026	0.0002	2.8995***	0.0012	-0.0010	0.7536***	0.0023	0.0005
	(0.1087)	(0.0043)	(0.0038)	(0.3481)	(0.0039)	(0.0036)	(0.0840)	(0.0041)	(0.0036)
Potential shape (km)	1.0258***			1.4143***			1.1035***		
	(0.0936)			(0.1184)			(0.1062)		
Observations	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414	6,414
F-test stat	11.55			10.52			8.66		
City fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Regression results from Eq. 1, 2 and 3 with log of $PM_{2.5}$ concentration as the dependent variable measured in micrograms per cubic meter ($\frac{\mu g}{m^3}$). Column (1), (4) and (7) report the results for the first stage. Column (2), (5) and (8) report results for the second stage. Column (3), (6) and (9) report results for the pooled OLS. Standard errors are clustered at the city level. *** p<0.01, ** p<0.05, * p<0.1.

The effect of city shape on actual commuting patterns is estimated as follows:

$$log(y_{ids}) = \beta_1 shape_{ids} + \beta_2 ln(area_{ids}) + \alpha_s + \epsilon_{ids}, \tag{A1}$$

where the dependant variable y_{is} is the share of people in city i in district d in state s that commute $>$ 10km. In alternative specifications, the dependant variable refers to the share of people that (i) commute at all or (ii) commute with a specific mode of transport.

The variable $shape_{ids}$ is the respective shape property, that measures the compactness (in km) of city i in district d in state s, where a higher value indicates a less compact city shape. I further control for the log of the total area of a city and state fixed effects α_s. Standard errors are clustered at the state-level. The first and second stage for the IV approach are implemented analogously to Eq. 2 and 3 in Section 2.