RESEARCH ARTICLE

A Gini approach to spatial CO_2 emissions

Bin Zhou[1,2]*, Stephan Thies[1], Ramana Gudipudi[1], Matthias K. B. Lüdeke[1], Jürgen P. Kropp[1,3], Diego Rybski[1,4]

1 Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Potsdam, Germany,
2 Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Beersheba, Israel,
3 Department of Geo- and Environmental Sciences, University of Potsdam, Potsdam, Germany,
4 Department of Environmental Science Policy and Management, University of California Berkeley, Berkeley, CA, United States of America

* zhoub@post.bgu.ac.il
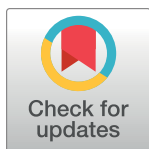
Abstract

Combining global gridded population and fossil fuel based CO_2 emission data at 1 km scale, we investigate the spatial origin of CO_2 emissions in relation to the population distribution within countries. We depict the correlations between these two datasets by a quasi-Lorenz curve which enables us to discern the individual contributions of densely and sparsely populated regions to the national CO_2 emissions. We observe pronounced country-specific characteristics and quantify them using an indicator resembling the Gini-index. As demonstrated by a robustness test, the Gini-index for each country arise from a compound distribution between the population and emissions which differs among countries. Relating these indices with the degree of socio-economic development measured by per capita Gross Domestic Product (GDP) at purchase power parity, we find a strong negative correlation between the two quantities with a Pearson correlation coefficient of -0.71. More specifically, this implies that in developing countries locations with large population tend to emit relatively more CO_2, and in developed countries the opposite tends to be the case. Based on the relation to urban scaling, we discuss the implications for CO_2 emissions from cities. Our results show that general statements with regard to the (in)efficiency of large cities should be avoided as it is subject to the socio-economic development of respective countries. Concerning the political relevance, our results suggest a differentiated spatial prioritization in deploying climate change mitigation measures in cities for developed and developing countries.

Introduction

Urbanization is an ongoing process in many parts on the globe. It is projected that due to rural-urban migration much of the future urbanization is going to take place in developing and transition countries. This leads to ever more mega-cities [1, 2]. In parallel, humanity is facing another challenge, namely climate change. To date, cities, despite occupying less than 1% of the global land area, account for more than 70% of the anthropogenic green house gas (GHG) emissions [3]. Therefore, cities are often identified as the key focal areas for global

mitigation actions. While a large contribution of the global CO$_2$ emissions is commonly attributed to cities [4], the CO$_2$ reduction role of further urbanization is also discussed with the argument of efficiency gains associated with the high densities in cities [5]. Moreover, cities are known to perform more efficiently in addressing the basic needs of human beings [5]. Hence, a diversified view on cities is needed and in view of climate change mitigation, a better understanding of the interplay between urbanization, origin of CO$_2$ emissions, and socio-economic development is of great interest.

Globally, cities are characterized by higher population densities compared to rural areas. Recent literature has identified the crucial role played by population density in either increasing or decreasing the emission efficiency in cities [6–10]. The impact of population density on reducing/increasing CO$_2$ emissions in these studies is either calculated based on specific assumptions made to calculate the city specific CO$_2$ emissions or through the construction of city clusters using a clustering algorithm, see [11, 12]. However, most of these studies are limited to a specific country or a region. Therefore, there is a gap in the existing literature about the sub-national origin of CO$_2$ emissions at a global scale. Bridging this gap would provide better insights as to whether population density is a crucial factor in improving/decreasing emission efficiency and would identify other factors that influence CO$_2$ emissions at a country scale.

Here, we investigate how the spatial origin of CO$_2$ emissions relates to the spatial distribution of population. We address the questions, to which extent locations of large population also emit the most CO$_2$ and if there is any dependence on human development. In order to avoid discussions about the proper city definition, the correlations are analyzed on the level of grid cells—keeping in mind that locations of high population are likely to belong to cities. Thus, we analyze population and CO$_2$ emissions by employing a quasi-Lorenz curve that relates the cumulative population and cumulative emissions for entire countries on a grid-cell level (the Lorenz curve was originally used to describe unequal income distribution).

The shape of these curves explains whether the emissions are concentrated in locations of high or low population. Inspired by the apparent similarity, we extend the well-known Gini-index. Based on the data employed, we find that within many countries, locations with high or low population exhibit different relative emissions. We thus compare the extended "Gini-index" with the economic strength of the considered countries (as captured by the GDP per capita) which can be to some extent interpreted as a measure for the stage of development. We further hypothesize that the development stage of respective countries plays an important role in explaining this relationship.

Earlier studies attempted to address the emission efficiency of densely populated regions by means of urban scaling, where an urban indicator is plotted against the city size in terms of population [13]. The exponent, estimated as the slope of a linear regression in the log-log representation, quantifies efficiency gains of large or small cities. However, in case of urban CO$_2$ emissions, published results from urban scaling leave an inconclusive picture (for an overview we refer to [14, 15]). In the present work we address this issue by combining high resolution, global population and CO$_2$ emission data sets in order to quantify whether locations with high or low population emit more or less CO$_2$. We further discuss an analytic link between our approach and urban scaling.

Materials and methods

Population data

We used the Gridded Population of the World, version 4 (GPWv4) population count data for the year 2010 [16]. GPWv4 data allocate the population counts of census units collected
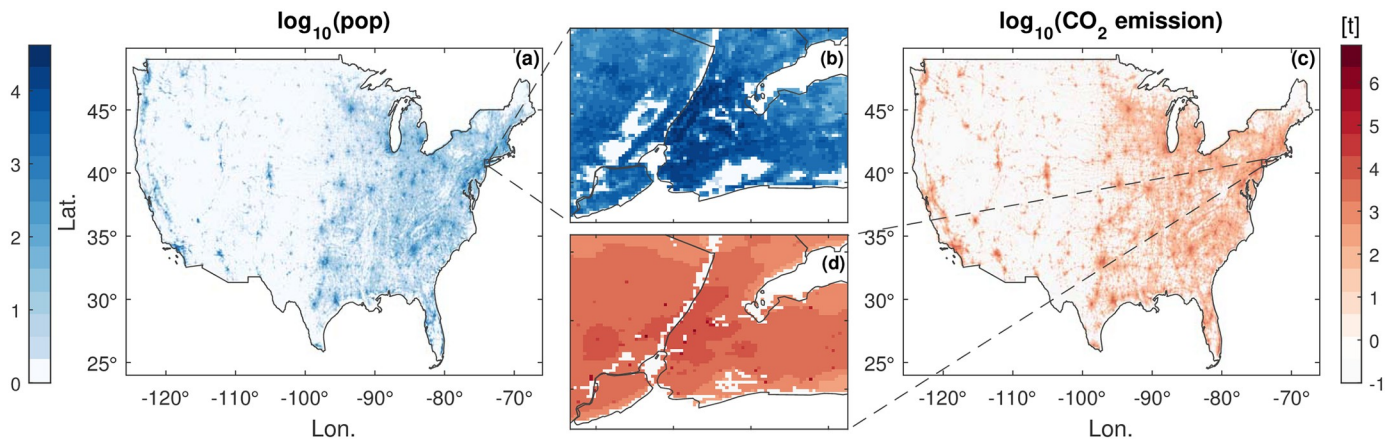
Fig 1. Population and CO₂ emissions for the contiguous USA. (a) Gridded Population of the World, version 4, GPWv4 [17] at 1×1 km² spatial resolution in 2010 and (c) total anthropogenic CO₂ emission from ODIAC data [18] for the same region, year, and resolution. (b) and (d) depict magnified views of population and CO₂ emission for the New York metropolitan region, respectively. Visually, large agglomerations of population coincide with large amounts of emissions. To which extent they relate proportionally is the subject of this paper.

globally from various institutions into standard 1×1 km² grid cells by means of an areal-weighting interpolation [17]. Fig 1(a) illustrates the GPWv4 data in the year 2010 for the contiguous US. The distribution of population in the US exhibits an inhomogeneity. The metropolitan urban agglomerations accommodate a large share of population in the US, whereas the states in the Mountain West are generally sparsely populated.

CO₂ emissions data

Fossil fuel based CO₂ emission estimates are obtained from the Open source Data Inventory of Anthropogenic CO₂ (ODIAC) emissions of version ODIAC2015a available globally at 1×1 km² grid for the year 2010 [18]. In the ODIAC dataset, point sources, i.e. power plant emissions obtained from the database CARMA (Carbon Monitoring and Action) are directly assigned to the grids, while non-point sources (e.g. emissions from transport, industrial, residential, and commercial sectors) are disaggregated based on global and national emission estimates made by the Carbon Dioxide Information Analysis Center (CDIAC) [19], using remotely sensed nightlight data as a proxy. An exception is the emissions from cement production which have point source origins but are spatially disaggregated as non-point sources. Non-land emissions, such as those from international bunkers (international aviation and maritime shipping), are assigned to the non-point emissions.

Compared with conventional population-based approaches, the nightlight data can trace the human activities more appropriately [20, 21]. Worthy of special mention is that the gridded emission data of ODIAC used in this study is not disaggregated using population density as a proxy. Therefore, the two datasets depict distinguishing zonal patterns, as shown in the example for the New York metropolitan region in Fig 1(b) and 1(d). Without relying on the time-consuming update of census data, emissions allocated using nightlights can be updated more frequently and may be of particular importance for developing countries where conducting census is still a challenge. Fig 1(c) shows the gridded total anthropogenic CO₂ emissions (in tons) for the year 2010 for the contiguous US, analogous to the population data shown in Fig 1(a). As observed, the emissions also exhibit pronounced inhomogeneities.

In order to check the consistency of the results obtained, we further compare our results obtained from the ODIAC data with other CO₂ emission datasets, namely the Fossil Fuel Data

Assimilation System (FFDAS) version 2.0 [22] and the Emission Database for Global Atmospheric Research (EDGAR) version 4.3.2 [23]. Both data are for the year 2010. For the subnational analysis we also analyze the Vulcan data, which has been analyzed before [12]. However, we focus on ODIAC, since it has the highest resolution, and we discuss the results of other datasets in comparison to the ODIAC results.

The fundamentals of creating the four gridded CO_2 emission inventories used in this study have been compared and discussed in detail in [24]. In general, they differentiate themselves in terms of 1) the energy statistics used which determines the sectors included in calculating the total national CO_2 emissions, and 2) the approach to disaggregating and allocating the CO_2 emissions to a regular grid.

Dissimilarities among the inventories may be dominated by the disaggreation method. FFDAS applies the Kaya identity to balance CO_2 emissions across regions, relying on population and nightlight data [22, 25] (see also [15] for further information on the Kaya identity). Viewed as the most accurate emissions inventory, a bottom-up method has been used for the Vulcan data allocate large point sources, road-specific emissions, and non-point emissions to census tracts, and further resampled to a 10-km grid [26]. However, since the subnational emissions data are not always available, the Vulcan data is restricted to the USA at the moment.

Inhomogeneity index of CO_2 emissions G_e

In order to characterize the relation between country-wise population and CO_2 emissions, we plotted the cumulative quantities against each other. We sorted the grid cells of a country by population in ascending order and calculated the cumulative share of population and CO_2 emissions arising therefrom. Then we interpreted the cumulative emissions as a function of the corresponding cumulative population.

The plotted curves resemble the so-called *concentration curves* used to describe socio-economic inequalities. The most popular concentration curve is the Lorenz curve usually employed to visualize income inequalities. Other applications of concentration curves include for example the analysis of socio-economic inequalities in the health sector (e.g. [27]). Since the curves we compute here, do not agree exactly with the classical definition of a concentration curve we will refer to them as *quasi-Lorenz curves*. We justify the choice of this method by its simplicity—it only requires sorting—and the fact that it does not require any parameters or assumptions on functional forms.

To quantify the curves, we break the shape of each curve down to a single number. As it is well known and used in this context, we generalize the Gini coefficient, which originally has been introduced to quantify income inequality [28]. As illustrated in Fig 2, we distinguish between curves above or below the dashed line with a slope of 45˚—the line of equality. For the blue quasi-Lorenz curve, we defined an inhomogeneity index as the ratio of the area between the curve and the line of equality (marked as A) to the total area below the line of equality ($A + B$). Analogously, the inhomogeneity index of the green curve is $-A'/(A' + B')$. We arbitrarily assign the inhomogeneity index for the curves above the line of equality negative, and below positive.

Results

Inhomogeneity of emissions across countries

Fig 3 shows the quasi-Lorenz curves for a few countries. The slopes of the curves depend on per capita emissions. In the case of constant emissions per capita, the cumulative share of population and CO_2 emissions are proportional to each other and follow the diagonal. This is
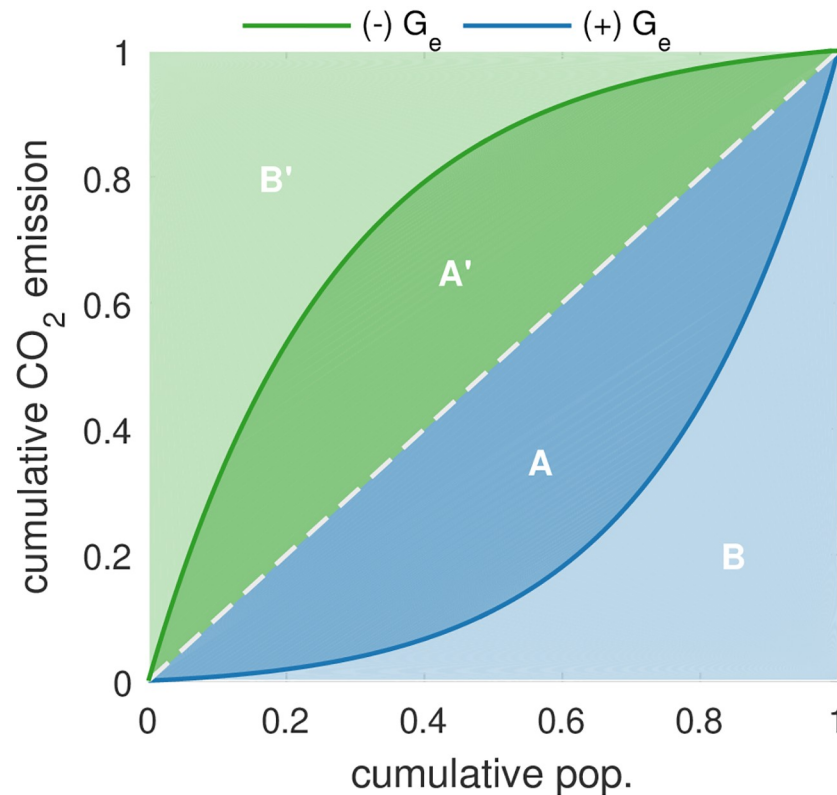
Fig 2. Illustration of the inhomogeneity index G_e. Quasi-Lorenz curves (solid lines) and the calculation of G_e, which is inspired by the Gini coefficient: $G_{e+} = A/(A + B)$ and $G_{e-} = -A'/(A' + B')$.

approximately the case for Germany, Fig 3(e). In the case of the USA, Fig 3(a), and the UK, Fig 3(b), the curves are bent to the upper left corner, i.e. the grid cells with small population already include a large amount of emissions (large slope). On the contrary, the curves, e.g., for Uganda, Fig 3(h), and Kenya, Fig 3(g), are bent to the lower right corner. Many grid cells with small population are necessary to include a fair amount of emissions (small slope). Accordingly, curves bent to upper left indicate high per capita emissions in sparsely populated cells and comparably lower per capita emissions in densely populated cells, and vice versa.

In Germany, per capita CO₂ emissions of large cities are smaller than those of small ones, but the difference seems to be minor [29]. In contrast, per capita CO₂ in the UK emissions remarkably diverge between large and small cities, ranging from 25.6 tonnes per capita in Middlesbrough to 5.4 tonnes per capita in London in 2012, reflecting the impact of industrial base [30, 31].

Interestingly, in Fig 3 developed countries seem to belong to the group where the curves extend to the upper left corner and less developed countries seem to belong to the group where the curves extend to the opposite corner.

G_e versus GDP per capita at trans-national level

In order to verify whether there is a systematic relationship between the curve type and the level of countries' economic development, we plotted the values of the inhomogeneity index G_e for a large number of countries against the logarithm of GDP Purchasing Power
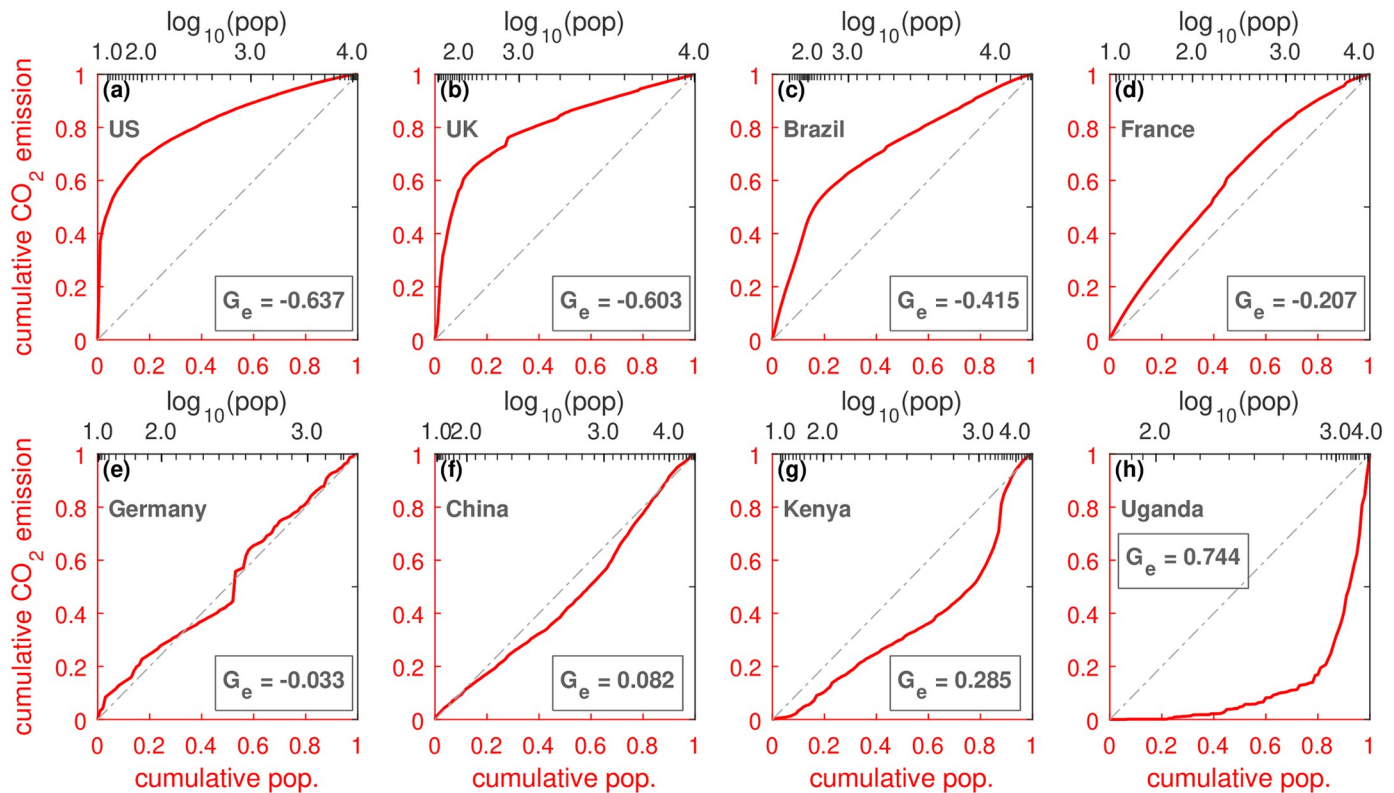
Fig 3. Quasi-Lorenz curves and corresponding inhomogeneity index G_e for selected countries. The country-specific curves are drawn by plotting the accumulated population (in ascending order) on the horizontal axis against the accumulated share of CO_2 emissions of the corresponding grid cells. The panels show the curves for (a) USA, (b) UK, (c) Brazil, (d) France, (e) Germany, (f) China, (g) Kenya, and (h) Uganda. If the curves follow the diagonal, then low and high densities have the same emissions per capita. If the curves are bent to the lower right corner, then cells of small density exhibit relatively low emissions and high population cells exhibit relatively high emissions. Curves in the upper left corner indicate the opposite behavior. The inhomogeneity index G_e is positive or negative. It can be seen, that various countries exhibit non-proportional relations between population and emissions. The inhomogeneity index G_e seems to be related to the development of the country.

Parity (PPP) per capita obtained from the World Bank, an important indicator for economic development.

As observed in Fig 4, the two quantities correlate (with a Pearson correlation coefficient $\rho = -0.71$, $p \leq 0.01$). In general, for developed countries G_e tends to have smaller values, and for developing ones it tends to have larger values. Thus, we generalize that in economically developing countries, high population densities are more emission intense and the opposite is the case in economically developed countries.

We repeated the analysis for the FFDAS and the EDGAR data. For reasons of consistency, we also analyze the ODIAC data aggregated to 10×10 km^2 resolution. For the three datasets, the resulting G_e-values are plotted against the GPD per capita, analogous to Fig 4.

Fig 5(b) reveals a very similar development dependence when FFDAS data are applied as for ODIAC [Fig 5(a)]. In contrast, for the EDGAR data [Fig 5(c)] the development dependence vanishes and is even slightly inverted ($\rho = 0.29$, $p \leq 0.01$). Differences between the G_e-values of the EDGAR and ODIAC or FFDAS data are most pronounced for developing countries. The difference in these results could be attributed to the poor quality of population census, high demographic dynamics, and insufficient geo-spatial data in developing countries. However, further investigation is needed in order to understand which of these methodological differences factor more with regard to the pronounced dissimilarities in the results. EDGAR relies
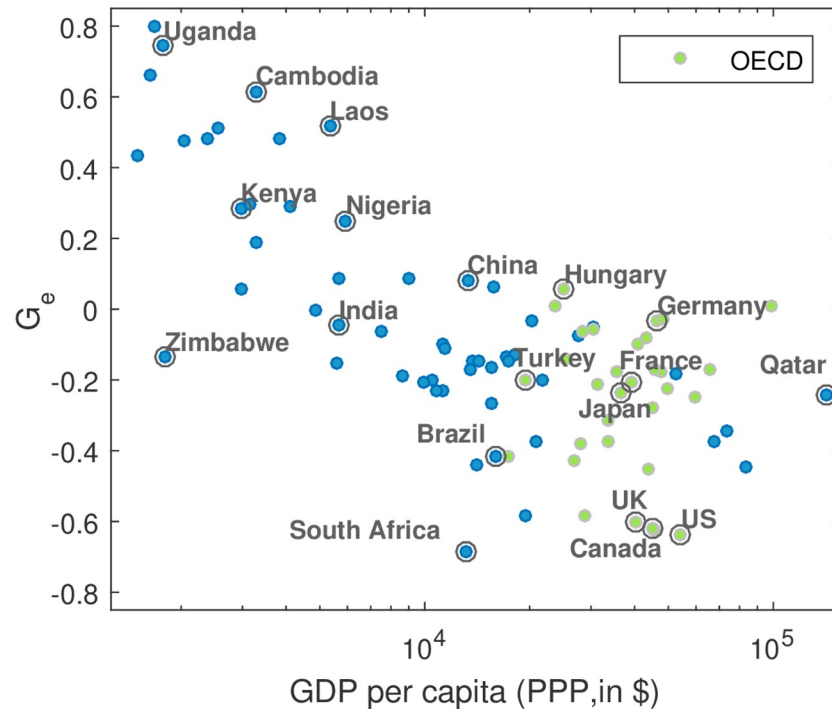
Fig 4. Development dependence of CO$_2$-population-inhomgeneity. The inhomogeneity index G_e is plotted vs. Gross Domestic Product (GDP) per capita (PPP) for 94 countries on a semi-logarithmic scale. For better readability only the symbols of a sub-set of countries are labeled. As can be seen, the G_e correlate with the economic development. The Pearson correlation coefficient between G_e and GDP on a logarithmic scale is $\rho = -0.71$ ($p \leq 0.01$). In more developed countries high population densities have lower emissions as low densities. The GDP data were obtained from the World Bank (http://data.worldbank.org), measured in USD of the year 2010.

https://doi.org/10.1371/journal.pone.0242479.g004

on road networks, population density and agriculture land use data to downscale the national emissions, which renders it more sensible to errors embedded in the proxy datasets. In comparison, FFDAS uses, besides population density, nightlight data to disaggregate emissions.

Moreover, since ODIAC and FFDAS are at least partly based on nightlight data for the subnational disaggregation [25, 33, 34], one may argue that the development dependence in Fig 5 (a) and 5(b) are simply due to such an effect in the nightlight data. Thus, we also analyzed the nightlight data from the Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) data [32] in an analogous way as the emissions data and the results are displayed in Fig 5(d). As can be seen, for nightlight data, we do not see any correlations between G_e-values and the GDP per capita. Accordingly, we conclude that the development dependence found in the ODIAC and FFDAS data is not stemming from the nightlight data. Overall, G_e-values tend to be negative for nightlights, indicating that locations of low population have a relatively strong contribution.

G_e versus GDP per capita at sub-national level

Next we analyzed whether the correlations between G_e (for ODIAC) and GDP per capita among countries also appear within a country. Taking China as an example, we disaggregate the national data into provinces. Analogously as for the countries, we calculate cumulative emissions vs. cumulative population and determine the inhomogeneity index at the province level. In Fig 6(a) the G_e-values are plotted vs. the corresponding GDP per capita values, as in
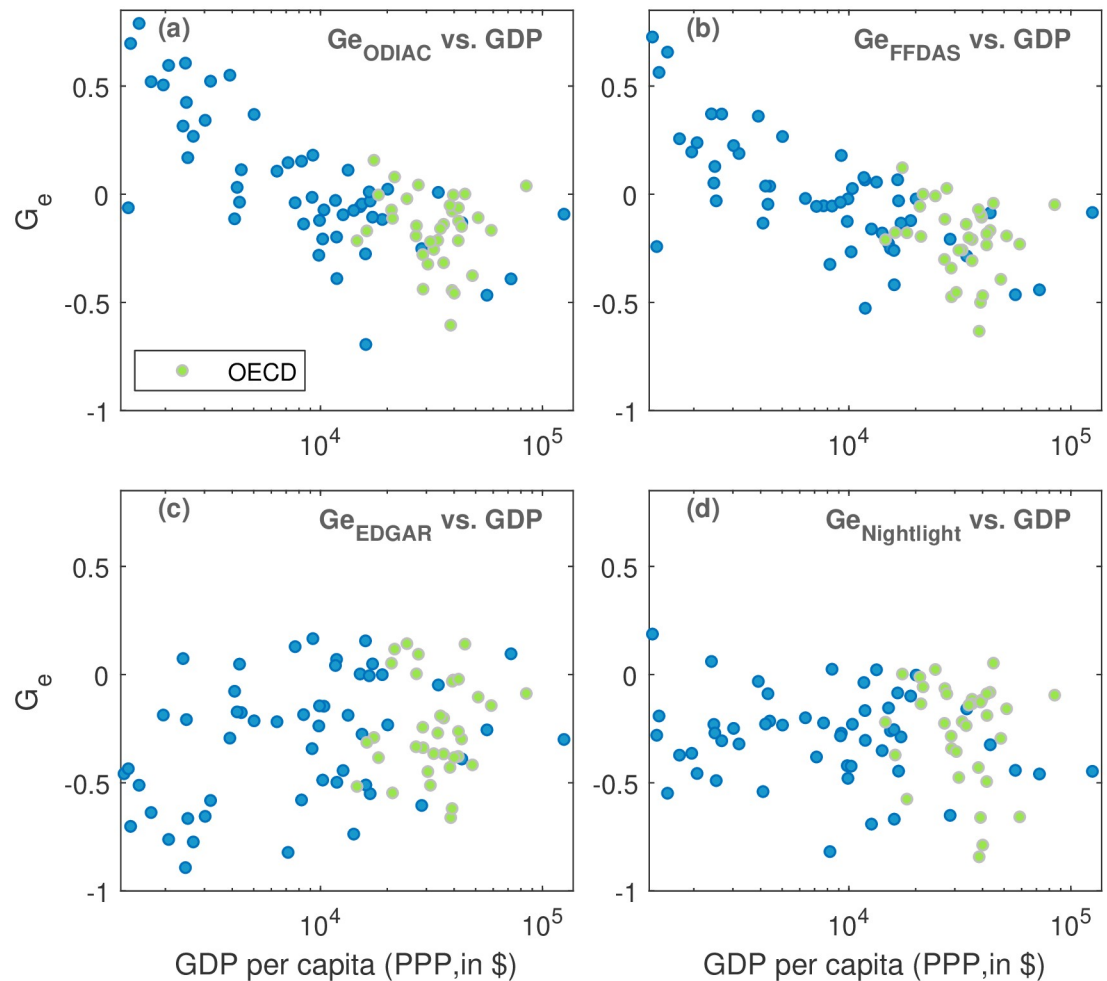
Fig 5. Comparison of CO$_2$-population-inhomgeneity for different CO$_2$ datasets and nightlights. (a) ODIAC, (b) FFDAS, (c) EDGAR, (d) nightlights [32]. Each panel is analogous to Fig 4, but for consistency of spatial resolution the underlying ODIAC data in (a) has been aggregated to 10 km resolution.

Fig 4 but now for provinces. Similar to the country analysis and even more pronounced, we find statistically significant correlations ($\rho = -0.87$, p-value: <0.01).

We performed the corresponding sub-national analysis for the USA on the state level. However, we could not find significant correlations (see S1 Fig in S1 File). Despite this lack of correlations, we find a spatial pattern in the USA. States at the west coast and in the North-east tend to have larger G_e-values. This is also the case for other states at the east coast and in the Midwest. States in the south as well as Montana, North Dakota, South Dakota tend to have more extreme G_e-values. Repeating the analysis for the Vulcan data, which might be considered the most detailed data, still no correlations between G_e and GDP per capita within the USA are found (S2 Fig in S1 File). However, the analysis does show weak correlations between the G_e-values of Vulcan and ODIAC data. This may imply that, albeit based on a relatively simple disaggregation scheme, the ODIAC datasets are able to describe the spatial inhomogeneity of CO2 emissions at a large scale comparably well as a more complex bottom-up based CO2 emission data, particularly in countries where an accurate CO2 inventory is available.
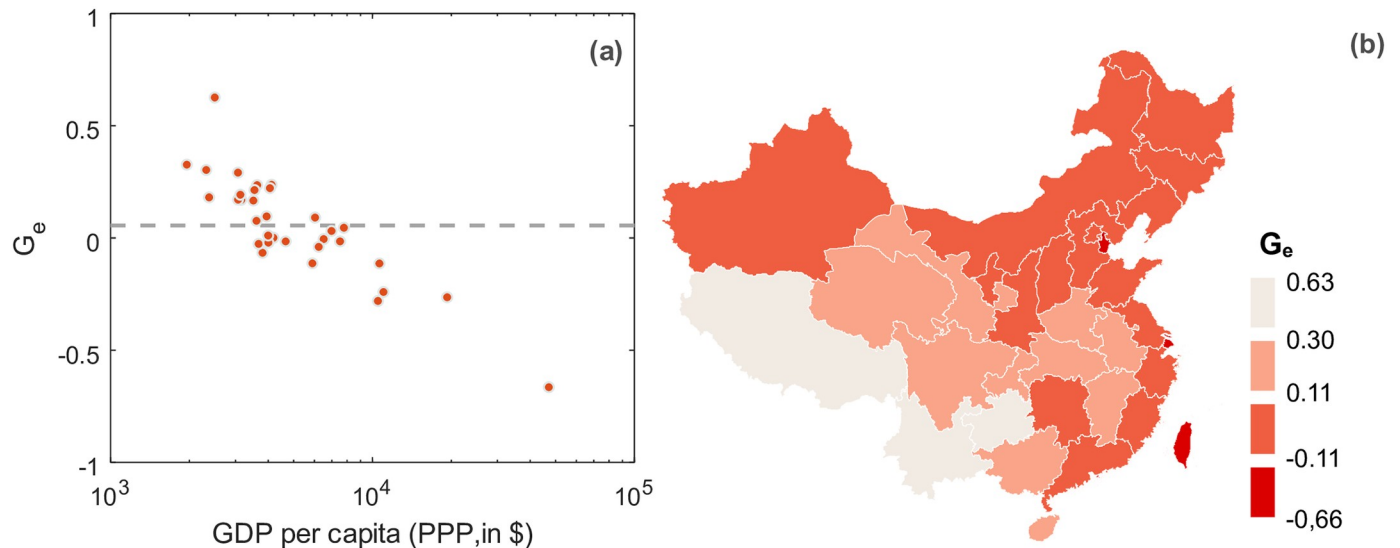
Fig 6. Sub-national inhomogeneity index G_e. We calculated the G_e on the province level for China. In (a) the G_e-values are plotted against the corresponding province GDP per capita values on a logarithmic scale, analogous to Fig 4. The dashed line indicates the country-level mean G_e. Panel (b) shows a map of China where the provinces are color-coded according to the inhomogeneity index G_e. It can be seen that the development dependence as found in Fig 4 does also hold on the sub-national scale—at least in China. Provinces with lowest and highest G_e-values are Hong Kong and Tibet, respectively. Note, however, that for the USA we do not find sub-national correlations (S1 Fig in S1 File). (Data source of China level-1 administrative boundaries: https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-1-states-provinces/).

https://doi.org/10.1371/journal.pone.0242479.g006

Robustness of G_e

Lastly we checked the robustness of the G_e coefficient. We explored different forms of sampling and randomization. In order to check the influence of outliers, we create random sub-samples of the ODIAC data. We constructed a set with 50% of the original size by randomly selecting pairs of population and emissions values from the original set without replacement for 1000 iterations. We calculated the cumulative quantities as before and determined the inhomogeneity index. Repeating the procedure we can assess the statistical spread. As observed in Fig 7, the resampling has minor influence on the shape of the curve and the resulting G_e-values. For Germany, Fig 7(a), 95% of the realizations lead to G_e-values in the range of -0.131 to 0.063, with a median and mean of -0.032, which is very close to the measured value -0.033. The sub-sampled robustness check for the UK led to analogous findings [Fig 7(b)].

Another way to randomize is to shuffle. Since in the analysis we have already sorted the data, we now shuffle only the emissions data and destroy the correlations between emissions and population. Then we perform the whole analysis and obtain cumulative emissions and population curves as well as G_e-values. Repeating the procedure we can assess the statistical spreading. The results are also displayed in Fig 7, and we find that the curves for the shuffled data are very different from the original curves which shows that the actual shapes in Fig 3 are due to the correlations between emissions and population. The shape of the curves for the shuffled data differs between Germany and the UK, Fig 7(a) and 7(b). Since shuffling destroys any correlations, the actual form of the curves can be attributed to a combination of the probability distributions of the population and emissions which differ among the countries.

Relation to urban scaling

The analysis of CO_2 efficiency that is carried out here using quasi-Lorenz curves can be related to the urban scaling approach as advocated in [13]. The urban scaling approach aims to
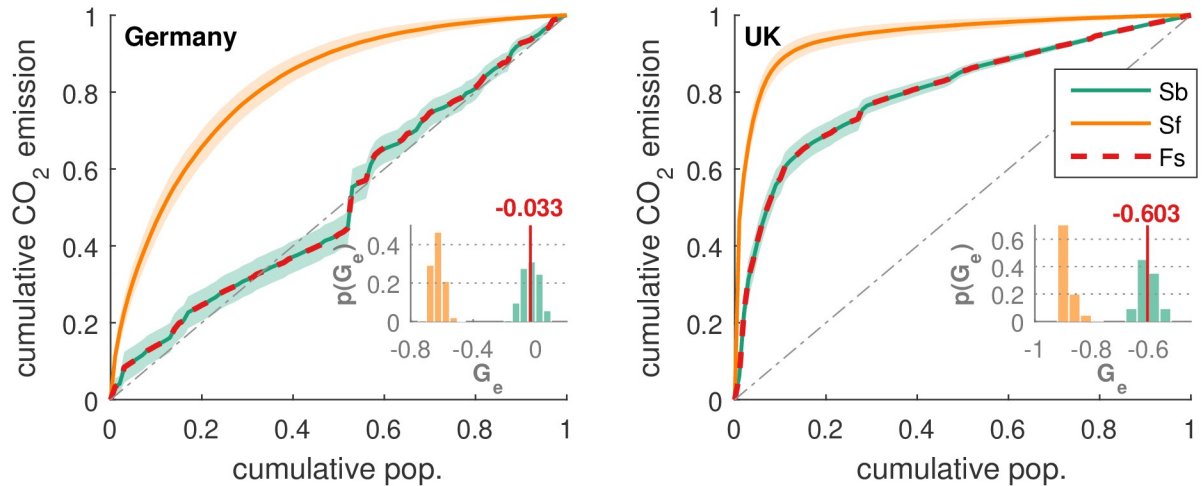
Fig 7. Robustness of G_e. In order to illustrate the robustness of the curves in Fig 3 we compare them with curves when the data is sub-sampled or shuffled. The panels for (a) Germany and (b) UK include the curves for the full samples (Fs), median and envelop for the sub-sampled data (Sb, green), and the median and envelop for the shuffled data (Sf, orange). The insets show histograms of the corresponding inhomogeneity indices. It can be seen that sub-samples of the data lead to similar results as for the full sample so that the results are not due to individual pixels. The G_e values from the shuffling approach -1, as the correlations between population and CO_2 are destroyed.

establish a parametric relationship between the urban population P_u of a city and the respective emissions E_u. In our analysis we do not analyze urban population and urban emissions explicitly but examine gridded population P_g and emission data E_g within countries. Since urban areas are usually characterized by high population densities (depending on the pixel size), one could transfer the idea of urban scaling to our setting and assume the scaling relationship $E_g \sim P_g^\beta$. The case of $\beta < 1$ indicates CO_2 efficiency gains with increasing population (density) while $\beta > 1$ is associated with efficiency losses. Here it is of interest how the non-parametric quasi-Gini coefficient G_e is related to the parametric scaling exponent β.

Generally there is no simple association between β and G_e. Empirically, the β coefficient is usually estimated as the slope of a linear regression of the logarithmic quantities. Hence, it depends on the correlations among the logarithmic quantities cor{log P_g, log E_g} and on the variance of log P_g and log E_g only. By contrast, G_e as a non-parametric estimator depends on the exact form of the joint distribution of P_g and E_g. However, it is possible to determine a specific expression for the relationship between G_e and β under certain conditions. The coefficients are related via [35]

$$G_e = \frac{\beta - 1}{2\lambda - \beta - 1} , \tag{1}$$

if P_g is Pareto distributed with shape parameter $\lambda > 1$ and a scaling relation of the form $E_g \sim P_g^\beta$ with $\beta < \lambda$ holds *exactly*.

For a detailed derivation see Sec.2 in SI. The formula shows that a scaling coefficient $\beta > 1$ is associated with $G_e > 0$. Equivalently, $\beta < 1$ implies $G_e < 0$. If $E_g \sim P_g^\beta$ holds only approximatively, Eq (1) should still give a reasonable approximation.

Under this scenario, our finding of development dependent G_e-values implies a corresponding development dependence of the scaling exponent β. Accordingly, in developing countries large cities are typically less emission efficient and vice versa in developed countries.

Discussion and conclusions

In summary, we have analyzed the correlations between the spatial distribution of population with CO$_2$ emissions using high resolution datasets. In order to understand these correlations we employed the quasi-Lorenz curve. The shape of the curve indicates to which extent locations of high or low population emit relatively more or less CO$_2$. We characterized the inhomogeneity by a generalized Gini coefficient. For the ODIAC and FFDAS data it depends on the socio-economic development of the considered country (developing countries exhibit relatively more emissions in locations of high population). For the EDGAR data there is no development dependence (overall relatively more emissions in locations of low population). Within China, the development dependence persists for the ODIAC data, but within the USA it vanishes for the ODIAC and Vulcan data. Sub-sampling and shuffling supports the robustness of our analysis.

There is a well-known association between urbanization, economic development, and carbon emissions. However, the quantitative relations behind this association are less understood. Here we show that also the location of emissions is influenced by the economic development. We conclude that during the course of development a spatial separation of emission source and population happens, based on the results for the ODIAC and FFDAS data. This means to some extent high-emitting sources relocate away from locations of large population. A possible explanation could be an increasing environmental consciousness and adoption of cleaner technologies—a trend similar to the environmental Kuznets curve (EKC). Another possibility could be altering composition of economic sectors from agriculture over emission intensive industry to service [36]. While a majority of national mitigation strategies target specific sectors, our results suggest a complementary spatial perspective to prioritize mitigation actions. Depending on the considered scope of emissions, these would be sparsely populated regions in developed countries and densely populated regions in developing and transition countries. Particular attentions should be paid to the latter, as these countries are projected to become more urbanized in the upcoming decades, which entails further rural-urban migration.

The difficulty in explaining the observed phenomenon of country-specific inhomogeneity indices may be attributable to a complex interplay of human activity on local, country, and international scale which entails more evaluation. Concentration or dispersion of human activities is strongly linked to the extent of urban sprawl. Such structural properties certainly affect both the population and the emissions. Moreover, as mentioned earlier, the proxies used to downscale national level CO$_2$ emissions and the sectors included while calculating the national level emission data will also impact the spatial inhomogeneity of the origin of CO$_2$ emissions. In addition, the location of point sources is an important aspect that can hardly be generalized on the national or even international scale. Maybe, a starting point could be a better understanding of the spatial characteristics of CO$_2$ efficiency. Explaining the presented phenomenon—i.e. development dependent concentration of emissions in locations of high or low population—remains a challenge for future research.

Our results for the ODIAC and FFDAS data are consistent with previously reported findings [14], according to which in developing countries large cities are comparably less efficient in terms of CO$_2$ emissions, and in developed ones small cities are less efficient. On the one hand, the present study provides stronger empirical evidence, e.g. because it is based on more data and the signatures are more pronounced. On the other hand, the methodology of the present study does not rely on any city definition [37, 38] or any assumption about the functional form of the correlations between population and emissions [39].

We argue that the affirmation "large cities are less green" [11] needs to be revised. According to our results only in developing countries large cities are less green. In developed

countries, including the USA, the opposite is the case, relatively more emissions stem from small cities. Anyways, we find it misleading to speak about "green cities" in the context of urban CO$_2$ emissions [7], since greenness usually refers to urban vegetation or metaphorically to pollution (while CO$_2$ is a colorless gas which as a GHG contributes to global warming).

Certainly, our analysis also has some potential caveats which we want to discuss briefly. The analysis stands and falls with the employed input data, so we cannot exclude to obtain other results if we use other population or emissions data as inputs. Why the EDGAR data leads to different results compared to ODIAC and FFDAS is an interesting problem requiring further research. Moreover, our curves, such as in Fig 3, can have (multiple) crossings with the diagonal, and the index G_e cannot capture to a full extent more complex shapes of the curves.

Another aspect that could be addressed in future studies is the role of the population density [6, 12, 14, 40, 41]. Here we avoid any discussion about city definitions by simply taking gridded data. Since the grid cells are approximately of equal area, the population count and the density are approximately identical. In order to investigate the influence of the density, a suitable city definition—joining grid-cells—will be necessary.

Supporting information

S1 File.
(PDF)

Acknowledgments

We thank M. Barthelemy for useful discussions.

Author Contributions

Conceptualization: Bin Zhou, Diego Rybski.

Data curation: Bin Zhou.

Formal analysis: Bin Zhou, Stephan Thies.

Funding acquisition: Jürgen P. Kropp.

Investigation: Bin Zhou, Diego Rybski.

Methodology: Bin Zhou, Stephan Thies, Ramana Gudipudi, Diego Rybski.

Project administration: Jürgen P. Kropp, Diego Rybski.

Software: Bin Zhou.

Supervision: Matthias K. B. Lüdeke, Jürgen P. Kropp, Diego Rybski.

Validation: Stephan Thies, Matthias K. B. Lüdeke.

Visualization: Bin Zhou.

Writing – original draft: Bin Zhou, Ramana Gudipudi, Diego Rybski.

Writing – review & editing: Bin Zhou, Ramana Gudipudi, Matthias K. B. Lüdeke, Jürgen P. Kropp, Diego Rybski.

References

1. United Nations. World Urbanization Prospects 2014: The 2014 Revision, (ST/ESA/SER.A/366); 2015.

2. Wigginton NS, Fahrenkamp-Uppenbrink J, Wible B, Malakoff D. Cities are the Future. Science. 2016; 352(6288):904–905. https://doi.org/10.1126/science.352.6288.904 PMID: 27199407

3. UN-HABITAT. Cities and Climate Change: Global Report on Human Settlements 2011; 2011. Available from: http://urban.ias.unu.edu/index.php/cities-and-climate-change/.

4. Seto KC, Güneralp B, Hutyra LR. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proc Natl Acad Sci U S A. 2012; 109(40):16083–8. https://doi.org/10.1073/pnas.1211658109 PMID: 22988086

5. Dodman D. Blaming cities for climate change? An analysis of urban greenhouse gas emissions inventories. Environ Urban. 2009; 21(1):185–201. https://doi.org/10.1177/0956247809103016

6. Newman PWC, Kenworthy JR. Gasoline Consumption and Cities. J Am Plann Assoc. 1989; 55(1):24–37. https://doi.org/10.1080/01944368908975398

7. Glaeser EL, Kahn ME. The greenness of cities: Carbon dioxide emissions and urban development. J Urban Econ. 2010; 67(3):404–418. https://doi.org/10.1016/j.jue.2009.11.006

8. Brown MA, Southworth F, Sarzynski A. The geography of metropolitan carbon footprints. Policy and Society. 2009; 27(4):285–304. https://doi.org/10.1016/j.polsoc.2009.01.001

9. Jones C, Kammen DM. Spatial Distribution of U.S. Household Carbon Footprints Reveals Suburbanization Undermines Greenhouse Gas Benefits of Urban Population Density. Environ Sci Technol. 2014; 48 (2):895–902. https://doi.org/10.1021/es4034364 PMID: 24328208

10. Kennedy C, Steinberger J, Gasson B, Hansen Y, Hillman T, Havránek M, et al. Greenhouse Gas Emissions from Global Cities. Environ Sci Technol. 2009; 43(19):7297–7302. https://doi.org/10.1021/es900213p PMID: 19848137

11. Oliveira EA, Andrade JS, Makse HA. Large cities are less green. Sci Rep. 2014; 4:4235. https://doi.org/10.1038/srep04235 PMID: 24577263

12. Gudipudi R, Fluschnik T, Ros AGC, Walther C, Kropp JP. City density and CO2 efficiency. Energ Policy. 2016; 91:352–361. https://doi.org/10.1016/j.enpol.2016.01.015

13. Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB. Growth, innovation, scaling, and the pace of life in cities. Proc Nat Acad Sci USA. 2007; 104(17):7301–7306. https://doi.org/10.1073/pnas.0610172104 PMID: 17438298

14. Rybski D, Reusser DE, Winz AL, Fichtner C, Sterzel T, Kropp JP. Cities as nuclei of sustainability? Environ Plan B. 2017; 44(3):425–440. https://doi.org/10.1177/0265813516638340

15. Gudipudi R, Rybski D, Lüdeke MKB, Zhou B, Liu Z, Kropp JP. The efficient, the intensive, and the productive: Insights from urban Kaya scaling. Applied Energy. 2019; 236(November 2018):155–162. https://doi.org/10.1016/j.apenergy.2018.11.054

16. CIESIN. Gridded Population of the World, Version 4 (GPWv4): Population Count.; 2016. Available from: http://dx.doi.org/10.7927/H4NP22DQ.

17. Doxsey-Whitfield E, MacManus K, Adamo SB, Pistolesi L, Squires J, Borkovska O, et al. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. Pap Appl Geogr. 2015; 1(3):226–234. https://doi.org/10.1080/23754931.2015.1014272

18. Oda T, Maksyutov S. ODIAC Fossil Fuel CO2 Emissions Dataset (version ODIAC2015a). Center for Global Environmental Research, National Institute for Environmental Studies, Downloaded from http://db.cger.nies.go.jp/dataset/ODIAC/. (2017/12/12).

19. Boden TA, Marland G, Andres RJ. Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., USA. 2016. https://doi.org/10.3334/CDIAC/00001_V2016

20. Elvidge CD, Baugh KE, Dietz JB, Bland T, Sutton PC, Kroehl HW. Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements. Remote Sens Environ. 1999; 68(1):77–88. https://doi.org/10.1016/S0034-4257(98)00098-4

21. Elvidge CD, Imhoff ML, Baugh KE, Hobson VR, Nelson I, Safran J, et al. Night-time lights of the world: 1994–1995. ISPRS J Photogramm Remote Sens. 2001; 56(2):81–99. https://doi.org/10.1016/S0924-2716(01)00040-5

22. Asefi-Najafabady S, Rayner PJ, Gurney KR, McRobert A, Song Y, Coltin K, et al. A multiyear, global gridded fossil fuel CO2 emission data product: Evaluation and analysis of results. Journal of Geophysical Research: Atmospheres. 2014; 119(17):10,213–10,231. https://doi.org/10.1002/2013JD021296

23. Janssens-Maenhout G, Crippa M, Guizzardi D, Muntean M, Schaaf E, Dentener F, et al. EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970-2012. Earth System Science Data Discussions. 2017;(August):1–55. https://doi.org/10.5194/essd-2017-79

24. Hutchins MG, Colby JD, Marland G, Marland E. A comparison of five high-resolution spatially-explicit, fossil-fuel, carbon dioxide emission inventories for the United States. Mitigation and Adaptation Strategies for Global Change. 2016. https://doi.org/10.1007/s11027-016-9709-9

25. Raupach MRR, Rayner PJJ, Paget M. Regional variations in spatial structure of nightlights, population density and fossil-fuel CO2 emissions. Energy Policy. 2010; 38(9):4756–4764. https://doi.org/10.1016/j.enpol.2009.08.021

26. Gurney KR, Mendoza DL, Zhou Y, Fischer ML, Miller CC, Geethakumar S, et al. High Resolution Fossil Fuel Combustion CO 2 Emission Fluxes for the United States. Environmental Science & Technology. 2009; 43(14):5535–5541. https://doi.org/10.1021/es900806c

27. Kakwani N, Wagstaff A, van Doorslaer E. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. J Econom. 1997; 77(1):87–103. https://doi.org/10.1016/S0304-4076(96)01807-6

28. Gini C. Measurement of Inequality of Incomes. Econ J. 1921; 31(121):124–126. https://doi.org/10.2307/2223319

29. Schubert J, Wolbring T, Gill B. Settlement Structures and Carbon Emissions in Germany: The Effects of Social and Physical Concentration on Carbon Emissions in Rural and Urban Residential Areas. Environ Policy Gov. 2013; 23(1):13–29. https://doi.org/10.1002/eet.1600

30. Centre for Cities. Cities Outlook 2015; 2015. Available from: http://www.centreforcities.org/wp-content/uploads/2015/01/Cities_Outlook_2015.pdf.

31. United Kingdom Government. 2014 Local Aauthority Carbon Dioxide Dmissions; 2016. Available from: https://www.gov.uk/government/statistics/local-authority-emissions-estimates.

32. Elvidge CD, Baugh K, Zhizhin M, Hsu FC, Ghosh T. VIIRS night-time lights. International Journal of Remote Sensing. 2017; 38(21):5860–5879. https://doi.org/10.1080/01431161.2017.1342050

33. Rayner PJ, Raupach MR, Paget M, Peylin P, Koffi E. A new global gridded data set of CO 2 emissions from fossil fuel combustion: Methodology and evaluation. Journal of Geophysical Research. 2010; 115 (D19):D19306. https://doi.org/10.1029/2009JD013439

34. Oda T, Maksyutov S, Andres RJ. The Open-source Data Inventory for Anthropogenic CO2, version 2016 (ODIAC2016): A global monthly fossil fuel CO2 gridded emissions data product for tracer transport simulations and surface flux inversions. Earth System Science Data. 2018; 10(1):87–107. https://doi.org/10.5194/essd-10-87-2018 PMID: 31662803

35. Pfähler W. Relative Concentration Curve: Functional Form and Measures of Non-Proportionality. Bulletin of Economic Research. 1985; 37(3):201–211. https://doi.org/10.1111/j.1467-8586.1985.tb00194.x

36. Lutz R, Spies M, Reusser DE, Kropp JP, Rybski D. Characterizing the development of sectoral gross domestic product composition. Phys. Rev. E 2013; 88(1):012804. https://doi.org/10.1103/PhysRevE.88.012804 PMID: 23944515

37. Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M. Constructing cities, deconstructing scaling laws. J R Soc Interface. 2014; 12(102):20140745. https://doi.org/10.1098/rsif.2014.0745

38. Rozenfeld HD, Rybski D, Andrade JS Jr, Batty M, Stanley HE, Makse HA. Laws of Population Growth. Proc Natl Acad Sci U S A. 2008; 105(48):18702–18707. https://doi.org/10.1073/pnas.0807435105 PMID: 19033186

39. Leitão JC, Miotto JM, Gerlach M, Altmann EG. Is this scaling nonlinear? R Soc Open Sci. 2016; 3(7). https://doi.org/10.1098/rsos.150649 PMID: 27493764

40. Rybski D. Relating urban scaling, fundamental allometry, and density scaling. arXivorg e-Print archive. 2016;arXiv:1609.01217 [physics.soc-ph].

41. Ribeiro HV, Rybski D, Kropp JP. Effects of changing population or density on urban carbon dioxide emissions. Nat Commun. 2019; 10:3204. https://doi.org/10.1038/s41467-019-11184-y PMID: 31324796