Urban pollution: A global perspective

Rainald Borck
Philipp Schrauth

CEPA

Center for Economic Policy Analysis

CEPA Discussion Papers

Urban pollution: A global perspective[*]

Rainald Borck
University of Potsdam, CESifo, DIW Berlin
Philipp Schrauth
University of Potsdam

ABSTRACT

We use worldwide satellite data to analyse how population size and density affect urban pollution. We find that density significantly increases pollution exposure. Looking only at urban areas, we find that population size affects exposure more than density. Moreover, the effect is driven mostly by population commuting to core cities rather than the core city population itself. We analyse heterogeneity by geography and income levels. By and large, the influence of population on pollution is greatest in Asia and middle-income countries. A counterfactual simulation shows that $PM_{2.5}$ exposure would fall by up to 36% and NO_2 exposure up to 53% if within countries population size were equalized across all cities.

Keywords: population density, air pollution, gridded data
JEL Codes: Q53, R12

Corresponding author:
Rainald Borck
University of Potsdam
Chair of Public Economics
August-Bebel-Str. 89
14482 Potsdam, Germany
Email: rainald.borck@uni-potsdam.de

1 Introduction

Pollution is an important determinant of urban quality of life. Households have flocked to cities over the last centuries and decades, attracted by various agglomeration economies, such as higher productivity and wages. However, city life has been and still is, to different extents, plagued by agglomeration costs stemming from crime, congestion, and pollution. Besides, urbanization and environmental degradation are not evenly spread throughout the world. While developed countries were already more than 50% urbanized by 1950, this threshold has been reached by less developed countries only in 2020.[1] Accordingly, the bulk of the recent and imminent increase in world urbanization will occur in developing regions. This is also where urban air pollution is most severe. For example, taking the average $PM_{2.5}$ concentration value from the WHO air quality database from 2022,[2] 20 of the 25 dirtiest cities were located in India, China, Bangladesh or Pakistan, with the remaining in Cameroon, Iran, Mongolia, Madagascar and Afghanistan. Therefore, the relationship between agglomeration and pollution is also a question of socio-economic development. Reigning in pollution, especially in large cities, will be important as developing countries thrive to improve their citizens' well-being. Yet, while there is an extensive literature on the benefits stemming from agglomeration economies, there is much less research on the costs of agglomeration (Ahlfeldt and Pietrostefani, 2019).

In this paper, we contribute to filling this gap. We use global gridded data on air pollution and population to analyse how agglomeration, in the form of large and densely populated cities, affects exposure to $PM_{2.5}$ and NO_2 pollution.

In theory, population density might increase or decrease pollution concentration in cities. Borck and Schrauth (2021) present a model where residents of a monocentric city pollute due to commuting and residential energy use for heating, electricity, etc. They show that population density increases pollution concentration. The reason is that larger and more densely populated cities have more aggregate commuting and that residential energy use increases as well, even though residents live in smaller dwellings on average.[3] However, there are some countervailing forces. For instance, public transit is more viable in large and densely populated cities due to economies of density, and denser housing is more energy efficient. Therefore, the relation between density and pollution is theoretically ambiguous.

Similar opposing forces determine whether cities with larger total population (as opposed to density) are more polluted (see Borck and Pflüger, 2019; Borck and Tabuchi, 2019). Further, the relation between population density and pollution is likely to depend

---

[1] See UN Urbanization Prospects, `https://population.un.org/wup/`.

[2] See WHO Air Quality Database from April, 2022, `https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database`.

[3] In the model, total pollution increases more than urban area, which means that pollution concentration increases.

on many factors that vary between regions, such as geography, institutions (environmental policies) etc. Therefore, an interesting question that we look at is how pollution and its relation with density varies between regions with different characteristics.

We use 11-16 years (depending on the pollutant) of gridded satellite data to document the distribution of pollution over space and time. There are several main findings. First, we show that about 3/4 of the world population and about 79 percent of city dwellers live in places with particulate pollution above thresholds as recommended by the WHO. Thus it seems that pollution is especially severe in cities. We go on to estimate the elasticity of pollution with respect to population density for $PM_{2.5}$ and NO_2. Using OLS regressions with country fixed effects, we find elasticities of 0.15-0.16 for NO_2 and 0.02-0.03 for $PM_{2.5}$. To tackle concerns of reverse causality and omitted variables, we also instrument population density using historical populations from different periods in time. Doing so has only a very small effect on the estimated elasticities.

We present our results using both grid cells and cities (Functional Urban Areas, FUA) as units of observation. Examining cities allows us to explicitly differentiate between the different effects of agglomeration size versus population density on exposure. For cities, we find that population size seems to be more important than density. Furthermore, using the definition of FUAs allows us to differentiate between the core city and its surrounding commuting zone. In fact, it turns out that pollution exposure is not significantly affected by core city population, but does rise significantly with population living in FUAs' commuting zone.

Moreover, we study how the pollution-density relationship varies over continents and by income. For the rasterized global data, we find that the pollution-density relation is strongest in middle income countries and in Asia. For the city data, population/density affect pollution most in upper middle and high income countries as well as in Europe and North America.

We also present outcomes on a more local level by estimating the effect of within-city variations in density. Again, we find positive effects of density on exposure, but the effects are mostly smaller in size. Additionally, we estimate spatial first difference regressions, where the estimated elasticities are based on changes between neighbouring grid cells (Druckenmiller and Hsiang, 2018). The corresponding coefficient estimates turn out to be positive, but again smaller in size.

Lastly, we perform a simple counterfactual simulation. Using the exposure–population elasticity from our city analysis, estimated separately for each country, we ask how each country's total exposure would be affected by an equal redistribution of population across cities. We find that for $PM_{2.5}$, exposure falls by 36.5% for the country with the largest drop (Indonesia), which has a large estimated elasticity. Conversely, there are some

countries with negative elasticities, so exposure would rise in this counterfactual by a maximum of 22.5% in Senegal.

The study contributes to a small but growing economic literature on urban pollution generally, and on the relation between agglomeration and pollution in particular. Empirical papers in fields other than economics have largely been confined to cross-sectional studies.[4] However, omitted variables and reverse causality are difficult to tackle in these settings. Among the few serious efforts to identify the causal effect of population density on pollution are Borck and Schrauth (2021) and Carozzi and Roth (2022). Borck and Schrauth (2021) use panel data from German districts, while Carozzi and Roth (2022) use cross-sectional data from US metropolitan areas. Both papers instrument density with a variety of historical and geological instruments. Castells-Quintana et al. (2021) and Aldeco et al. (2019) also study global pollution. Aldeco et al. (2019) focus on studying the effect of various policies using a spatial equilibrium model. Castells-Quintana et al. (2021) is also closely related to our paper, but there are several differences. They study emissions in a global panel of cities, while we analyse exposure in both cities and raster cells, which allows for a truly global analysis and lets us study the urban-rural pollution gradient in addition to cross-city differences. Moreover, we do a variety of heterogeneity analyses, and instead of emissions, we look at pollution exposure which is more tightly linked to local welfare.

The paper is organized as follows. The next section presents our data, descriptive analyses and empirical approach. Section 3 shows the results. In Section 4, we simulate how total exposure would change if, within countries, we were to redistribute population equally among all cities. The last section concludes the paper.

2 Data and estimation

2.1 Data

Most data sets we use are derived from satellites and are provided as a grid of raster cells covering the entire world. Those rasterized grid maps come in different resolutions, mostly between 0.01 and 0.25 decimal degrees. We transform all data to 0.25 degree raster cells, which is a compromise between the different levels of aggregation of the native data and moreover alleviates concerns about auto-correlation at finer scales. At the equator, a quarter degree grid corresponds to 27.8 kilometres into one direction or

---

[4]See, e.g. Sarzynski (2012) and Lamsal et al. (2013). See also Borck and Schrauth (2021) for further references.

roughly 775 square kilometres overall.[5] For our analyses, we use the years 2000, 2010 and 2015. In the following, the different data sets are described in more detail.[6]

2.1.1 Units of observation

In the analyses we use two types of observational units. The first are raw grid cells from the raster data. The second are cities or "functional urban areas" (FUA) as defined by Moreno-Monroy et al. (2020). These FUA are cities and their surrounding areas with strong internal commuting links. We view these two data sets as providing complementary results. On the one hand, defining cities gets us closer to measuring activity in economically meaningful areas. On the other hand, using all grid cells – even very thinly populated ones – allows us to measure an urban-rural gradient of pollution. Thus, our paper differs from and complements other papers that have mostly studied cities only (e.g. Carozzi and Roth, 2022; Castells-Quintana et al., 2021).[7]

Raster data. The first units of observations in our analysis are grid cells of a rasterized world map. The majority of the data we use is provided as raster maps, which then can be matched to each other geographically. The advantage of looking at grid cells is that we abstract from defining cities or urban areas and that there is an increasing database of worldwide data covering a wide range of topics. In addition, it will allow us to measure an urban-rural pollution gradient since observations take into account any type of inhabited land and do not depend on city definitions. We use grids of 0.25 decimal degrees and aggregate all the other raster maps to this size. This leaves us with more than 240,000 cells that fall on land to which we make some minor adjustments.[8] The chosen grid size is a compromise between data that is available at relatively fine grid scale, and data that is available at coarser levels only. It also mitigates concerns about spatial auto correlations.[9]

---

[5]Moving away from the equator means that equally sized grids cover smaller areas due to the curvature of the earth. At the 45th degree of latitude for example, which crosses South Dakota, Mongolia, France and Italy, 0.01 decimal degrees are equal to 787.1 meters in one direction. The value approaches zero at the poles. Most of human activity takes place between the 50th parallel south and the 60th parallel north.

[6]The NO_2 data is only available from the years 2000 to 2012, of which we use the years 2000 and 2010. The $PM_{2.5}$ data is available until 2015.

[7]Castells-Quintana et al. (2021) also look at the effect of density and polycentricity on pollution at the country level.

[8]We drop grid cells which cannot be assigned unambiguously to one single country. As a consequence, about 21,400 grid cells that lie at country borders are dropped from the sample. Furthermore, we harmonize the country composition of our city and grid cell samples. Thus, all countries which do not contain at least one Functional Urban Area are dropped.

[9]For variables available at higher grid resolutions, e.g. 0.1 decimal degrees, we re-project the data to $\frac{1}{4}$ degrees using an appropriate function: For continuous variables, we calculate either the mean of all smaller grids within the respective quarter degree grid (pollution exposure for example is mean

Cities. There are several reasons why we want to define city delineations. First, defining cities allows us to distinguish between city size and density. In a grid with equally sized grid cells, density would be strictly proportional to population. While basic urban economic theory also predicts a positive relation between population size and density (e.g. Brueckner (1987)), in practice the two vary independently, for instance, due to differences in zoning policies across cities. Since different agglomeration economies and diseconomies may operate at different spatial scales, population size and density might then affect pollution differently (Ahlfeldt and Pietrostefani, 2019; Cheshire and Magrini, 2008). Second, some of the $PM_{2.5}$ pollution stems from sources not directly attributable to daily human activities such as volcanoes or wildfires (see e.g. NASA Earth Observatory (2015)). While this may be interesting in its own right (if it leads to rural areas being dirtier than they would otherwise be), abstracting from these types of events by focusing on urban areas allows us to concentrate on the effect of human activity in cities on pollution. Third, we can conduct between city analyses to supplement the rural-urban gradient. This type of city size effect helps us connect the empirical analysis with theoretical considerations about optimal city size (see e.g. Borck and Tabuchi (2019)). Fourth, we can detect within city differences. Thus, we can study whether there is a core-periphery gradient of pollution exposure within cities and we can compare it to between-city effects or the urban-rural gradient of pollution exposure. Lastly, our historical population instruments consist of geo-coded city locations. Directly instrumenting urban areas rather than grid cells thus seems more adequate.

We define cities as Functional Urban Areas (FUA) following Moreno-Monroy et al. (2020). They use population and travel time data in 2015 to define unique urban centres including their commuting zones.[10] Thus, our city definitions do not vary over time. A FUA consists of an urban core with at least 50,000 inhabitants and the surrounding commuting zone, which is constructed using travel times. FUA were originally defined by the OECD for OECD countries and Colombia. Moreno-Monroy et al. (2020) use those OECD-defined FUAs to estimate city boundaries for the rest of the world.

Figure 1 shows our two main units of analysis and the population distribution as provided by LandScan for the north-eastern USA. More precisely, the figure visualizes the New York, Philadelphia, Baltimore and Washington, D.C. area and the area's FUAs. Grid-cell analyses consider all the non-white grid cells that fall on land. Densely populated city cores are shown in red, and less dense suburbs and rural areas are shown in yellow and blue. The greyish transparent polygons overlaying the population grids depict the resulting FUAs.

---

exposure within a quarter degree grid) or sum the values of these finely scaled grids, as appropriate. For categorical variables we take the modal value within a quarter degree grid.

[10]The geo-package with the FUA shapefile is publicly available at https://ghsl.jrc.ec.europa.eu/download.php.

Overall, there are 9,031 FUAs in 188 countries and about 245,000 grid cells in 185 countries. Since the main part of analyses contains within-country effects, we mostly restrict the samples such that very small countries with very few raster cells or cities are dropped.

Figure 1: Grid-cell units, Functional Urban Areas and LandScan population. New York, Philadelphia, Washington D.C.
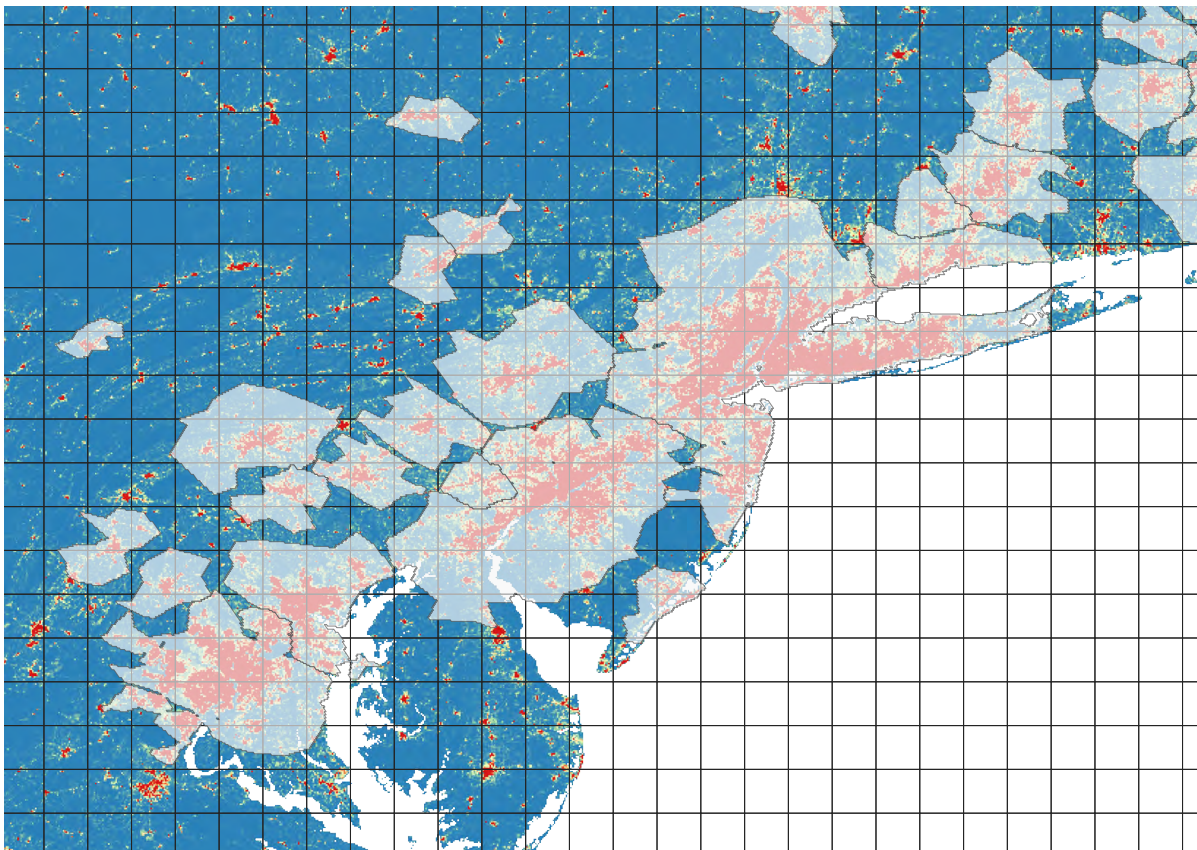
Note: The graph depicts a small section of the entire sample. White areas are water surfaces (lakes and oceans) that are excluded from estimations. The overlying grid corresponds to the raster units, while the greyish transparent polygons show single FUAs. Within both observation units, LandScan population data are shown in different colour gradations. Blue indicates very low population density and red indicates very high population density.

2.1.2 Air pollution data

The most direct measure of ground-level pollution concentration would be measurements from in-situ monitors. However, these are not widely available, especially in many lower income countries. Even among high income countries, only selected areas contain monitoring stations. To get coverage of world wide pollution we use satellite data, which captures air pollution concentration as vertical column densities in the troposphere. For

ground-level observations, we resort to data sets that use chemical transport models to translate satellite measures into ground-level pollution.[11] The data products are annual means of dust and sea-salt removed $PM_{2.5}$ from 2000 until 2015 at 0.01x0.01 decimal degree (dd) resolution and annual means of NO_2 without corrections from 2000 until 2010 at 0.1x0.1 dd resolution.[12] For most of our analysis, we use the data that are weighted using geographically weighted regression (GWR), but outcomes are not sensitive to using the non-weighted data.[13]

This native data serves to construct our pollution measure of interest, which is pollution exposure (see e.g. Carozzi and Roth (2022) or Aldeco et al. (2019)). Population-weighted pollution exposure E in grid cell G is then given by:

$$E_G = \sum_{i \in G} P_{G,i} \times \frac{N_{G,i}}{\sum_{j \in G} N_{G,j}},$$

where i indexes small grid cells within a large grid cell G.[14] Average pollution concentration in cell i is given by $P_{G,i}$ and population in i is represented by $N_{G,i}$. Average pollution exposure is therefore the sum of grid-specific pollution exposure divided by overall population in a large unit G. We also repeat the analysis with our FUA dataset, where G represents a city instead of a raster cell.

2.1.3 Population density

Present population and density. For measures of population and population density, we use LandScan data (LandScan, 2018). Population in this data set is provided on a very fine spatial scale (30 arc-seconds)[15], obtained from censuses and other sources worldwide. The data aims to show where people are located on average over the course of 24 hours. Thus, it includes place of residence and of work in its estimations.[16] However, there is no information about how exactly the information is implemented in the population grid estimates. Henderson et al. (2021) provide a "ground-truthing" exercise

---

[11]See Hammer et al. (2020) and Van Donkelaar et al. (2016) for $PM_{2.5}$ and Lamsal et al. (2008) for NO_2. This data is available online (Atmospheric Composition Analysis Group, 2018).

[12]Satellite measurements of pollution are captured between 8:30 a.m. and 11:00 a.m. local time, depending on the satellite. The cell size means that a raster contains about one square kilometre at the equator compared to the 120 km^2 NO_2 raster cells.

[13]GWR uses in-situ (on the ground) monitors to detect regional biases of satellite optical depth measurements. This bias is estimated and then corrected using different predictors like land cover or elevation difference (Van Donkelaar et al., 2015). The advantage of GWR is the more accurate representation of ground-level $PM_{2.5}$.

[14]For example, a $PM_{2.5}$ observation in a cell of 0.01x0.01 dd within a larger grid cell that spans over 0.25x0.25 dd.

[15]30 arc seconds correspond to a little less than 0.01x0.01 dd, which is about 1km at the equator.

[16]By contrast, other data sets distribute population obtained from administrative sources equally in space or use buildings as a proxy for where people live without distinguishing whether those buildings are commercial or residential ones.

and conclude that LandScan data perform well and are suitable for analyses on a global scale.[17]

Historical population measures. We also instrument population or density using historical population data, following a large literature in urban and regional economics since Ciccone and Hall (1996).

The data comes from Reba et al. (2016), who provide geo-referenced population data worldwide ranging from 3700 B.C. to 2000 A.D. using historical, archaeological, and census-based estimates. This data set comprises about 1500 settlements worldwide. We use it to construct two different instruments. The main instrument will be the population in 1900. This is the year with most observations in the historic population sample prior to 1914 and represents the population in industrialized times before the two world wars. The second instrument is population in the last year of observation before 1750, and thus in pre-industrialized times. In Section 2.3 below, we will come back to the issue of instrument relevance and exogeneity.

In the raster analysis, we assign historical population to single grid cells. Thus, each grid cell is instrumented with the historical population data of the settlement that lies within this grid cell. Some grid cells contain several data points. In this case, we sum up the population over these data points. The drawback of instrumenting with historical population counts is that the estimation sample is drastically reduced and that there is a concentration of settlements in economically developed countries. Since using those instruments therefore deprives us of much valuable information, we mainly present the IV results in order to compare them with OLS outcomes on a harmonized sample. For most of this study, we will concentrate on within-country OLS estimates.

2.1.4 Controls

We use a number of variables in order to control for potential observable factors that may be correlated with population density and pollution. Income is an obvious candidate variable that correlates with population density (Combes and Gobillon, 2015) and affects pollution. To control for economic development, we therefore use GDP from the dataset provided by Kummu et al. (2018). It is based on subnational accounts, like states in the U.S. or districts in Germany. In some specifications we control for the presence of coal-fired and other highly polluting power plants in a grid cell.[18] Since power plants may be close to dense areas, this may be one channel through which density affects pollution.

---

[17]This finding is confirmed by Galdo et al. (2019) who use machine learning combined with human judgment to identify urban areas in India and compare their outcomes with LandScan data.

[18]The data is available at https://datasets.wri.org/dataset/globalpowerplantdatabase.

We also control for a variety of topological and climatological variables that may be correlated with density and pollution, such as ruggedness, temperature, wind speed, and precipitation. We compute ruggedness following Nunn and Puga (2012), which is roughly the grid-cell average difference in elevation between a point and the terrain surrounding it. Our ruggedness measure is calculated over land surface only, leaving out water.[19] Temperature and precipitation are taken as long-run averages over the 30 year period between 1960 and 1990 (FAO/IIASA, 2012). Wind data is retrieved from the global wind atlas (Davis et al., 2019).

In addition to weather, we control for variables that might influence pollution through their suitability for trade on the one hand and through their climatological impact on the other. These controls consist of three dummy variables (whether a major river, a large lake, or a coastline is within 25 kilometres of a grid centroid) and a continuous variable (distance of grid centroid to coast). In our city-level analysis, we measure the distance of a city border to the respective first nature characteristic. Data for coastlines, rivers and lakes come from Natural Earth (2018).[20]

Moreover, we control for the agricultural suitability of a raster cell or city (compare Henderson et al., 2018). Locations that are densely populated due to their fertile soil may also suffer from high pollution, since agriculture is a strong producer of particulate matter pollution. We thus take into account land suitability (as continuous variable), and a set of biome indicators. Land suitability for agriculture is based on measures of climate and soil and predicts the probability of land to be cultivated (Ramankutty et al., 2002). Biomes describe the ecological system of an area and its dominant natural vegetation. These categories include for instance "tundras", "tropical and subtropical dry broadleaf forests", or "Deserts & Xeric Shrublands". The 14 biome indicators we use are taken from Olson et al. (2001).[21]

We want to assess a variety of country features that may influence the relation between population and pollution. In order to do so, we make use of a large set of national indicators from the World Bank that contains economic performance, energy use or demographic characteristics (World Bank, 2019). This data set ranges from 1960 to 2018, which allows us to calculate for instance means of urbanization rates and their

---

[19]In order to calculate the fraction of water surface we use the world map gridded at 1 km resolution provided by Lloyd et al. (2017), which is based on a global water mask dummy variable gridded at very high resolution (Feng et al., 2016).

[20]We take the "high" resolution datasets from Natural Earth (2018). Rivers are categorized into 10 size ranks, where 1 are the largest and 10 the smallest rivers. We only consider rivers between the ranks 1-6. Large lakes are those with a surface area greater than 5,000 square kilometres, excluding unnatural dams. This leaves us with 29 major lakes.

[21]Just like Henderson et al. (2018) we combine the categories "tropical and subtropical dry broadleaf forests" with "tropical and subtropical coniferous forests" as well as "tropical and subtropical grasslands and savannas and shrublands" with "flooded grasslands and savannas". Furthermore, we drop areas historically covered by ice or rocks from the analysis. Doing so does not, however, change our results.

growth, residents of large agglomerations or energy use over different time periods. In our analysis below, we will correlate some of these measures with the density-elasticity of pollution.

2.2 Descriptive analysis

Human health is vulnerable to pollution. The WHO has set short- and long-run thresholds to indicate very high pollutant concentration levels that presumably pose major threats to human health. Short-run refers to the average value over 24 hours for PM_{10} and one hour for NO_2, while long-run means the annual mean. In our sample, the long-term threshold of $40\mu g/m^3$ for NO_2 is not exceeded in a single raster or city, which would seem to suggest that this pollutant does not constitute a major health threat.[22] However, Borck and Schrauth (2021) show that NO_2 levels for both the short- and the long-run thresholds are transgressed in Germany on a more local level. The respective data is taken from in-situ monitors and therefore provides a more accurate local measure of air pollution. Indeed, the raster size we have available for NO_2 does not allow for a very local consideration of this pollutant. Furthermore, we only have available annual means, so we cannot analyse short-run threshold transgressions.

Things look different for $PM_{2.5}$-pollution. Table 1 shows the share of population that lives in raster cells or cities where annual mean $PM_{2.5}$ pollution exceeds the short-run $(25\mu g/m^3)$ or long-run $(10\mu g/m^3)$ WHO thresholds. In 2015, out of the 7.26 billion people in the sample, about 5.52b lived in raster cells with mean $PM_{2.5}$-levels beyond the long-term threshold of 10 $\mu g/m^3$. This corresponds to around 76 percent of the overall world population. Approximately 39 percent were even permanently exposed to concentrations beyond 25 $\mu g/m^3$, which is the WHO 24-hour mean and therefore recommended to be avoided over periods longer than a day. Note, however, that within a raster cell there is variation in pollution concentrations such that not everybody is actually exposed to those pollution concentrations. Therefore, the actual number of people permanently exposed to such high concentrations may lie below 75 percent.

If we only consider FUA, there are about 4 billion people living in urban areas (this is in line with the UN's estimate of worldwide city population) of whom about 79% live in cities with long-term mean $PM_{2.5}$ pollution beyond $10\mu g/m^3$. About 44% are in cities where average annual urban pollution even exceeds the short-term threshold of $25\mu g/m^3$.

There are marked differences between continents. In Asia, 92 percent of the population face an annual average $PM_{2.5}$-pollution beyond $10\mu g/m^3$. In Europe the corre-

---

[22] The WHO has recently updated its thresholds (for example, the annual thresholds are now 10 $\mu g/m^3$ for NO_2 and 5 $\mu g/m^3$ for $PM_{2.5}$.), but we use the thresholds that were published during the period we study.

Table 1: Descriptive statistics

	Raster		Cities	
	Mean (Std)	Min Max	Mean (Std)	Min Max
PM$_{2.5}$	8.07 (8.51)	0 104.94	31.46 (24.82)	.9 119.57
NO$_2$	.38 (0.77)	0 28.7	1.70 (2.31)	0.01 17.85
Population	29,779 (158,719)	0 11,501,275	435,016 (1,384,120)	50,079 36,471,787
Overall population	7.26 billion		3.93 billion	
Share population exposed to:				
PM$_{2.5}$ > 10 $\mu g/m^3$	76%		79%	
PM$_{2.5}$ > 25 $\mu g/m^3$	39%		44%	
Countries	185		188	
Observations	244,649		9031	

Note: Own calculations. The table shows descriptives of unweighted pollution provided by Atmospheric Composition Analysis Group (2018) for the years 2010 (NO$_2$) and 2015 (PM$_{2.5}$). Population data is shown for the year 2015 and is taken from LandScan (2018). For the calculation of PM$_{2.5}$-exposure with values larger than 10 $\mu g/m^3$ or 25 $\mu g/m^3$ we sum the population in grid cells (FUAs) that match a pollution cell that is higher than the respective value in 2015 and divide it by the overall population in the respective sample.

sponding number is 67 percent, in Africa 62 percent, in North-America 37 percent, and in South America 32 percent.[23]

Figure 2 shows the geographical distribution of PM$_{2.5}$ in 2015 and its change from 2000 to 2015. Dark grey/black areas in panel (a) are highly polluted with values close to or larger than 25 $\mu g/m^3$ while light grey/white ones may have values close to or below the annual WHO threshold of 10 $\mu g/m^3$. In Panel (b), dark grey/black areas have seen an increase in PM$_{2.5}$-concentrations between 2000 and 2015, while light grey/white areas experienced little change or even a decline. Apparently, many of the highly polluted areas in 2015 either developed into highly polluted areas or became even more polluted over the course of 15 years. The pollution problem has become much more serious especially in India and China, but also in Africa south of the equator (net of dust and sea salt). Many areas in the U.S., especially in the east, have improved their air quality over time, which is also the case in parts of Western Europe.
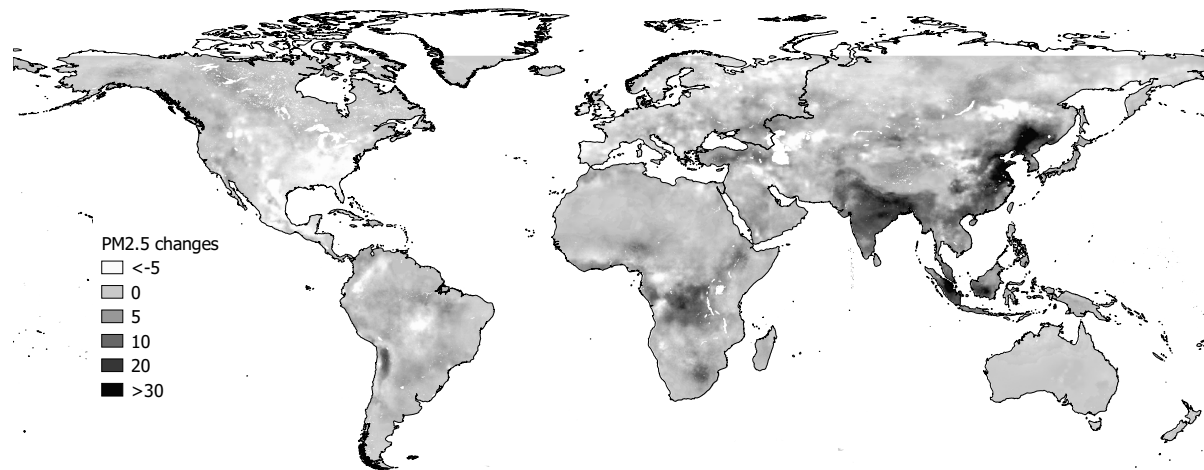
Pollution of NO$_2$ is much more concentrated in a few areas as shown in Figure 3a. The highest concentration exposure is found in North-eastern China, Middle Europe,

---

[23]These numbers are based on the raster data.

Figure 2: PM$_{2.5}$ concentration and development over time
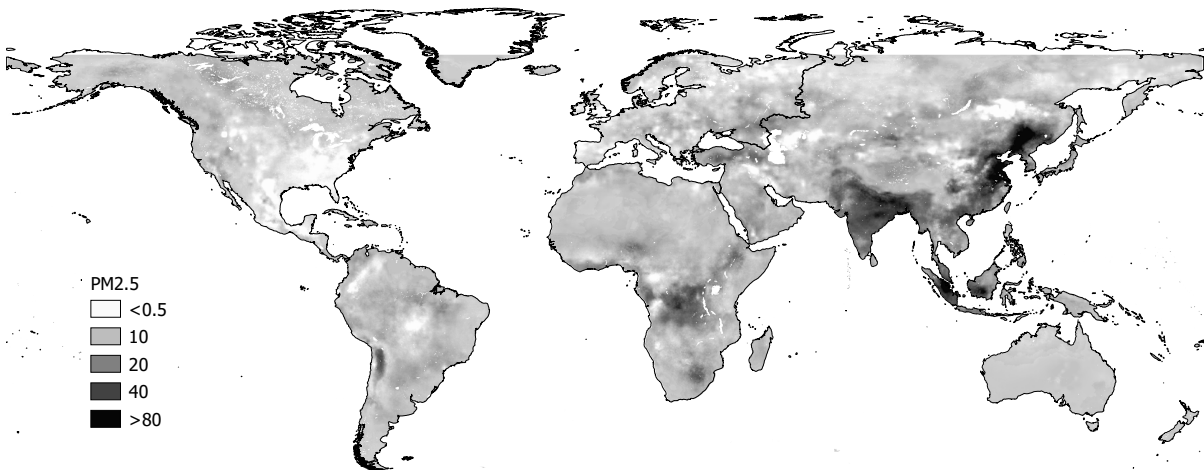
PM2.5
<0.5
10
20
40
>80

(a) PM$_{2.5}$ concentration in 2015

PM2.5 changes
<-5
0
5
10
20
>30

(b) PM$_{2.5}$ change 2000-2015

Note: Figure 2a shows the worldwide distribution of PM$_{2.5}$ concentrations in 2015. White/light grey depicts low and dark grey/black high concentration levels. Figure 2b shows developments of PM$_{2.5}$ levels between 2000 and 2015. White/light grey areas saw negative, little or no change in pollution concentration; dark grey/black areas have experienced large increases in PM$_{2.5}$ pollution.

parts of the United States and parts of Russia. The range of values is much smaller for NO$_2$. The maximum value reached is 30 $\mu g/m^3$ in 2010. Figure 3b shows the change in NO$_2$ concentration levels. Again, dark grey/black areas are those where pollution most strongly increased from 2000 until 2010. Some parts of the U.S. and Europe have experienced air quality improvements. In Africa, Australia, and South America, NO$_2$ concentrations barely changed. Predominantly densely populated metropolitan areas such as Santiago de Chile, Cairo, or São Paulo seem to have higher pollution levels in 2015 compared to 2010.

Figure 3: NO$_2$ concentration and change over time
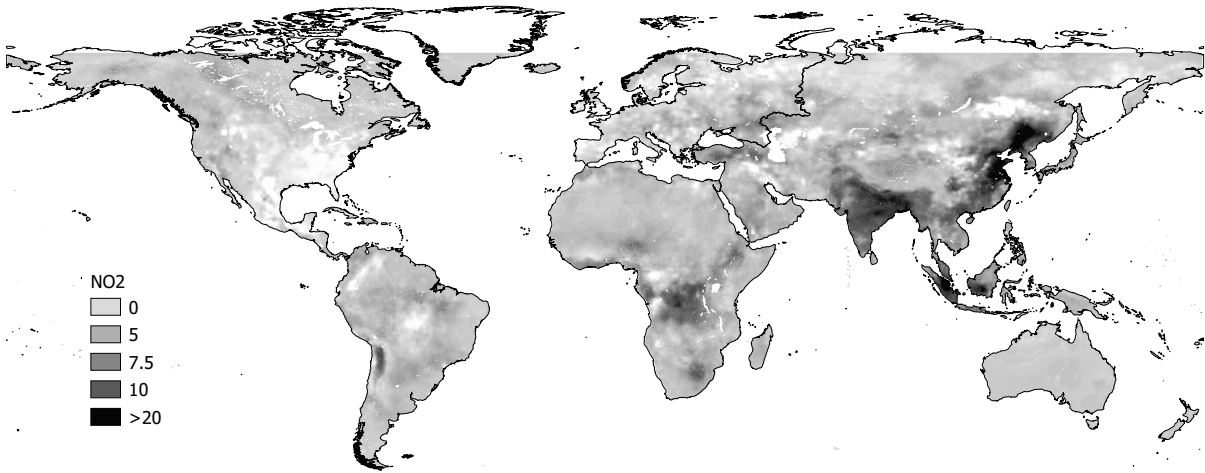

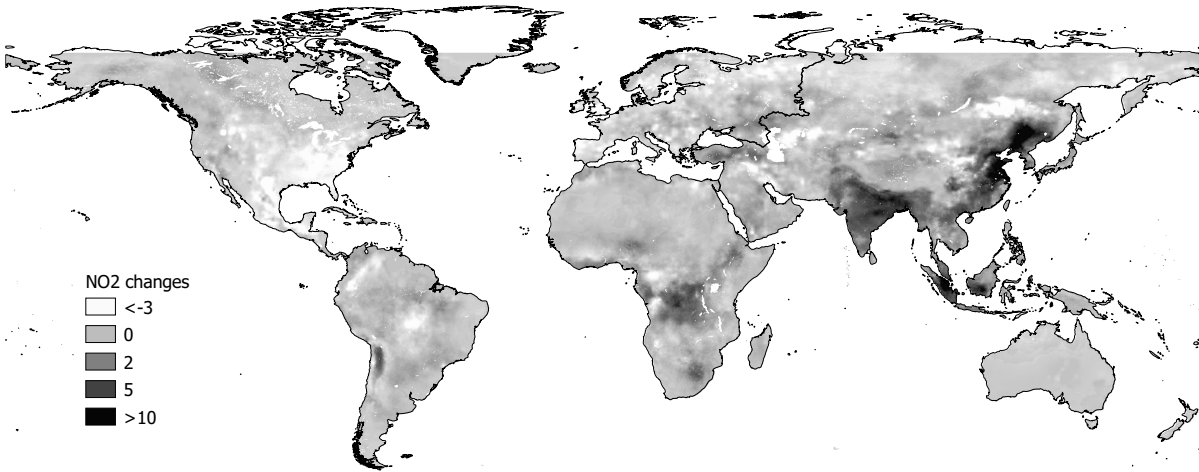(a) NO$_2$ concentration in 2010

(b) NO$_2$ change 2000-2010

Note: Figure 3a shows the worldwide distribution of NO$_2$ concentrations in 2010. White/light grey depicts low and dark grey/black high concentration levels. Figure 3b shows changes of NO$_2$ levels between 2000 and 2010. White/light grey areas saw negative, little or no change in pollution concentration; dark grey/black areas have experienced large increases in NO$_2$ pollution.

We now turn to our regression framework for estimating the effect of population on pollution exposure.

2.3 Estimation

In a first step, we run simple Ordinary Least Square (OLS) regressions of air pollution exposure on population (density) and control variables. To mitigate concerns about spatial autocorrelation, we analyse pollution within relatively large $\frac{1}{4}$ decimal degree grids and we cluster standard errors within three-by-three squares of grid cells times

13

year of analysis (following Henderson et al., 2018). This clustering approach accounts for the potential correlation of pollution in space since particulates for instance disperse spatially with the wind. In a second step, we restrict the sample to cities as defined above.

The OLS regression equation is:[24]

$$\ln(E_{GtS)}) = \beta + \rho \ln(D_{GtS}) + \gamma X_{GtS} + \alpha_t + \epsilon_{GtS}, \tag{1}$$

where E_{GtS} is exposure to NO_2 or $PM_{2.5}$ in grid cell/FUA G and year t in country S (in our baseline regressions we only include the last year of observation, i.e. 2010 for NO_2 and 2015 for $PM_{2.5}$). Our parameter of interest, ρ, measures the elasticity of pollution exposure with respect to population density D_{GtS} (or population N_{GtS}, depending on the specification). X_{GtS} is a vector of control variables. As explained above, these contain the log of GDP, several variables about the suitability for trade (whether the raster/city lies on a river, on a lake or on the coast) and agriculture (land suitability, and biome indicators), temperature, precipitation (both as 1960-1990 long term means), wind speed, the presence of dirty power plants, ruggedness and latitude.

We will compare OLS results to within-country estimates, which include a country dummy θ_S. These within-country regressions compare raster cells/cities within a country to each other. The rationale is to control for any countrywide unobserved geographic or political features that may be correlated with both pollution and population (density).

Even though country fixed effects already account for a large portion of unobservables, OLS regressions may still be biased due to reverse causality or omitted variables. Economic theory and empirical evidence suggests that households would want to move to cleaner areas (Chen et al., 2022). Hence, population would be endogenous to pollution exposure. Moreover, within countries, there may be unobservable differences in policies, attitudes, and the like that are correlated with population measures and pollution. We therefore follow the urban economics literature in instrumenting population with historical population levels.[25]

The main assumptions for using historical population data have been widely discussed in the urban economic literature. First, the distribution of population and economic activity tends to be persistent over time (see for instance Davis and Weinstein, 2002). This is intuitive, since infrastructure and buildings are durable and thus population changes are sluggish. Therefore, historical population is a good predictor of current agglomerations. Second, the exclusion restriction states that historical population should affect

---

[24]Note that in the main analysis, we only look at a cross-section. We keep the time index because we run fixed-effects panel regressions as a robustness check.

[25]See, e.g. Combes et al. (2010) for a classic application and discussion of the issues in estimating agglomeration economies. See Borck and Schrauth (2021) and Carozzi and Roth (2022) for the same approach in estimating density effects on pollution in Germany and the US.

pollution only through its effect on current population. The argument is that, if we go back in time far enough, structural change will have led to a reshaping of local economies such that historic population levels should be exogenous to current pollution levels. Suppose, for instance, that a city formed close to a river in pre-industrial times in order to benefit from the trade advantage conferred by the river. It may have grown into a densely populated and highly polluted place nowadays due to industrial and traffic pollution. Then, the exogeneity assumption would be satisfied, since today's agglomeration pattern and its effect on pollution (namely, motorized traffic and industrial production) differs from the historic one (trade).

In constructing historical instruments, there is a trade-off: on the one hand, the exogeneity argument forces us to go back sufficiently long in time. On the other hand, availability constraints force us to use more recent data in order to have a sufficient number of observations.[26] With respect to exogeneity, we believe that population counts prior to 1750 have the stronger arguments compared to more recent population instruments. Before the industrial revolution, which started in the second half of the 18th century, air pollution was probably not a decisive factor for migration decisions, whereas during industrialization, there seems to be already some evidence of sorting with respect to pollution (Heblich et al., 2021).[27] Using population in 1900, however, provides us with at least twice as many observations than more historical population counts. We will use population in 1900 as our main instrument, but outcomes do not differ much using population from pre-industrialized times.[28]

We instrument population (density) as follows:

$$\ln(D_{GtS}) = \ \alpha + B_2 X_{GtS} + B_3 Z_{GtS} + \eta_{GtS}, \tag{2}$$

where the instrument Z is historical population. The predicted values for population, $\ln(\hat{D}_{GtS})$, are then used in the second stage instead of actual measures of population in equation (1).

We also present results from long-difference estimations. These regress the changes in exposure between the last and first year of observation on changes in population/density. The idea here is that there may be some unobserved differences between units that simultaneously affect population density and pollution. For instance, sorting of "green" individuals into large cities might lead to a negative correlation between density and

---

[26]We also experimented with soil quality and other natural causes as instruments like Borck and Schrauth (2021) and Combes et al. (2010). However, we were not able to find strong instruments that could explain agglomerations all around the world.

[27]Heblich et al. (2021) argue that more polluted parts of cities in England were poorer as the rich sorted into less polluted areas. The authors find that those sorting patterns have persisted until today.

[28]Borck and Schrauth (2021) analyse German data and show that historically dense places have no more industrial employment than less dense ones. Since industry was a prime polluter following the industrial revolution, this lends some credibility to the exclusion restriction.

exposure. This kind of heterogeneity is, given that it is time invariant, differenced out in the long-difference estimation.

We also estimate city level regressions, using FUAs as units of observation. This approach compares cities (between city estimates) within countries, but in addition also allows us to look at within-city effects. In within-city grid-level regressions we estimate

$$\ln(E_{GtC}) = \beta + \rho \ln(D_{GtC}) + \alpha_t + \theta_C + \epsilon_{itC}, \tag{3}$$

where C is the city index. We drop all the city-specific control variables since we now control for city fixed effects θ_C and only look at within-city differences.

In addition, we estimate spatial first difference (SFD) models, following the approach proposed by Druckenmiller and Hsiang (2018). This transfers the idea of first differences in time into physical space. In short, SFD regresses the differences in outcomes between neighbouring grid cells on the differences in controls between these same cells. Since the estimation differences out any unobserved factors that are common to neighbouring cells (such as possibly geographic and institutional factors that may be correlated with population density and pollution), this mitigates omitted variable bias. On the downside, some interesting variation is lost by only considering variation between neighbouring cells. Further details are in Appendix C.

3 Results

Before presenting our main outcomes using within-country OLS, we briefly compare OLS and IV coefficients first. We estimate IV regressions in order to gauge the magnitude and direction of potential biases. The reason for not using IV results as our favoured outcomes, as described above, is that we are only able to instrument a small subsample of all observed units, which moreover is primarily restricted to a developed world sample.[29]

To compare OLS and IV results, we harmonize the sample to those cities or grid cells for which the corresponding instrument is available. In all regressions we control for the full set of trade, agricultural, and weather variables as well as logged GDP, ruggedness, latitude and an indicator for the presence of a dirty power plant. We show results with population in 1900 as instrument in Table 2. The first insight is that the sample size is drastically reduced when considering instrumental variables. With raster cells as units of observation, only slightly more than 1000 grid cells of the roughly 200,000 total cells in the whole sample remain. Regarding FUAs, we have historic population for about 10% of all cities. The comparison of OLS and IV results shows only very small and

---

[29]The historical population data by Reba et al. (2016) only covers less than 0.7 percent of all inhabited grid cells in the sample when taking population from 1900 as instrument, and only slightly more than 0.3 percent of all cells when instrumenting with pre-industrial population.

Table 2: IV and OLS regressions

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Raster-level estimations								
log(Sum of population)	0.119***	0.0902***	0.0593***	0.0556***	0.435***	0.424***	0.375***	0.387***
	(0.0219)	(0.0263)	(0.0195)	(0.0168)	(0.0266)	(0.0356)	(0.0328)	(0.0318)
N	1022	1022	1022	1022	1024	1024	1024	1024
R^2	0.588	0.587	0.853	0.853	0.637	0.637	0.793	0.793
Countries	99	99	99	99	99	99	99	99
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel B: FUA-level estimations								
log(Sum of population)	0.0701***	0.0720***	0.0475***	0.0480***	0.294***	0.321***	0.274***	0.286***
	(0.0167)	(0.0199)	(0.0120)	(0.0123)	(0.0202)	(0.0290)	(0.0215)	(0.0243)
N	836	836	836	836	836	836	836	836
R^2	0.656	0.656	0.884	0.884	0.662	0.661	0.801	0.801
Countries	96	96	96	96	96	96	96	96
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV

Note: The table presents coefficients of OLS and IV regressions for both raster-level and city-level (FUA) results. The instrument used is population in 1900. Samples are harmonized such that OLS regressions only include those raster cells/cities for which we have the instrument available. The first two columns of each pollutant show results without country fixed effects, while the latter two of each show results including country FE. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

insignificant differences as soon as we include country fixed effects (columns 3,4,7, and 8). Using population before industrialization as instrument yields similar results (see Appendix Table A.1). The population instruments are exactly identified. Table 3 shows the first stage regressions. The F-statistic indicates that the instruments are strong.

Hence, it seems like omitted variable bias or reverse causality does not cause large biases in the estimates.[30] In the remainder of the paper we focus on within-country regressions using the whole sample available for both grid cells and cities.

3.1 Raster-level outcomes

We first consider raster-level outcomes. Table 4 compares outcomes between simple OLS and within country regressions using the entire sample for both pollutants, NO$_2$ and PM$_{2.5}$. As this will become important in our city-level results, we differentiate between the sum of population within a grid cell and population density of a cell. All specifications include our baseline covariates, i.e. weather, GDP, geographical characteristics,

---

[30]Note, however, that this conclusion holds only for the selective sample of cities where we have long-lagged historical population.

Table 3: First stages with Population instruments

	Population 1900				Population 1650-1750			
	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2	$PM_{2.5}$	NO_2
Population 1900	0.642***	0.624***	0.844***	0.830***				
	(0.0266)	(0.0265)	(0.0253)	(0.0251)				
Population 1650-1750					0.479***	0.464***	0.603***	0.584***
					(0.0605)	(0.0608)	(0.0826)	(0.0826)
N	1022	1024	836	836	357	358	322	322
Countries	99	99	96	96	73	73	70	70
First-stage Statistic	583.5	553.2	1114.4	1096.6	62.76	58.20	53.34	49.91

Note: The table presents first stage results for the population instruments for both raster-level and city-level first stage regressions. Columns 1/2 and 5/6 depict raster-level first stage results, while columns 3/4 and 7/8 show FUA-level coefficients. The source of historical population is Reba et al. (2016). Population 1900 is the population count in year 1900, while population 1650-1750 takes the last population count before 1750 observed in the dataset, which we refer to as pre-industrialized population. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Raster-level OLS regressions

	$PM_{2.5}$				NO_2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.102***		0.0294***		0.191***		0.157***	
	(0.00172)		(0.00149)		(0.00233)		(0.00275)	
log(pop density)		0.0943***		0.0180***		0.184***		0.146***
		(0.00169)		(0.00135)		(0.00179)		(0.00175)
log(GDP)	-0.105***	-0.111***	0.0468***	0.0440***	0.217***	0.213***	0.124***	0.121***
	(0.00541)	(0.00543)	(0.00672)	(0.00673)	(0.00526)	(0.00511)	(0.00876)	(0.00877)
Temperature	0.00521***	0.00672***	0.0249***	0.0265***	0.0657***	0.0672***	0.108***	0.110***
	(0.00107)	(0.00107)	(0.00101)	(0.00101)	(0.00118)	(0.00116)	(0.00151)	(0.00147)
Precipitation	-0.000918***	-0.000975***	-0.000286***	-0.000348***	-0.0000865	-0.000195	-0.00155***	-0.00165***
	(0.000117)	(0.000117)	(0.0000978)	(0.0000975)	(0.000135)	(0.000131)	(0.000143)	(0.000136)
Wind speed	-0.189***	-0.194***	-0.118***	-0.120***	-0.0100**	-0.0134***	0.00931**	0.00831**
	(0.00395)	(0.00398)	(0.00324)	(0.00325)	(0.00440)	(0.00414)	(0.00436)	(0.00393)
1(On coast)	-0.799***	-0.859***	-0.688***	-0.694***	-0.265***	-0.348***	-0.300***	-0.345***
	(0.0180)	(0.0183)	(0.0146)	(0.0147)	(0.0199)	(0.0159)	(0.0192)	(0.0143)
1(Close to lake)	-0.360***	-0.371***	-0.260***	-0.254***	0.198***	0.195***	0.131***	0.136***
	(0.0328)	(0.0335)	(0.0335)	(0.0336)	(0.0372)	(0.0301)	(0.0347)	(0.0273)
1(Close to river)	0.0275	0.0330*	-0.0224*	-0.0160	-0.118***	-0.113***	-0.0593***	-0.0551***
	(0.0172)	(0.0173)	(0.0128)	(0.0128)	(0.0206)	(0.0207)	(0.0180)	(0.0181)
1(Close to dirty powerplant)	0.0790***	0.105***	0.103***	0.131***	0.310***	0.328***	0.226***	0.243***
	(0.0190)	(0.0191)	(0.0139)	(0.0139)	(0.0242)	(0.0242)	(0.0214)	(0.0211)
Ruggedness	-0.00183***	-0.00119*	-0.00324***	-0.00309***	-0.0151***	-0.0143***	-0.00680***	-0.00617***
	(0.000702)	(0.000703)	(0.000658)	(0.000660)	(0.00103)	(0.00101)	(0.000965)	(0.000934)
Latitude	0.00288***	0.00261***	-0.0116***	-0.0116***	0.0596***	0.0580***	0.0686***	0.0668***
	(0.000799)	(0.000805)	(0.000957)	(0.000958)	(0.000855)	(0.000836)	(0.00125)	(0.00124)
N	175237	175235	175237	175235	178535	178507	178535	178507
R^2	0.386	0.378	0.607	0.606	0.604	0.652	0.675	0.732
Countries	161	161	161	161	160	160	160	160
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of OLS regressions including all relevant controls. Apart from those that are listed in the table, the estimations additionally account for all biome indicators, and land suitability. Columns 3,4,7 and 8 additionally include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

suitability for trade and agriculture, and a dummy for whether there is at least one highly polluting power plant within a grid cell. Examining the results shows that the coefficients for both total population and density are reduced in magnitude when we include

country fixed effects to the $PM_{2.5}$-exposure estimations, while the difference between the estimates is smaller for NO_2. In other words, the within-country effect of density on pollution exposure is much smaller than the overall effect. This suggests, for $PM_{2.5}$, that the effect of density is partly driven by certain highly polluted countries with densely populated grid cells. It might be, for instance, that some countries have policies that both limit migration to large cities and pollution.

Taking the within-country estimates, our main results show an elasticity of pollution exposure with respect to density of 0.02 in the $PM_{2.5}$ regressions and 0.15 in the NO_2 regressions. This implies that doubling population density would result in a 1.3 percent increase in $PM_{2.5}$ exposure and a 10.7 percent increase in NO_2 exposure.

Looking at the other coefficients, we find that once we control for country fixed effects, grid cells with higher GDP are more polluted.[31] Pollution exposure rises with temperature. By contrast, precipitation and wind speed are negatively correlated with pollution exposure. Exposure is also strongly affected by dirty power plants.

Using the log of pollution exposure leads to the treatment of all zero observations as missing. To avoid this, we repeat the estimation by replacing all zero values to the minimum non-zero values observed in the data.[32] Interestingly, the pollution-density elasticity for both pollutants becomes significantly higher (0.2 for $PM_{2.5}$ and 0.24 for NO_2 in the regressions with country fixed effects, see Tab. A.2).

That the elasticity is so much lower when we exclude grid cells with zero outcome points to a significant non-linearity in the effect of (log) density on (log) exposure. To address this question from a slightly different angle, we now present non-linear regressions, where we include categorical variables for large and densely populated areas instead of continuous variables. We thus attempt to more directly measure an urban-rural gap. In order to do so, we categorize grid cells with less than 50,000 inhabitants and those with density below 100 persons per sq. km as 'rural'.[33] The results are shown in Table A.3. The urban-rural gap is clearly evident: Going from rural to urban raster cells significantly increases pollution exposure. We redo the exercise with 4 instead of only 2 categories (see Table A.4). There is some variation in the effects; still, we find that the effect of going from what we call 'rural' to 'urban' is larger than the effect of going from one urban category to the next (e.g. from low to moderate density). In summary, there

---

[31]Interestingly, the coefficient on GDP is negative in the $PM_{2.5}$ regression without country fixed effects. This suggests that around the world, grid cells in higher income regions tend to be less exposed to pollution, but this effect is driven by the fact that these grid cells are predominantly located in less polluted high income countries.

[32]This follows Henderson et al. (2018), who use the approach of setting observations with a zero for night lights to the minimum value in the sample in their estimates presented as main results.

[33]This follows the density category 'extremely low' in
https://www.yourarticlelibrary.com/population/population-density-classification-of-the-spatial-distribution-of-population-density/19853.

is an urban-rural gap in pollution exposure which trumps the effect of increasing density within urban areas.

In the remainder of the raster-level results, we will mainly report within-country effects and focus on population density, which makes results more comparable to previous literature. The main estimates in Table 4 assume a homogeneous relationship between pollution and population in the entire world. In order to check whether this relationship changes with geography, country income, and the like, we now consider various interactions to analyse the heterogeneity of this effect.

Heterogeneity and robustness. We now turn to analysing heterogeneities in the pollution-density gradient across continents and countries at different stages of development. Figure 4 shows the results of running within-country regressions by country income groups and by continents, where income groups follow the World Bank classification into low, lower middle, upper middle, and high income. Population density is a significant determinant of pollution over all income groups and continents, but to a different extent. Figure 4a exhibits that the strongest within-country effects of density are found in low middle income countries regarding $PM_{2.5}$-exposure and in upper middle income countries for NO_2. Hence, it seems like the density effect is to some extent non-linear in income, and middle-income countries tend to have a stronger effect of density on pollution than both low and high-income countries. A potential reason is that density is not "dirty" (Carozzi and Roth, 2022) in low income countries, because there is little dirty activity such as driving and heating, whereas in high income cities, cleaner transport modes (e.g. public transport) and residential energy use (heating and cooking with electricity or "modern fuels") may mitigate the effects of density.[34] In contrast, middle income country agglomerations may be dirtier than low income ones because there is more driving and residential energy use, but technologies for these activities are not as clean as in high income country cities.

Fig. 4b shows differences by continent. For NO_2, the density effect is smallest in Africa and largest in Asia. In Fig. 5 and 6, we further show the density coefficients for each country from individual within-country regressions on world maps. The maps show some interesting ramifications. The NO_2 density elasticities in Fig. 5 suggest that China and India – where most of the biggest and most polluted cities in the world are located – seem to have the strongest effect of density on NO_2 pollution exposure in Asia. In North America, Mexico and the US have larger effects than Canada, while within Europe countries from the south seem to have higher elasticities than countries in the north.

---

[34]See Borck and Mulder (2022) for a model with dirty and clean energy use as well as transport modes in developing countries.

Figure 4: Heterogeneity of density effect by subgroup
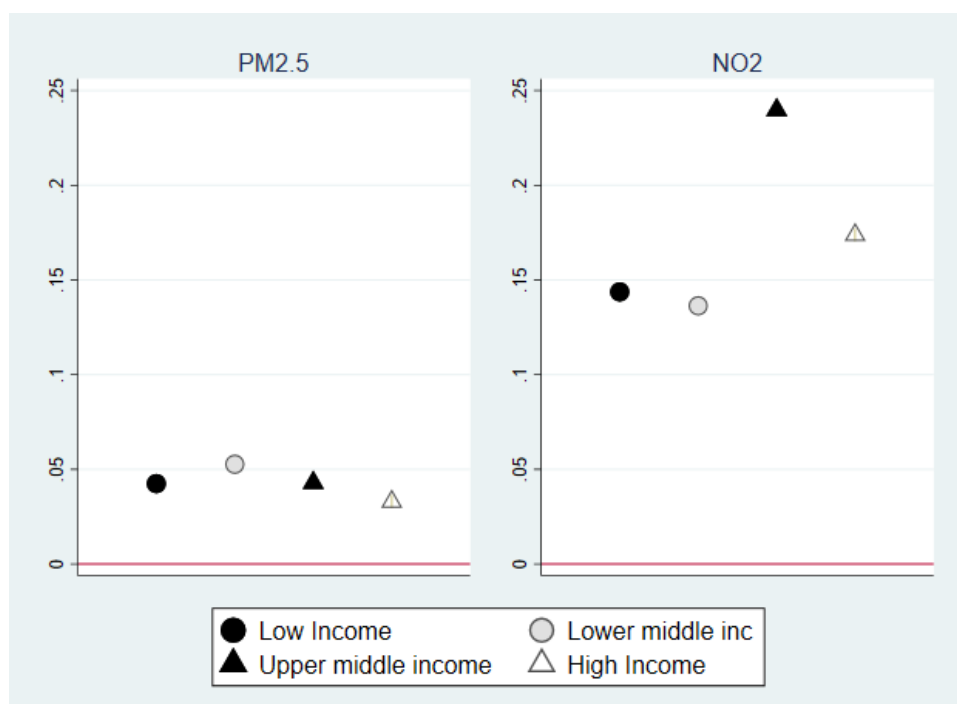

(a) By income group
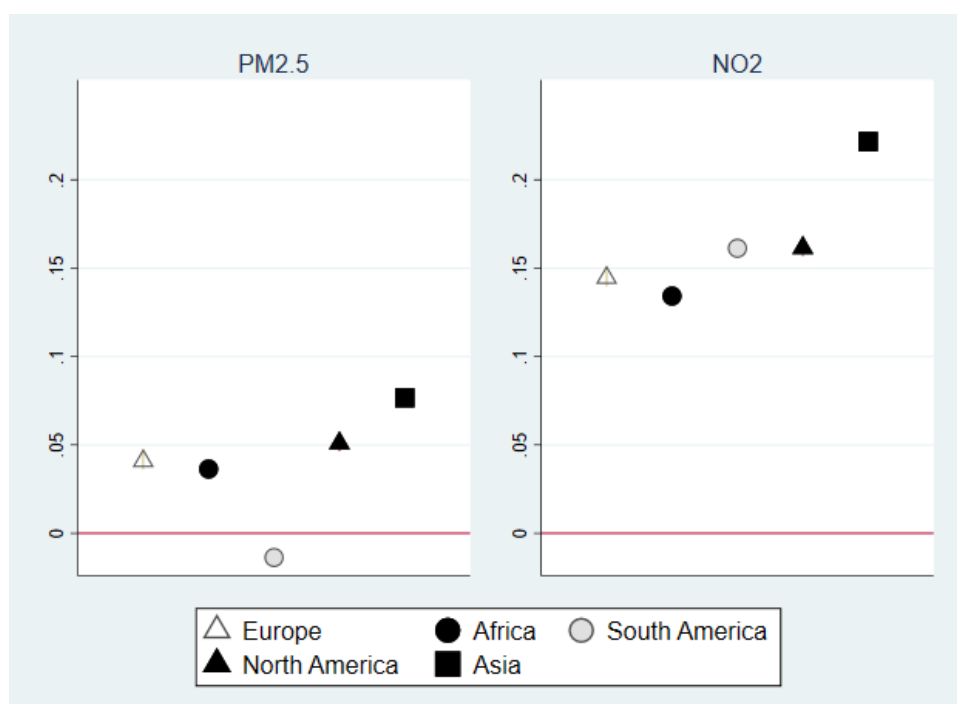
(b) By continent

Note: The graphs show coefficients of separate regressions for each of the respective subgroups. The $y-$axes represent the coefficient sizes. All estimations include within-country fixed effects, and all the control variables from our standard estimations (compare Table 4). Income group definitions are taken from the World Bank.

Fig. 6 shows the country-specific density elasticities for $PM_{2.5}$. In Asia and Europe, most of the countries that have large density elasticities for NO_2 also have large elasticities for $PM_{2.5}$, while the density effect in the Americas seems to be smaller for $PM_{2.5}$ than for NO_2.

Figure 5: Density coefficients for NO_2 by country
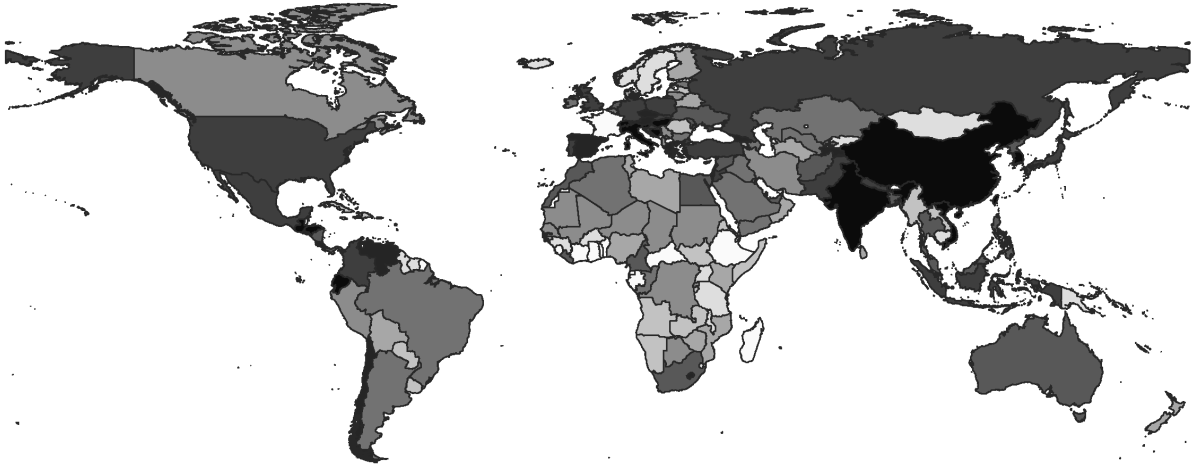
Note: The map shows a graphical representation of the effect of population density on NO_2, where regressions are performed for each country individually. Dark colors represent coefficients at the upper end of the density-pollution gradient, light colors show those at the lower end. All relevant control variables are included in the estimations.

Figure 6: Density coefficients for $PM_{2.5}$ by country
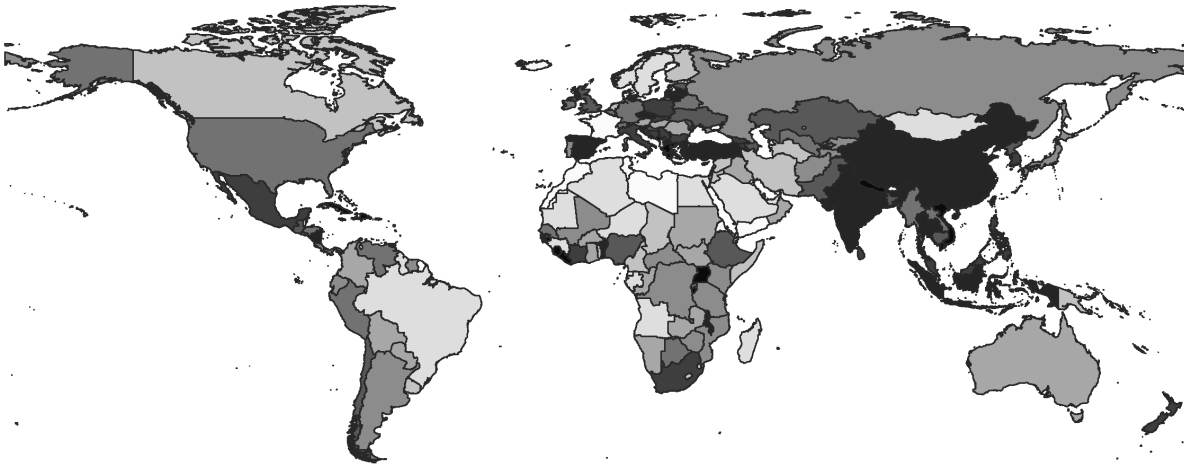
Note: The map shows a graphical representation of the effect of population density on $PM_{2.5}$, where regressions are performed for each country individually. Dark colours represent coefficients at the upper end of the density-pollution gradient, light colours show those at the lower end. All relevant control variables are included in the estimations.

The Landscan data distribute population to grid cells using certain grid characteristics, but the exact algorithm is not known. While Henderson et al. (2021) provide a ground-truthing exercise, it might still be the case that the data generation biases the results. As a robustness check, we therefore run the regressions using administrative areas as units of observations[35]. In these data, where population is smoothly distributed among all grid cells within an administrative unit.[36] We present the results in Table A.5. Again, they hardly differ from our previous results.[37]

Lastly, we run long-difference regressions of pollution changes between the last and first year in our sample on population or density changes in the same period. Thus, we control for any time-invariant unobserved heterogeneity between administrative units that might affect both population and pollution. For instance, it might be that within countries, population sorting leads to residents of dense cities being 'greener' on average, which would bias our estimates (downwards in this case). Analysing long differences within rasters differences out these time invariant unobserved heterogeneities. Results are shown in Table A.6.[38] In the long-difference estimates, all the variables that are time-constant drop out, so the only explanatory variable left besides the population data is GDP. As the Table shows, the long-difference estimates again show a positive and significant effect of both population and density on both pollutant-exposure measures. The magnitudes are now reversed, however: it seems that population changes now affect NO_2-exposure more strongly than $PM_{2.5}$. A potential explanation is that the variation in the cross section as well as over time is much lower for NO_2 than for $PM_{2.5}$. This implies that a given change in population over time affects changes in NO_2 less than changes in $PM_{2.5}$.

Channels. An interesting question in interpreting the findings is what mechanisms could be responsible for the observed relationship between density and pollution exposure (see also Borck and Schrauth, 2021; Carozzi and Roth, 2022). While a complete investigation is made difficult by the scarcity of available data at a worldwide scale, we nonetheless try to shed some light on these channels here. We follow the analysis in Borck and Schrauth (2021) and leave out some sets of explanatory variables. We then compare the density coefficient with and without these variables. The direction of change

---

[35]We use gridded population of the world (GPWv4) data on administrative (GADM) level, see `https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11`.

[36]This obviously introduces other biases. Nonetheless, it is reassuring to find that the results do not seem to be driven by biases in the computation of grid specific density.

[37]Using the gridded GPW data without aggregating it to GADM level does not change the results either (results not shown here).

[38]These estimates are also based on the GPWv4 data on the GADM (administrative unit) level. The reason is that the quality of LandScan data significantly improved over time and therefore comparisons over time should not be made, as stated by the data provider itself (see https://gistlandscan01.ornl.gov/frequently-asked-questions).

of the coefficient then allows us to determine how these variables affect pollution directly and indirectly through their correlation with density.

We report the results of leaving out, one by one, different groups of our explanatory variables in Tab. A.7. Column (1) shows the baseline regression results, col. (2) leaves out GDP, (3) the weather variables, (4) the trade variables (closeness to river, lake or coast) and (5) the agricultural variables (land suitability and biomes).

As can be seen in the table, the density coefficient rises in all columns compared to the baseline. For both pollutants, we find the largest increase when we leave out weather and agricultural suitability. This seems to imply that the density effect is driven most by the fact that dense areas are located, on average, in areas that have weather that is conducive to high pollution exposure (such as hot, dry areas with little wind). Moreover, dense areas seem, on average, to be located in places with a first nature that is advantageous to agriculture, which tends to increase pollution. Higher income and suitability for trade apparently drive the density effect to a lesser extent.

Figure 7: Distribution of country-specific density coefficients

(a) PM$_{2.5}$

(b) NO$_2$
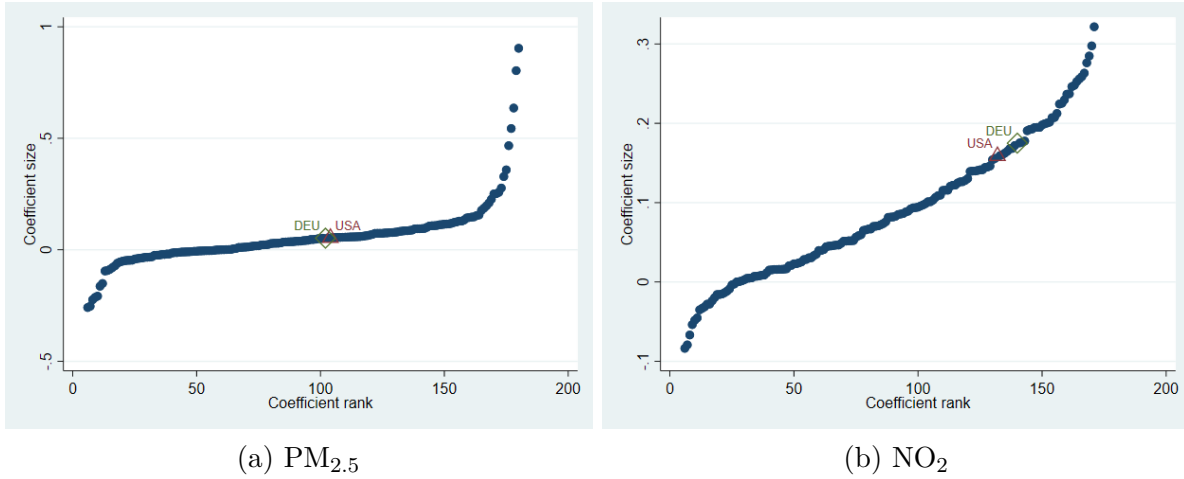
Note: The graphs show the distribution of the population density / pollution exposure gradient, when running estimations for each country in the world separately. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. The green hollow diamond is the coefficient for Germany (DEU), and the red hollow triangle the one for the United States (USA). These are highlighted in order to compare them with findings by Borck and Schrauth (2021) and Carozzi and Roth (2022) respectively. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

Raster-level outcomes and country characteristics.   We now look at how the pollution-density relation changes with country characteristics to get additional insights into its determinants. Figure 7 ranks all country-specific coefficients and plots them by

their size.[39] All coefficients plotted come from regressions including control variables as presented in Table 4. The green hollow diamond and the red hollow triangle show the coefficients for Germany (DEU) and the US, in order to compare them to prior papers in the field (Carozzi and Roth, 2022; Borck and Schrauth, 2021). Both coefficients are somewhat smaller than the ones in those papers. Both Carozzi and Roth (2022) and Borck and Schrauth (2021) run extensive tests to more credibly estimate a causal effect within one country; however, the samples differ from the one used here. In particular, Borck and Schrauth (2021) study German counties (*Kreise*) and Carozzi and Roth (2022) US CBSA. This difference notwithstanding, we think it is reassuring to see that the magnitude of the coefficients is roughly in line with previously estimated ones from studies that are better able to address causality issues than we are.[40]

About 70% of $PM_{2.5}$ coefficients and 75% of NO_2 coefficients lie in a range between 0 and .3. About 82% of $PM_{2.5}$ coefficients and about 85% of NO_2 coefficients are positive, where about half of the cases with negative coefficients have negative ones for both pollutants. It is interesting to briefly look at the outliers, i.e. countries with the 2.5 percent highest and lowest (negative) coefficients. In most instances, these turn out to be small island states such as Jamaica, Malta, East Timor (downward) or Bahamas, Barbados, Cap Verde (upwards).
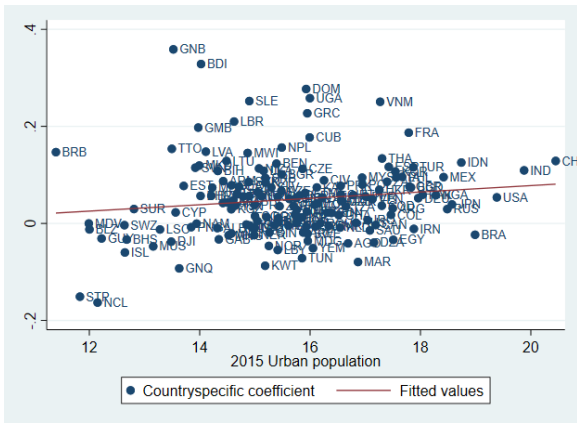
Figure 8 shows simple scatter plots between the country-specific density elasticity estimates of our within-country regressions and urbanization patterns (where, again, the density coefficients stem from regressions with all basic controls described above). In general, most of the correlations seem insignificant for $PM_{2.5}$, while we do find some interesting correlations for NO_2. As the figure shows, the more people live in urban areas, the stronger is the density effect on pollution, while the effect of the urbanization rate is also positive but somewhat weaker. This suggests that density is more likely to increase NO_2 pollution when many people live in cities, which underlines the non-linear effects described above. It also links the paper's results to the theory of city systems; indeed it seems like total exposure will be reduced by shifting individuals from denser to less dense regions (Borck and Tabuchi, 2019). Moreover and interestingly, the density coefficients are negatively correlated with renewable energy use (figure not shown). Intuitively, when energy use is relatively clean, packing residents densely together does not produce as much pollution as when countries rely largely on fossil fuels.

In the next subsection, we examine regression results when we explicitly consider cities defined as functional urban areas.

---

[39]We excluded the lowest and the highest 2.5% of the sample as outliers, thus removing 5 percent of country coefficients.

[40]See also Ahlfeldt and Pietrostefani (2019).

Figure 8: Country-specific density effects and urbanization

(a) Urban Population total (PM$_{2.5}$)

(b) Urban Population total (NO$_2$)

(c) Urbanization rate (PM$_{2.5}$)

(d) Urbanization rate (NO$_2$)

(e) Large Agglomerations (PM$_{2.5}$)

(f) Large Agglomerations (NO$_2$)

Note: Scatter plots of the country-specific population density effect on pollution exposure correlated with different World Bank indicators as specified by each subtitle. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

3.2 City-level outcomes

We now present results from regressions on the FUA sample. We use different variables to analyse the effect of population on pollution: (i) logged mean population density within a city polygon, and (ii) the logged total population. In addition, we can differentiate between the core city population (log(Pop urban centre)) and that of the surrounding commuting area (log(Pop commuting)). This allows us to move in the direction of considering mechanisms for the relation we study. Sprawling cities with many commuters and single family homes might have different pollution levels compared to dense cities with high-rise buildings and without much long-distance commuting.

Table 5: Within-country regressions for FUA

	$PM_{2.5}$			NO_2		
	(1)	(2)	(3)	(4)	(5)	(6)
log(pop density)	-0.00435			0.00227		
	(0.0134)			(0.0102)		
log(Sum of population)		0.0791***			0.0947***	
		(0.00512)			(0.00535)	
log(Pop urban center)			0.00963			0.0402***
			(0.00790)			(0.0109)
log(Pop commuting)			0.0752***			0.112***
			(0.00725)			(0.00815)
log(GDP)	0.0547***	0.0397**	0.0362**	0.331***	0.312***	0.313***
	(0.0163)	(0.0161)	(0.0159)	(0.0242)	(0.0236)	(0.0239)
Temperature	0.0130***	0.0148***	0.0151***	0.000608	0.00355	0.00275
	(0.00278)	(0.00273)	(0.00268)	(0.00389)	(0.00380)	(0.00387)
Precipitation	-0.000334	-0.000294	-0.000274	-0.00208***	-0.00204***	-0.00205***
	(0.000208)	(0.000206)	(0.000199)	(0.000162)	(0.000161)	(0.000169)
Wind speed	-0.0697***	-0.0679***	-0.0649***	-0.00759	-0.00511	-0.00472
	(0.00799)	(0.00788)	(0.00766)	(0.00950)	(0.00935)	(0.00973)
1(On coast)	-0.384***	-0.429***	-0.437***	0.0457**	-0.00609	-0.0175
	(0.0225)	(0.0226)	(0.0221)	(0.0231)	(0.0225)	(0.0225)
1(Close to lake)	-0.265***	-0.282***	-0.259***	0.192***	0.179***	0.179***
	(0.0701)	(0.0693)	(0.0642)	(0.0591)	(0.0570)	(0.0600)
1(Close to river)	0.0560***	0.0362**	0.0363***	0.0621***	0.0371**	0.0277
	(0.0146)	(0.0145)	(0.0140)	(0.0173)	(0.0170)	(0.0178)
1(Close to dirty powerplant)	0.140***	0.0773***	0.0663***	0.502***	0.429***	0.391***
	(0.0138)	(0.0139)	(0.0136)	(0.0207)	(0.0205)	(0.0206)
Ruggedness	-0.000298***	-0.000280***	-0.000286***	-0.000511***	-0.000486***	-0.000532***
	(0.0000287)	(0.0000281)	(0.0000287)	(0.0000354)	(0.0000354)	(0.0000378)
Latitude	0.0126***	0.0134***	0.0138***	0.00918***	0.0109***	0.0103***
	(0.00178)	(0.00176)	(0.00175)	(0.00248)	(0.00245)	(0.00249)
N	8871	8871	8249	8861	8861	8219
R^2	0.688	0.699	0.726	0.786	0.795	0.790
Countries	136	136	136	134	134	134
No. of cities	8871	8871	8249	8861	8861	8219
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents coefficients of OLS regressions with FUA as units of observations. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude log(GDP), and country-fixed effects (Country FE). Standard errors are robust. t statistics are in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5 compares the effects of population density with those of total city population and additionally differentiates between population in the urban centre in commuting zones. Interestingly, for both pollutants, the effect of population density is not significant, while total population positively affects pollution. The coefficients indicate that a 1% increase in total population increases $PM_{2.5}$ exposure by 0.08 percent and NO_2 exposure by 0.95 percent. Compared to raster-level results, the population coefficient is larger for $PM_{2.5}$ and smaller for NO_2. Hence, it seems that density *per se* does not drive higher pollution exposure; rather, the relation seems to be driven by the way that large populations are organized spatially within cities.

To elaborate on this theme, in columns (3) and (6), we distinguish between core city and commuting population. For $PM_{2.5}$, we find that the coefficient of core city population is insignificant, while that on commuting population is positive and significant. For NO_2, both are significant, but the coefficient on commuting population is about three times as large. Consequently, it seems that large cities *per se* are not more polluted than smaller ones. Rather, pollution seems to be significantly higher in cities with a large fraction of people commuting into the city from satellite cities. These findings thus shed more light on the link between density, population distribution and pollution exposure. It seems that, when we only look at cities, large and dense development does not need to be bad for the environment. This may be due to the fact that these urban features promote the use of clean public transport and energy efficient buildings, which may partly offset the increased pollution exposure stemming from a high concentration of polluting activities.[41]

Figures 9 and 10 again show the distribution of coefficients by continents and income groups. For $PM_{2.5}$, there is no clear trend by income. For NO_2, however, the population effect is strongest among upper middle and high income countries. The upper panel of Fig. 10 shows that for $PM_{2.5}$, there are no pronounced differences in the density effect between continents. For NO_2, the population and density effects are lowest in Africa and Asia. The effect of commuting population is also lowest in Africa.

City level outcomes and country characteristics. Just as for the raster-level results, we repeat the exercise of relating the country-specific population-pollution coefficients to country characteristics, such as urbanization rates and income. We again present scatter plots, where each point shows the country-specific coefficient of population on pollution exposure. Figure 11 shows the plots. The correlations for NO_2 again seem to be stronger than for $PM_{2.5}$. We find that agglomeration seems to affect pollution more the higher the urbanization rate and the share of population living in large agglomerations of more than 1 million people.

---

[41]See Brownstone and Golob (2009) for the effect of density on driving, Borck and Brueckner (2018) for the link between density and energy efficiency, and Carozzi and Roth (2022) on the interpretation of higher exposure due to the density of economic activity.

Figure 9: Population/density effect by income
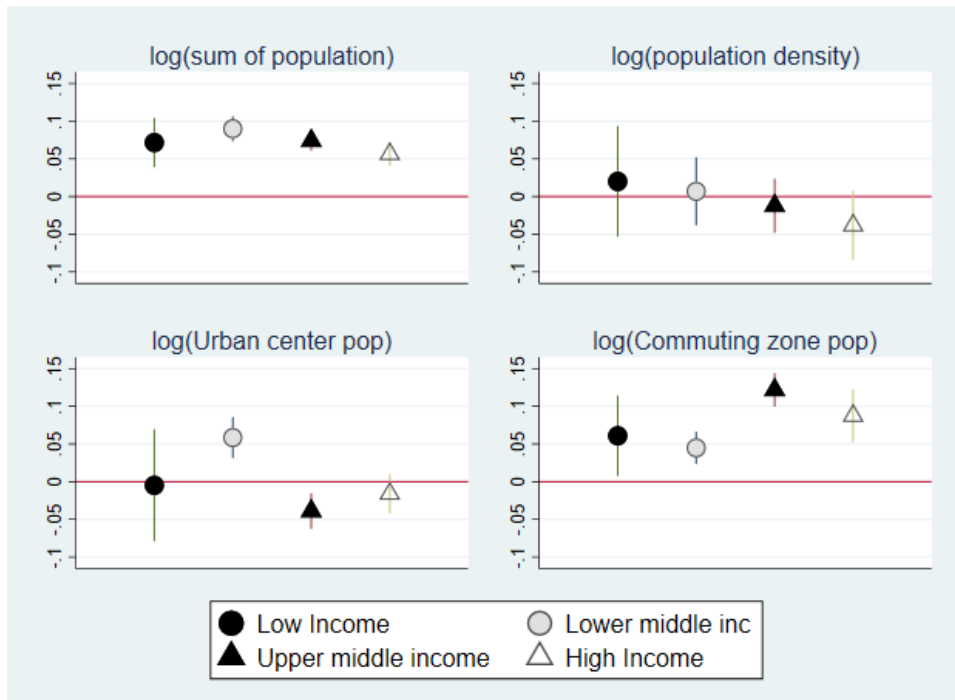
(a) PM$_{2.5}$
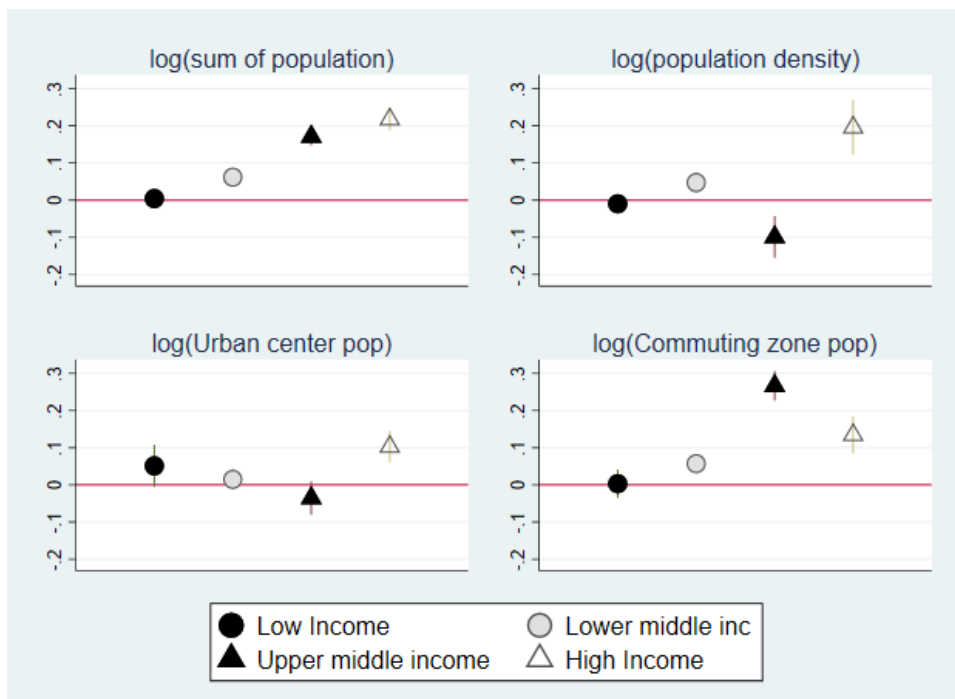
(b) NO$_2$

Note: Coefficients of different population measures on pollution exposure by subgroups. The coefficients are obtained from regressions for each subgroup separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby.

29

Figure 10: Population/density effect by continent
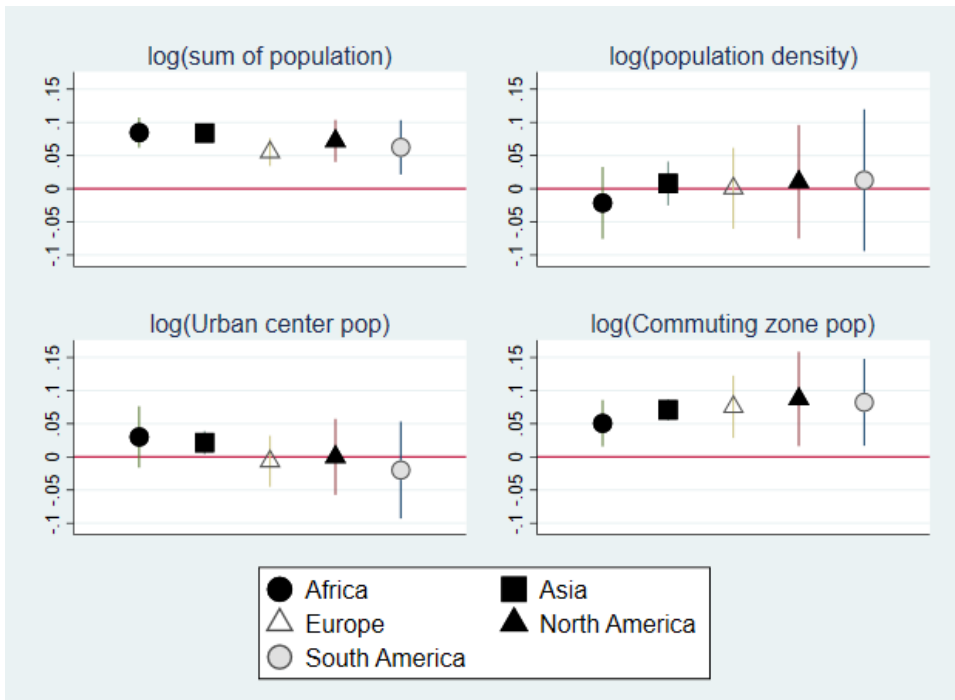
(a) PM$_{2.5}$

(b) NO$_2$

Note: Coefficients of different population measures on pollution exposure by subgroups. The coefficients are obtained from regressions for each subgroup separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby.

30

Figure 11: Country-specific population effects and urbanization

(a) Urban Population total (PM$_{2.5}$)

(b) Urban Population total (NO$_2$)

(c) Urbanization rate (PM$_{2.5}$)

(d) Urbanization rate (NO$_2$)

(e) Large Agglomerations (PM$_{2.5}$)

(f) Large Agglomerations (NO$_2$)

Note: Scatter plots of the country-specific population density effect on pollution exposure correlated with different World Bank indicators as specified by each subtitle. The coefficients are obtained from regressions for each country separately, controlling for trade variables (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and an indicator for a dirty power plant nearby. Outliers (2.5% highest and lowest coefficients) are excluded from the graphical representation.

3.2.1 Local level outcomes

Table 6: Within City regressions

	PM$_{2.5}$	NO$_2$
	(1)	(2)
log(pop density)	0.0215***	0.0425***
	(0.000351)	(0.000839)
N	66118	66127
R^2	0.995	0.971
Countries	187	175
City FE	Yes	Yes

Note: The table presents coefficients of within-FUA-effects. Each pollution cell is matched to a population cell. Since the within-city-fixed effects capture all relevant control variables we use in other tables, no additional variables are controlled for.
Standard errors are robust. t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Within city regressions. Tab. 6 shows the results of within-city regressions, that is, we compare raster cells that lie within the same FUA. Since we control for city fixed effects, all FUA-level variables are absorbed by them, so there are no additional control variables. The table shows that the population effect for PM$_{2.5}$ is about the same as in the baseline raster results. For NO$_2$, however, the coefficient is much lower. This may be due to the smaller variation within cities, or the fact that densely populated raster cells tend to lie in polluted cities. Beyond that, the smaller variation within cities for NO$_2$ is somewhat mechanically caused by the larger size of the grid cells.

Figure 12 shows differences of results by income groups and by continents. Within cities, the pollution-population gradient steadily increases with income. Hence, different from what we observed before, higher income countries seem to have especially strong pollution in densely populated areas within cities. Looking at differences by continents, American cities exhibit the highest and African cities the lowest pollution-population gradient.

Spatial First Differences. Tab. A.8 in the Appendix shows the results from spatial first differences (SFD) regressions. Again, we find that for both pollutants, the results remain positive and significant. This result is reassuring. The SFD estimate differences out any unobserved heterogeneity that is common to neighbouring cells. The result thus further lends credence to the relationship between density and pollution exposure at a local level.

However, the effect of density on pollution is much smaller when we consider differences between neighbouring cells than when we compare the entire sample, especially for NO$_2$. Intuitively, the variation in pollution exposure and density is much smaller

Figure 12: Population effect within city by sub-sample

(a) PM$_{2.5}$ by income

(b) NO$_2$ by income

(c) PM$_{2.5}$ by continent

(d) NO$_2$2 by continent
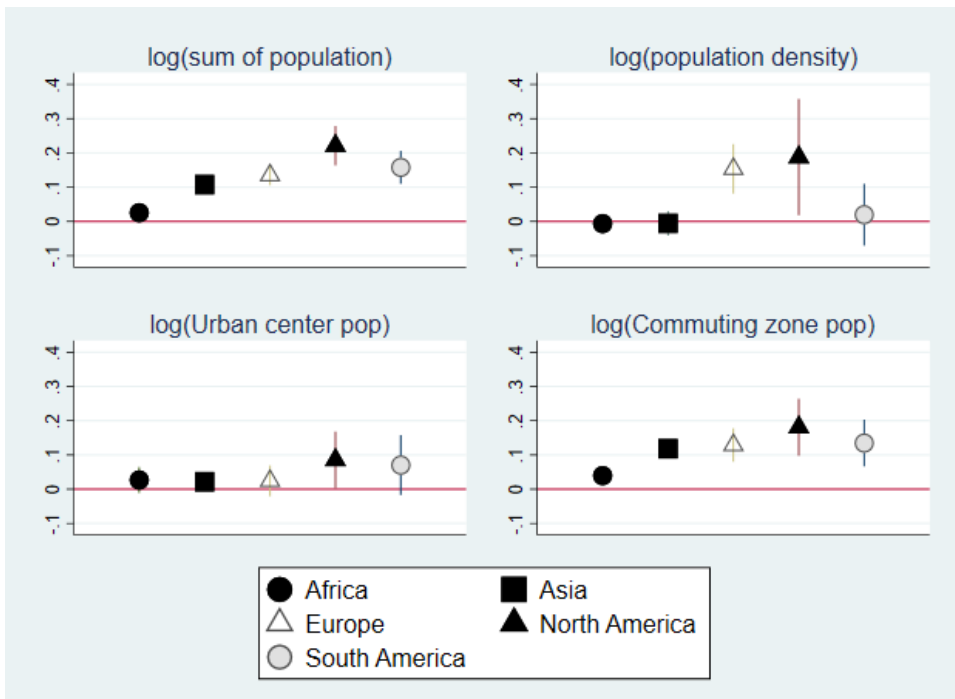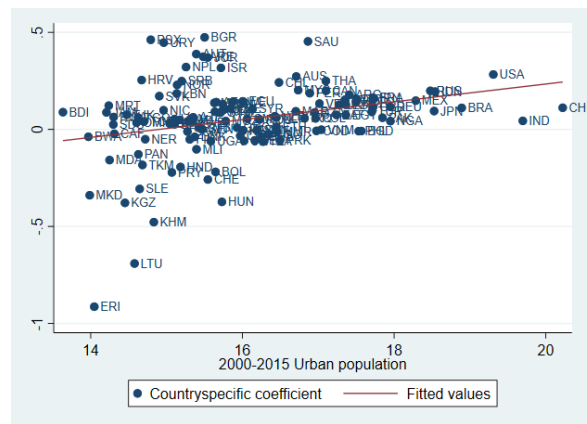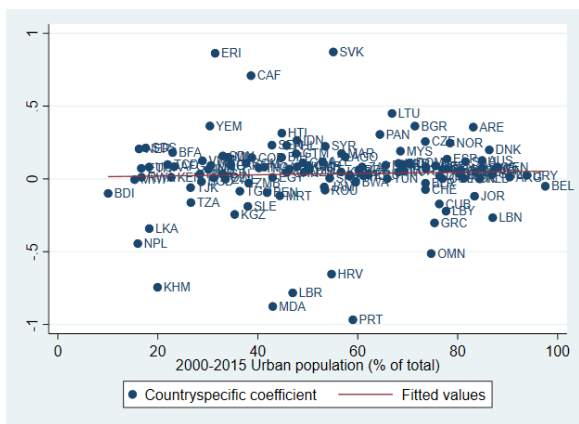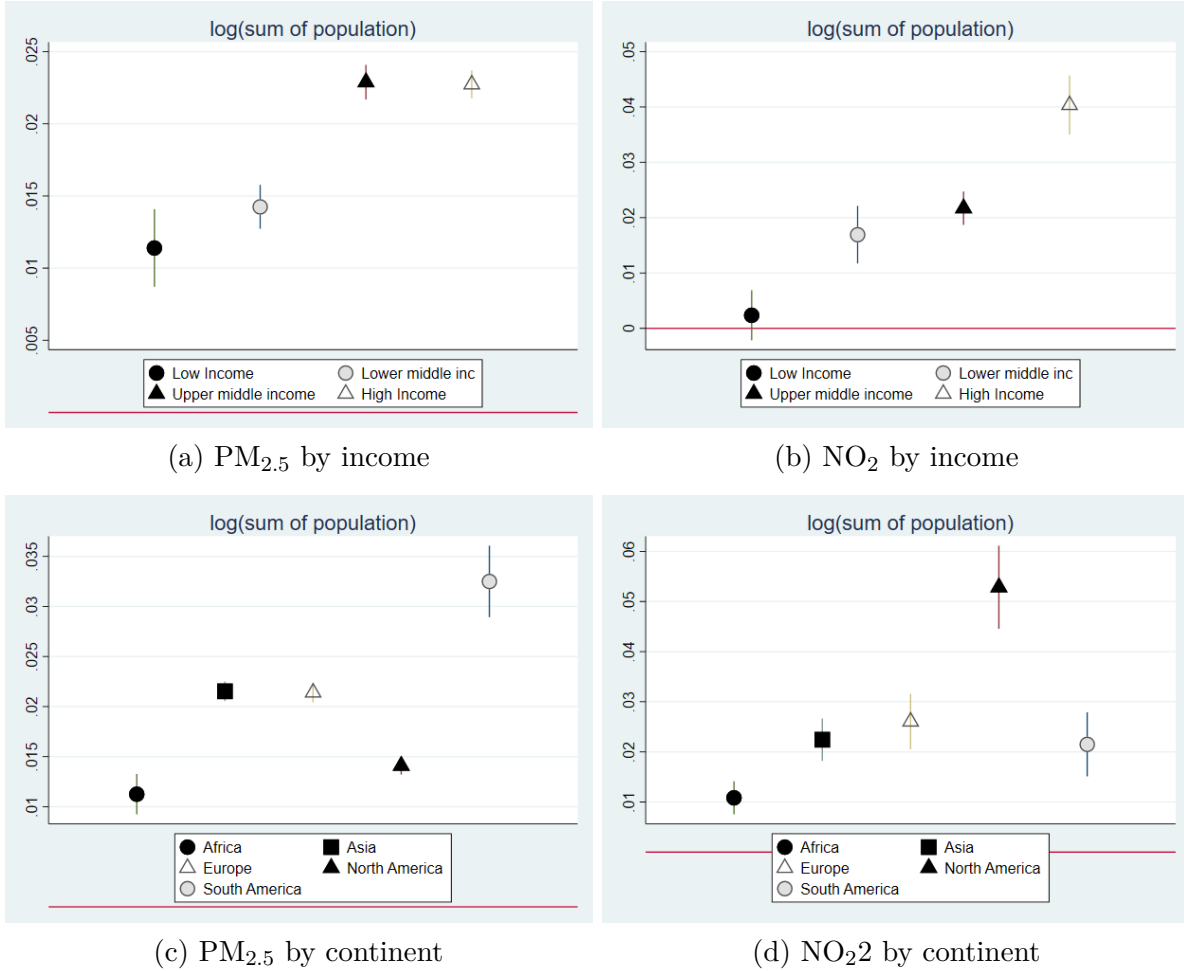
Note: Coefficients of population on pollution exposure by subgroups within cities. The coefficients are obtained from regressions for each subgroup separately. The results stem from within-city estimations, comparing pollution exposure of small grid cells within cities.

between neighbouring cells than in the entire sample, which likely explains the smaller effects.

4   Counterfactual simulation

In order to quantify the effects of the population exposure relation by country, we present results from a counterfactual simulation in this section, where we compute the country-specific change in exposure from equalizing population across all cities in a country. We do this counterfactual with the FUA sample. We thus assume that a country's population is given by its population living in FUAs. The counterfactual then answers the question: what would be the effect on total exposure if all cities had the same population?

Let country S have M_S cities with population N_i and total population $N_S = \sum_{j=1}^{M_S} n_i$. We want to compute the effect of redistributing population equally among cities within a country. Consequently, all cities have identical counterfactual population $N_i' = N_S/M_S$. From our estimates, we predict current total exposure for city i as $\tilde{E}_i = N_i^{1+\rho}$, where ρ is the estimated exposure-population elasticity. Total exposure is then $E = \sum_j \tilde{E}_j$.[42]

Now consider a counterfactual where city i's population is changed to $N_i' = N_S/M_S$, so all cities are equally large. The counterfactual exposure in city i is $E_i' = N_i'^{1+\rho} = (N_S/M_S)^{1+\rho}$. The proportional counterfactual change in city i is $\hat{E}_i = E_i'/\tilde{E}_i = \hat{N}_i^{1+\mu}$, where $\hat{N}_i = \frac{N/M_S}{N_i}$ is the proportional change in population.

Total counterfactual exposure in the country is $E' = \sum_j E_j'$. We can then compute the percentage change in exposure, $\Delta E = \hat{E} - 1 = E'/E - 1$.

We estimate ρ for all countries with at least 15 FUAs in our sample. We report the estimates of ρ along with total population, number of cities and the counterfactual change in exposure, ΔE, for the 10 countries with the smallest and largest change in Tab. 7 for $PM_{2.5}$ and in Tab. 8 for NO_2.

Since total exposure is convex in population if and only if $\rho > 0$, all countries with a positive estimated population elasticity would benefit from a reduction in total exposure induced by population smoothing. It is apparent from Tab. 7 and 8 that the countries with the largest percentage drop in total exposure are those with the largest population elasticity. There are, however, some differences in the composition of countries. For $PM_{2.5}$, we see the largest drops in total exposure in countries in East Asia/Pacific and Sub-Saharan Africa, plus two in Latin America/Caribbean. The countries with the largest increase in exposure (where the population elasticity is negative) tend to be lower income countries in Latin America/Caribbean and Sub-Saharan Africa, plus two in Middle East/North Africa.

For NO_2, the countries with the largest percentage drop in total exposure tend to be in East Asia/Pacific and Latin America/Caribbean, while the ones with increases in exposure are almost all in Sub-Saharan Africa, with the exception of North Korea.

5 Conclusion

This paper has studied the effect of population and population density on pollution exposure using worldwide gridded data. We find that population density increases exposure. Using city-level data, we find that population size, rather than density, increases exposure. Further, the reason seems not to be a large core city population, but rather a large

---

[42]Note that we do not use our usual control variables in these estimates, so their influence is subsumed in the effect of city population. Redoing the counterfactual with controls does not have a large effect on the results.

Table 7: Counterfactual change in exposure for PM$_{2.5}$

Country	Population (mill.)	# cities	ρ	ΔE	Rank
	A: 10 lowest ranked				
Indonesia	139.491	249	0.240	-36.494	1
Haiti	4.024	21	0.215	-31.816	2
Peru	16.998	41	0.194	-30.862	3
Nigeria	87.799	351	0.182	-30.166	4
Togo	2.953	15	0.205	-27.798	5
Vietnam	38.426	99	0.174	-27.086	6
D.R. Congo	23.493	125	0.159	-26.751	7
Malaysia	19.432	31	0.240	-22.821	8
Sudan	11.219	72	0.131	-22.080	9
Philippines	46.653	67	0.123	-21.646	10
	B: 10 highest ranked				
Colombia	28.753	85	-0.020	2.932	67
Tunisia	4.824	24	-0.035	3.586	68
Ecuador	9.361	29	-0.057	5.551	69
Yemen	6.285	23	-0.073	7.622	70
Venezuela	21.008	65	-0.129	9.436	71
Mozambique	5.143	48	-0.067	9.994	72
Uganda	5.300	21	-0.070	10.360	73
Angola	11.281	46	-0.072	13.524	74
Tanzania	10.234	40	-0.109	14.669	75
Senegal	6.291	29	-0.188	22.535	76

Note: ρ is the estimated within-country population elasticity of exposure. ΔE is the percentage change in exposure in the counterfactual relative to the baseline.

population commuting into the core city. Lastly, we find positive but smaller effects of density on pollution exposure at more local levels.

We also document heterogeneities of the density effects across countries. Using the entire rasterized data as observational units, the influence of population seems largest in Asia and in middle-income countries. In the FUA sample, population affects pollution most in upper middle and high income countries as well as in Europe and North America.

Finally, we study how reallocating population among cities within countries affects exposure. For most countries, exposure would fall if population were equalized across cities, since total city exposure is convex in city population. This allows us to connect the paper to the literature on optimal city size (Borck and Tabuchi, 2019). Using country specific exposure-population elasticities, we could in principle study how the distribution of optimal city size is determined by the trade-off between agglomeration benefits and costs, stemming from the increase in exposure.

Table 8: Counterfactual change in exposure for NO_2

Country	Population (mill.)	# cities	ρ	ΔE	Rank
A: 10 lowest ranked					
Peru	16.438	41	0.365	-53.175	1
China	567.218	1539	0.408	-46.429	2
Thailand	16.893	37	0.292	-45.821	3
Japan	105.882	88	0.298	-44.340	4
Guatemala	5.372	34	0.302	-43.668	5
Indonesia	133.417	249	0.269	-40.361	6
Iran	50.591	165	0.277	-38.800	7
Argentina	27.907	64	0.230	-37.419	8
Chile	12.235	31	0.275	-36.070	9
Australia	15.575	20	0.419	-34.052	10
B: 10 highest ranked					
Guinea	2.659	17	-0.024	3.059	66
Benin	2.934	21	-0.040	4.122	67
D.R. Congo	18.242	125	-0.023	4.476	68
Togo	1.924	15	-0.045	5.078	69
Burkina Faso	2.689	32	-0.031	5.558	70
Uganda	4.291	21	-0.042	6.011	71
North Korea	10.764	80	-0.073	6.400	72
Tanzania	7.210	40	-0.053	6.429	73
Cote d'Ivoire	6.953	35	-0.076	12.491	74
Ghana	8.090	51	-0.137	26.080	75

Note: ρ is the estimated within-country population elasticity of exposure. ΔE is the percentage change in exposure in the counterfactual relative to the baseline.

Some avenues for future research suggest themselves. First, it would obviously be of interest to study how differences in the population-pollution elasticities between countries are shaped by institutional determinants, urban structure, and other factors. Second, although we have tried to come close to estimating causal effects, we think it would be fruitful to merge this more descriptive global evidence with the more causal national evidence as in Borck and Schrauth (2021) and Carozzi and Roth (2022). As more data becomes available for more countries and longer time periods, more robust evidence on the agglomeration costs of pollution will certainly be forthcoming.

References

Ahlfeldt, G. M. and E. Pietrostefani (2019). The economic effects of density: A synthesis. *Journal of Urban Economics 111*, 93 – 107.

Aldeco, L., L. Barrage, and M. A. Turner (2019). Equilibrium particulate exposure. Working paper, International Growth Center.

Atmospheric Composition Analysis Group (2018). Surface no2 and surface pm 2.5 worldwide. `https://sites.wustl.edu/acag/datasets/`. Atmospheric Composition Analysis Group, Dalhousi University. Last access: November 2022.

Borck, R. and J. K. Brueckner (2018). Optimal energy taxation in cities. *Journal of the Association of Environmental and Resource Economists 5*(2), 481–516.

Borck, R. and P. Mulder (2022). Energy policies and pollution in developing country cities. Unpublished manudscript.

Borck, R. and M. Pflüger (2019). Green cities? Urbanization, trade, and the environment. *Journal of Regional Science 59*, 743–766.

Borck, R. and P. Schrauth (2021). Population density and urban air quality. *Regional Science and Urban Economics 86*, 103596.

Borck, R. and T. Tabuchi (2019). Pollution and city size: can cities be too small? *Journal of Economic Geography 19*(5), 995–1020.

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

Brueckner, J. K. (1987). Chapter 20 the structure of urban equilibria: A unified treatment of the muth-mills model. In *Urban Economics*, Volume 2 of *Handbook of Regional and Urban Economics*, pp. 821–845. Elsevier.

Carozzi, F. and S. Roth (2022). Dirty density: Air quality and the density of American cities. *Journal of Environmental Economics and Management*. Forthcoming.

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

Chen, S., P. Oliva, and P. Zhang (2022). The effect of air pollution on migration: Evidence from China. *Journal of Development Economics 156*, 102833.

Cheshire, P. and S. Magrini (2008, 10). Urban growth drivers in a Europe of sticky people and implicit boundaries. *Journal of Economic Geography 9*(1), 85–115.

Ciccone, A. and R. E. Hall (1996, March). Productivity and the Density of Economic Activity. *American Economic Review 86*(1), 54–70.

CIESIN Columbia University (2018). Gridded population of the world, version 4 (gpwv4): Population density adjusted to match 2015 revision un wpp country totals, revision 11. `https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11/data-download`. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), Center For International Earth Science Information Network (CIESIN). last access: November 2022.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2010). Estimating agglomeration economies with history, geology, and worker effects. In E. L. Glaeser (Ed.), *Agglomeration economics*, pp. 15–66. University of Chicago Press.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, Chapter 5, pp. 247–348. Elsevier.

Davis, D. R. and D. E. Weinstein (2002). Bones, bombs, and break points: the geography of economic activity. *American Economic Review 92*(5), 1269–1289.

Davis, N., J. Badger, A. N. Hahmann, B. Ohrbeck, H. Bjarke, T. Olsen, N. Gylling Mortensen, D. Heathfield, M. Onninen, G. Lizcano, and O. Lacave (2019). Global wind atlas v3. https://doi.org/10.11583/DTU.9420803.v1. Technical University of Denmark.

Druckenmiller, H. and S. Hsiang (2018). Accounting for unobservable heterogeneity in cross section using spatial first differences. Technical report, National Bureau of Economic Research.

FAO/IIASA (2012). Global agro-ecological zones (gaez v3. 0). *IIASA, Laxenburg, Austria and FAO, Rome, Italy.*

Feng, M., J. O. Sexton, S. Channan, and J. R. Townshend (2016). A global, high-resolution (30-m) inland water body dataset for 2000: First results of a topographic–spectral classification algorithm. *International Journal of Digital Earth 9*(2), 113–133.

Galdo, V., Y. Li, and M. Rama (2019). Identifying urban areas by combining human judgment and machine learning: An application to india. *Journal of Urban Economics*, 103229.

Hammer, M. S., A. van Donkelaar, C. Li, A. Lyapustin, A. M. Sayer, N. C. Hsu, R. C. Levy, M. Garay, O. Kalashnikova, R. A. Kahn, et al. (2020). Global estimates and long-term trends of fine particulate matter concentrations (1998-2018). *Environmental Science & Technology*.

Heblich, S., A. Trew, and Y. Zylberberg (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy 129*(5), 1508–1552.

Henderson, J. V., D. Nigmatulina, and S. Kriticos (2021). Measuring urban economic density. *Journal of Urban Economics 125*, 103188.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The global distribution of economic activity: Nature, history, and the role of trade. *The Quarterly Journal of Economics 133*(1), 357–406.

Kummu, M., M. Taka, and J. H. Guillaume (2018). Gridded global datasets for gross domestic product and human development index over 1990–2015. *Scientific data 5*, 180004.

Lamsal, L., R. Martin, D. D. Parrish, and N. A. Krotkov (2013). Scaling relationship for NO2 pollution and urban population size: a satellite perspective. *Environmental Science & Technology 47*(14), 7855–7861.

Lamsal, L., R. Martin, A. Van Donkelaar, M. Steinbacher, E. Celarier, E. Bucsela, E. Dunlea, and J. Pinto (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne ozone monitoring instrument. *Journal of Geophysical Research: Atmospheres 113*(D16).

LandScan (2018). Global population distribution data. `https://landscan.ornl.gov/`. OAK Ridge National Laboratory. Last access: November 2022.

Lloyd, C. T., A. Sorichetta, and A. J. Tatem (2017). High resolution global gridded data for use in population studies. *Scientific data 4*, 170001.

Moreno-Monroy, A. I., M. Schiavina, and P. Veneri (2020). Metropolitan areas in the world. delineation and population trends. *Journal of Urban Economics*, 103242.

NASA Earth Observatory (2015). Seeing through the smokey pall: Observations from an grim indonesian fire season. `https://earthobservatory.nasa.gov/features/IndonesianFires`. last access: November 2022.

Natural Earth (2018). Natural earth version 4.1.0. `https://www.naturalearthdata.com/`. Last access: November 2022.

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics 94*(1), 20–36.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, et al. (2001). Terrestrial ecoregions of the world: A new map of life on eartha new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience 51*(11), 933–938.

Ramankutty, N., J. A. Foley, J. Norman, and K. McSweeney (2002). The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Global Ecology and biogeography 11*(5), 377–392.

Reba, M., F. Reitsma, and K. C. Seto (2016). Spatializing 6,000 years of global urbanization from 3700 bc to ad 2000. *Scientific data 3*, 160034.

Sarzynski, A. (2012). Bigger is not always better: A comparative analysis of cities and their air pollution impact. *Urban Studies 49*(14), 3121–3138.

Van Donkelaar, A., R. V. Martin, M. Brauer, N. C. Hsu, R. A. Kahn, R. C. Levy, A. Lyapustin, A. M. Sayer, and D. M. Winker (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology 50*(7), 3762–3772.

Van Donkelaar, A., R. V. Martin, R. J. Spurr, and R. T. Burnett (2015). High-resolution satellite-derived pm2. 5 from optimal estimation and geographically weighted regression over north america. *Environmental science & technology 49*(17), 10482–10491.

World Bank (2019). World development indicators. `https://datacatalog.worldbank.org/dataset/world-development-indicators`. The World Bank. Last access: November 2022.

Appendix

A   Ground-level pollution data

Pollution data as measured by satellites comes in the form of Aerosol Optical Depth (AOD), which deduct pollution concentration from the intensity of light that is reflected into space. However, this measure is not the same as ground-level pollution concentration, which reflects actual pollution exposure. As we are interested in actual exposure, we use ground-level pollution as described by Hammer et al. (2020) for $PM_{2.5}$. They deduct their estimates by using a GEOS-chem chemical transport model, relating measures from ground-level monitors with satellite measured AOD. The $PM_{2.5}$-data comes in two versions: one that applies geographically weighted regression (GWR) and one that does not. If GWR is applied, then the correlation of ground-monitor measurements with AOD measurements is even higher than without the GWR application (with a slope of up to .97 using GWR compared to one of 0.90 without GWR). We stick to the data product, which is closest to ground-based monitors, which are the GWR products.

B Additional results

Table A.1: IV and OLS regressions with population before 1750 as IV

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Raster-level estimations								
log(Sum of population)	0.172***	0.251***	0.0822***	0.111***	0.380***	0.498***	0.304***	0.371***
	(0.0262)	(0.0600)	(0.0226)	(0.0356)	(0.0299)	(0.0891)	(0.0378)	(0.0799)
N	357	357	357	357	358	358	358	358
R^2	0.580	0.445	0.879	0.837	0.765	0.489	0.890	0.815
Countries	73	73	73	73	73	73	73	73
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel B: FUA-level estimations								
log(Sum of population)	0.116***	0.207***	0.0816***	0.0339	0.300***	0.333***	0.298***	0.258***
	(0.0240)	(0.0530)	(0.0234)	(0.0333)	(0.0281)	(0.0607)	(0.0365)	(0.0564)
N	322	322	322	322	322	322	322	322
R^2	0.598	0.578	0.876	0.873	0.768	0.767	0.887	0.886
Countries	70	70	70	70	70	70	70	70
Country FE	No	No	Yes	Yes	No	No	Yes	Yes
Est. method	OLS	IV	OLS	IV	OLS	IV	OLS	IV

Note: The table presents coefficients of OLS and IV regressions for both raster-level and city-level (FUA) results. The instrument used is population between 1650 and 1750. Samples are harmonized such that OLS regressions only include those raster cells/cities for which we have the instrument available. The first two columns of each pollutant show results without country fixed effects, while the latter two of each show results including country FE. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.2: Raster-level OLS regressions with non-zero pollution

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.278***		0.242***		0.253***		0.258***	
	(0.00348)		(0.00401)		(0.00698)		(0.00730)	
log(pop density)		0.251***		0.206***		0.238***		0.239***
		(0.00345)		(0.00390)		(0.00672)		(0.00691)
log(GDP)	-0.0581***	-0.0767***	0.0340**	0.0228	0.322***	0.312***	0.229***	0.220***
	(0.00767)	(0.00773)	(0.0148)	(0.0149)	(0.0179)	(0.0178)	(0.0382)	(0.0382)
Temperature	-0.00297*	0.00199	0.0220***	0.0279***	0.0873***	0.0904***	0.173***	0.177***
	(0.00179)	(0.00180)	(0.00233)	(0.00233)	(0.00525)	(0.00524)	(0.00727)	(0.00725)
Precipitation	0.000235	0.0000804	0.000832***	0.000623***	-0.00748***	-0.00765***	-0.00747***	-0.00767***
	(0.000167)	(0.000168)	(0.000186)	(0.000186)	(0.000678)	(0.000678)	(0.000755)	(0.000755)
Wind speed	-0.180***	-0.196***	-0.114***	-0.130***	-0.259***	-0.268***	-0.273***	-0.282***
	(0.00637)	(0.00648)	(0.00680)	(0.00689)	(0.0200)	(0.0199)	(0.0213)	(0.0213)
1(On coast)	-1.425***	-1.593***	-1.329***	-1.461***	-0.998***	-1.115***	-0.756***	-0.867***
	(0.0335)	(0.0347)	(0.0342)	(0.0354)	(0.0932)	(0.0932)	(0.0890)	(0.0885)
1(Close to lake)	-0.934***	-0.969***	-0.869***	-0.888***	0.189*	0.159	0.338***	0.308***
	(0.0836)	(0.0859)	(0.0819)	(0.0839)	(0.0984)	(0.0993)	(0.105)	(0.105)
1(Close to river)	-0.0731**	-0.0562	-0.103***	-0.0845**	-0.00942	0.00141	-0.0592	-0.0493
	(0.0367)	(0.0369)	(0.0364)	(0.0365)	(0.0444)	(0.0444)	(0.0427)	(0.0428)
1(Close to dirty powerplant)	-0.227***	-0.138***	-0.259***	-0.163***	0.0922	0.138**	-0.0389	0.00367
	(0.0258)	(0.0262)	(0.0252)	(0.0253)	(0.0668)	(0.0668)	(0.0582)	(0.0582)
Ruggedness	-0.00243*	-0.000409	0.00137	0.00238*	-0.158***	-0.157***	-0.131***	-0.130***
	(0.00128)	(0.00138)	(0.00138)	(0.00138)	(0.00820)	(0.00820)	(0.00781)	(0.00780)
Latitude	-0.000609	-0.000914	-0.0145***	-0.0157***	0.0551***	0.0537***	0.0918***	0.0896***
	(0.00115)	(0.00116)	(0.00190)	(0.00190)	(0.00353)	(0.00353)	(0.00641)	(0.00641)
N	180489	180453	180489	180453	181663	181629	181663	181629
R^2	0.276	0.260	0.310	0.296	0.196	0.196	0.271	0.271
Countries	161	161	161	161	161	161	161	161
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of OLS regressions including all relevant controls. Apart from those that are listed in the table, the estimations additionally account for all biome indicators, and land suitability. Columns 3,4,7 and 8 additionally include country-fixed effects (Country FE). All pollution observations that are zero are replaced by the minimum pollution value in the sample. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.3: Raster-level OLS regressions - urban-rural gradient

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Urban population	0.710***		0.294***		0.988***		0.761***	
	(0.0109)		(0.00792)		(0.0151)		(0.0153)	
Urban pop density		0.717***		0.265***		1.012***		0.774***
		(0.0117)		(0.00855)		(0.0159)		(0.0156)
N	175390	175388	175390	175388	179484	179466	179484	179466
R^2	0.370	0.366	0.605	0.604	0.211	0.215	0.261	0.266
Countries	177	177	177	177	171	171	171	171
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population (density) categories instead of continuous population density or the sum of population. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: Raster-level OLS regressions with population categories

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
50k to 100k	0.473***		0.211***		0.760***		0.560***	
	(0.0109)		(0.00778)		(0.0144)		(0.0136)	
100k to 500k	0.820***		0.334***		1.083***		0.861***	
	(0.0132)		(0.00972)		(0.0182)		(0.0180)	
500k to 1m	1.144***		0.533***		1.367***		1.219***	
	(0.0252)		(0.0178)		(0.0378)		(0.0305)	
>1m	1.123***		0.569***		1.767***		1.637***	
	(0.0364)		(0.0269)		(0.0614)		(0.0521)	
Low Density		0.553***		0.196***		0.822***		0.608***
		(0.0116)		(0.00853)		(0.0150)		(0.0142)
Moderate Density		0.792***		0.278***		1.084***		0.851***
		(0.0166)		(0.0118)		(0.0215)		(0.0198)
High Density		1.024***		0.436***		1.340***		1.065***
		(0.0216)		(0.0160)		(0.0298)		(0.0244)
Very High Density		1.064***		0.496***		1.535***		1.382***
		(0.0279)		(0.0217)		(0.0425)		(0.0366)
N	175390	175388	175390	175388	179484	179466	179484	179466
R^2	0.376	0.371	0.607	0.605	0.212	0.216	0.262	0.268
Countries	177	177	177	177	171	171	171	171
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population (density) categories instead of continuous population density or the sum of population. The population categories are: 0 to 50.000 (50k) [base category], 50k-100k, 100k-500k, 500k-1million or more than 1 million inhabitants. The density categories are: very low density [base category], low density, moderate density, and very high density. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km and continuous distance to coast measure), agricultural ones (biome indicators, land suitability for agriculture), weather (wind speed, temperature, precipitation) as well as ruggedness, latitude, log(GDP), and and indicator for a dirty power plant nearby. Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.5: GADM-level OLS regressions

	PM$_{2.5}$				NO$_2$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Sum of population)	0.108***		0.0289***		0.143***		0.114***	
	(0.00203)		(0.00162)		(0.00257)		(0.00323)	
log(pop density)		0.112***		0.0569***		0.245***		0.207***
		(0.00172)		(0.00144)		(0.00194)		(0.00214)
log(GDP)	-0.0276***	-0.0374***	0.101***	0.0933***	0.363***	0.357***	0.228***	0.202***
	(0.00403)	(0.00391)	(0.00522)	(0.00516)	(0.00463)	(0.00405)	(0.00679)	(0.00590)
Temperature	-0.00859***	-0.0120***	-0.0163***	-0.0163***	-0.00983***	-0.0232***	-0.0115***	-0.0113***
	(0.000839)	(0.000803)	(0.000920)	(0.000882)	(0.00119)	(0.000982)	(0.00150)	(0.00129)
Precipitation	-0.000105***	-0.000118***	0.0000690***	0.0000710***	-0.000141***	-0.000152***	-0.000141***	-0.000132***
	(0.00000828)	(0.00000828)	(0.00000847)	(0.00000819)	(0.00000881)	(0.00000760)	(0.0000103)	(0.00000893)
Wind Speed	-0.181***	-0.172***	-0.0778***	-0.0683***	0.0270***	0.0588***	-0.0175***	0.0167***
	(0.00314)	(0.00311)	(0.00311)	(0.00305)	(0.00398)	(0.00344)	(0.00388)	(0.00348)
1(On coast)	-0.380***	-0.385***	-0.178***	-0.192***	-0.164***	-0.222***	-0.128***	-0.175***
	(0.0102)	(0.0101)	(0.00760)	(0.00749)	(0.0116)	(0.0101)	(0.0104)	(0.00906)
1(Close to lake)	-0.0936***	-0.00463	0.0964***	0.0928***	-0.327***	-0.223***	0.0621**	0.0536**
	(0.0306)	(0.0297)	(0.0196)	(0.0190)	(0.0336)	(0.0279)	(0.0306)	(0.0268)
1(Close to river)	-0.00559	0.110***	-0.0176**	0.00774	-0.136***	0.0472***	-0.134***	-0.0376***
	(0.0107)	(0.0106)	(0.00690)	(0.00676)	(0.0136)	(0.0110)	(0.0115)	(0.00985)
1(Dirty power plant in district)	0.0452***	0.188***	0.103***	0.0881***	0.00618	0.145***	0.129***	0.0837***
	(0.0127)	(0.0126)	(0.00956)	(0.00928)	(0.0162)	(0.0138)	(0.0147)	(0.0126)
N	44933	44933	44933	44933	44560	44560	44560	44560
R^2	0.397	0.407	0.760	0.769	0.637	0.722	0.775	0.823
Countries	154	154	154	154	152	152	152	152
Country FE	No	No	Yes	Yes	No	No	Yes	Yes

Note: The table presents coefficients of estimations using population data by global administrative areas (GADM) rather than LandScan data. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude, and log(GDP). Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors are robust. t statistics are in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.6: GADM-level long differences regressions

	PM$_{2.5}$		NO$_2$	
	(1)	(2)	(3)	(4)
log(Sum of population)		0.146***		0.0591***
		(0.0115)		(0.00759)
log(pop density)	0.147***		0.0597***	
	(0.0115)		(0.00763)	
N	90396	90396	90010	90010
R^2	0.136	0.136	0.087	0.087
Countries	159	159	157	157

Note: The table presents coefficients of long differences estimations on sub-national administrative level comparable to NUTS-3 regions. This corresponds to districts in Germany. All estimations include the following control variables: Trade controls (river, lake, coastline within 25km), land suitability for agriculture, temperature, precipitation as well as ruggedness, latitude, and log(GDP). Columns 3,4,7 and 8 include country-fixed effects (Country FE). Standard errors are robust. t statistics are in parentheses. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.7: Raster-level OLS regressions with varying sets of control variables

	(1)	(2)	(3)	(4)	(5)
Panel A: $PM_{2.5}$ ***Estimates***					
log(Sum of population)	0.0294***	0.0357***	0.0481***	0.0326***	0.0517***
	(0.00149)	(0.00144)	(0.00148)	(0.00147)	(0.00134)
N	175237	175274	175237	175237	175237
R^2	0.607	0.605	0.589	0.582	0.592
Countries	161	162	161	161	161
Country FE	Yes	Yes	Yes	Yes	Yes
Panel B: NO_2 ***Estimates***					
log(Sum of population)	0.157***	0.172***	0.209***	0.173***	0.222***
	(0.00275)	(0.00262)	(0.00263)	(0.00261)	(0.00244)
N	178535	178579	178535	178535	178535
R^2	0.675	0.666	0.614	0.665	0.606
Countries	160	161	160	160	160
Country FE	Yes	Yes	Yes	Yes	Yes

Note: The table presents coefficients of OLS regressions including different sets of control variables. The first column of each panel is the reference outcome as presented in Table 4, that includes all control variables. Column 2 excludes GDP, column 3 excludes weather variables (temperature, wind, precipitation), and column 4 does not take into account trade-specific controls (distance to lake, river, coast). Finally, column 5 excludes the biome indicators. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

C   Spatial First Differences

The SFD estimation considers only differences between neighbouring cells. The crucial assumption to hold in this context is that

$$E[y_i|\mathbf{x}_{i-1}] = E[y_{i-1}|\mathbf{x}_{i-1}]\forall\,\{i, i-1\},$$

which states that pollution y in adjacent neighbouring grid cells i and $i-1$ would be equal if they had the same population density x_{i-1}. The authors refer to this as the "Local Conditional Independence Assumption" (LCIA).

We estimate the following equation:

$$\Delta y_i = \Delta x_i \beta_{SFD} + \theta_c + \Delta\epsilon_i,$$

where Δ is the difference operator. To make sure that the LCIA assumption holds, we add country fixed effects θ_c to our regressions. In a nutshell, we then compare neighbouring cells within countries.

Since there are neighbouring cells in North-South and in East-West direction, we will always run two separate regressions: one for horizontal neighbours (East-West direction) and one for vertical neighbours (North-South). We consider the SFD results as a lower bound on the density effects, since it does not allow to account for many potentially important differences between non-neighbouring cells.

Table A.8: Spatial first differences

	NO$_2$				PM$_{2.5}$			
	(1) WE	(2) NS	(3) WE	(4) NS	(5) WE	(6) NS	(7) WE	(8) NS
log(pop density)	0.0313***	0.0613***	0.0313***	0.0368***	0.0266***	0.00986***	0.0203***	0.00596***
	(0.00149)	(0.00225)	(0.00127)	(0.00137)	(0.00179)	(0.00188)	(0.00169)	(0.00182)
log(GDP)			0.0409***	0.277***			-0.00301	0.0829***
			(0.0149)	(0.0180)			(0.0138)	(0.0143)
Temperature			0.0453***	0.0741***			-0.00588***	-0.00111
			(0.00308)	(0.00323)			(0.00186)	(0.00205)
Precipitation			-0.000344*	0.00159***			0.000798***	-0.000262*
			(0.000197)	(0.000265)			(0.000144)	(0.000159)
Wind Speed			-0.0257***	-0.0341***			-0.127***	-0.105***
			(0.00416)	(0.00389)			(0.00403)	(0.00405)
1(Close to river)			0.0199*	-0.00441			-0.00935	0.00519
			(0.0121)	(0.0138)			(0.0130)	(0.0134)
1(Close to lake)			0.0696***	0.0348			-0.107***	-0.0808**
			(0.0233)	(0.0315)			(0.0259)	(0.0368)
1(On coast)			0.00119	-0.0371***			-0.313***	-0.245***
			(0.00976)	(0.00971)			(0.0155)	(0.0156)
Ruggedness			-0.00428***	-0.00130***			0.000767*	-0.000117
			(0.000530)	(0.000459)			(0.000443)	(0.000425)
N	162268	163530	159674	160907	158506	159646	156380	157456
R^2	0.019	0.060	0.177	0.204	0.009	0.004	0.063	0.025
No. of grids	162268	163530	159674	160907	158506	159646	156380	157456

Note: The table presents coefficients of Spatial first differences regressions including different sets of control variables. For all control variables, coefficients are provided in the table. Standard errors (in parentheses) are clustered within three-by-three squares of grid cells times year. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.