



**HAL**  
open science

# A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas

Quentin Lepetit, Vincent Viguié, Charlotte Liotta

► **To cite this version:**

Quentin Lepetit, Vincent Viguié, Charlotte Liotta. A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas. *Data in Brief*, 2023, 47, pp.108962. 10.1016/j.dib.2023.108962 . hal-04433873

**HAL Id: hal-04433873**

**<https://hal.science/hal-04433873>**

Submitted on 2 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contents lists available at [ScienceDirect](#)

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas



Quentin Lepetit<sup>a</sup>, Vincent Viguié<sup>a,\*</sup>, Charlotte Liotta<sup>a,b</sup>

<sup>a</sup> Centre International de Recherche sur L'Environnement et le Développement (CIRED), 45bis, Av de La Belle Gabrielle, Nogent-sur-Marne F-94736, France

<sup>b</sup> TU Berlin, Straße des 17. Juni 135, Berlin D-10623, Germany

## ARTICLE INFO

## Article history:

Received 29 September 2022

Revised 13 January 2023

Accepted 2 February 2023

Available online 9 February 2023

Dataset link: [A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas \(Original data\)](#)

## Keywords:

Cities

Urban form

Land cover

Urban economics

Real estate

Transportation

## ABSTRACT

This work presents a gridded dataset on real estate and transportation in 192 worldwide urban areas, obtained from the Google Maps API and the web scraping of real estate websites. For each city of the sample, these data have been associated with the corresponding population density and land cover data, extracted from the GHS POP and ESA CCI data respectively, and aggregated on a 1 km resolution grid, allowing for an integrated analysis. This dataset is the first to include spatialized real estate and transportation data in a large sample of cities covering 800 million people in both developed and developing countries. These data can be used as inputs for urban modeling purposes, transport modeling, or between-city comparisons in urban forms and transportation networks, and allow further analyses on e.g. urban sprawl, access to transportation, or equity in housing prices and access to transportation.

© 2023 Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

DOI of original article: [10.1016/j.regsciurbeco.2022.103832](https://doi.org/10.1016/j.regsciurbeco.2022.103832)

\* Corresponding author.

E-mail address: [viguie@centre-cired.fr](mailto:viguie@centre-cired.fr) (V. Viguié).

Social media: [@VincentViguie](#) (V. Viguié), [@CharlotteLiotta](#) (C. Liotta)

<https://doi.org/10.1016/j.dib.2023.108962>

2352-3409/© 2023 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

Subject	Economics.
Specific subject area	Urban form, Urban sprawl, Density, Real Estate, Land cover, Transportation studies, Environmental studies, Urban economics.
Type of data	Tables.
How the data were acquired	Transportation data have been obtained through the Google Maps and Baidu APIs, using Python. Real estate data have been obtained from the web scrapping of real estate websites, using Python. Population density data are taken from the GHS-POP dataset [1], and land cover data are taken from the ESA land cover CCI [2]. Data formatting and aggregation have been performed with QGIS and R.
Data format	Raw; Analyzed.
Description of data collection	A global sample of 192 cities has been selected to cover cities on all continents, while maximizing their diversity in terms of location, culture, or history. For each city, a 1km-resolution grid encompassing the urban area has been defined, on which all data (population density, land cover, transportation, and real estate), obtained as detailed in the previous sections, have been aggregated.
Data source location	CIREDE (center International de Recherche sur l'Environnement et le D��veloppement), Nogent-sur-Marne, France. Data on 192 cities worldwide. Sources of the real estate data and transportation data can be found in table <i>DataSources.xlsx</i> , available in the Zenodo repository. Population density and land cover data have been obtained from GHS-POP [1] and ESA Land cover CCI [2].
Data accessibility	Repository name: Zenodo. Data identification number: 6,821,394. Direct URL to data: <a href="https://zenodo.org/record/6821394">10.5281/zenodo.6821394</a>
Related research article	C. Liotta, V. Vigi��, Q. Lepetit, Testing the monocentric standard urban model in a global sample of cities, Reg. Sci. Urban Econ. 97 (2022) 103,832. <a href="https://doi.org/10.1016/j.regsciurbeco.2022.103832">10.1016/j.regsciurbeco.2022.103832</a> . [3]

## Value of the Data

- This dataset is the first dataset containing spatialized data on transportation and real estate in a large sample of cities in both developed and developing countries. With the corresponding land cover and density data, it allows working on city structures with a global perspective.
- These data can be used as inputs for urban modeling purposes. In particular, transportation, real estate, population density, and land cover data can be used for urban modeling in urban economics, as urban economics theories rely on the hypothesis that urban structures, and in particular housing supply and prices and population spatial distribution, depend on land use constraints and transportation costs. Transportation and population density data can also be used as inputs for transport modeling.
- These data can be used for between-city comparisons in urban forms and transportation networks, for instance in geographical studies and urban economics, and allow further analyses on e.g. urban sprawl, access to transportation, or equity in housing prices and access to transportation. These data can also be used as examples of differences between cities in urban forms and transportation networks for teaching purposes.
- By suggesting a standardized protocol, we make our dataset expandable to other countries and cities in the world, enabling reuses in transport, environmental or economic studies.

## 1. Data Description

The dataset is composed of four files (stored in both R data format .rds and Comma Separated Values .csv), which share a similar structure: each line corresponds to a pixel in a city. The first columns of each file are the same (Table 1), but the other columns are specific. The files are:

**Table 1**

Common records.

Variable	Type	Description
'ID'	[integer]	Spatial ID of the grid pixel.
'X'	[float]	X coordinates of the grid pixel in the city's corresponding UTM projection in meters.
'Y'	[float]	Y coordinates of the grid pixel in the city's corresponding UTM projection in meters.
'Area'	[float]	Pixel land area in m <sup>2</sup> .
'City'	[string]	City name.
'Country'	[string]	Country name.
'Continent'	[string]	Continent name.
'GridEPSG'	[integer]	Spatial Reference System EPSG code of the corresponding city. It gives the spatial projection used for the grid.
'dCenter'	[float]	Geographical distance to the Central Business District (CBD) in meters.

**Table 2**

Land cover data records (LandCoverData).

Variable	Type	Description
'OpenedToUrb'	[float]	Fraction of the grid cell that is unconstrained, i.e. opened to urbanization.
'ClosedToUrb'	[float]	Fraction of the grid cell that is constrained, i.e. closed to urbanization.
'ESACCI10' to 'ESACCI230'	[float]	ESA CCI land cover data in m <sup>2</sup> (see file EsacciReclassification.xlsx or land cover CCI user guide <a href="https://climate.esa.int/media/documents/CCI_Land_Cover_PUG_v2.0.pdf">https://climate.esa.int/media/documents/CCI_Land_Cover_PUG_v2.0.pdf</a> , accessed 08/10/2021).

**Table 3**

Population density data records (PopulationDensityData).

Variable	Type	Description
'PopDensitySource'	[string]	Population density source name.
'PopDensityYear'	[integer]	Population density year.
'PopDensity'	[float]	Population density inside the corresponding pixel for the corresponding year and population source.

**Table 4**

Transport data records (TransportData).

Variable	Type	Description
'TransportSource'	[string]	Source of the transport data.
'RushHour'	[string]	Night rush hour corresponding to the selected transport source and city.
'TransportYear'	[integer]	Year the transport data were collected.
'DistanceDriving'	[float]	Distance to the CBD during the city's rush hour, by private car, in meters.
'DurationDriving'	[float]	Travel time to the CBD during the city's rush hour, by private car, in seconds.
'DistanceTransit'	[float]	Distance to the CBD during the city's rush hour, by public transportation, in meters.
'DurationTransit'	[float]	Travel time to the CBD during the city's rush hour, by public transportation, in seconds.

- LandCoverData: Land cover data records (Table 2).
- PopulationDensityData: Population density data records (Table 3).
- RealEstateData: Real estate data records (Table 5).
- TransportData: Transport data records (Table 4).

Tables 1–5 present the variables in these files.

We provide two additional tables:

- *DataSources.xlsx* displays the detailed sources and collection dates of transport and real estate data, including the transport data source (column "Transport Data Source"), the rush hour at which the transport data have been collected (column "Rush Hour"), the real es-

**Table 5**

Real estate data records (RealEstateData). Note: suffixes “boxplotOutliers”, “percentilesOutliers” or “hampelOutliers” indicate the outliers’ exclusion method.

Variable	Type	Description
‘Currency’	[string]	ISO 4217 currency code of the corresponding country.
‘TransactionType’	[string]	Type of the transaction of corresponding observation. It can be ‘Rent’ or ‘Sale’.
‘TransactionSource’	[string]	Name of the data source.
‘TransactionMonth’	[string]	Month the data were collected.
‘TransactionYear’	[integer]	Year the data were collected.
‘avgSize’	[float]	Average size of dwellings inside the corresponding pixel, for the corresponding transaction, source, and date, in m <sup>2</sup> .
‘avgPriceSqm’	[float]	Average price per m <sup>2</sup> of dwellings inside the corresponding pixel, for the corresponding transaction, source, and date, in local currency.
‘medSize’	[float]	Median size of dwellings inside the corresponding pixel, for the corresponding transaction, source, and date, in m <sup>2</sup> .
‘medPriceSqm’	[float]	The median price per m <sup>2</sup> of dwellings inside the corresponding pixel, for the corresponding transaction, source, and date, in local currency.
RegPriceSqm	[float]	Price per m <sup>2</sup> of dwellings inside the corresponding pixel, for the corresponding transaction, source, and date, in local currency, obtained by regressing prices on dwelling sizes in the pixel.
‘nRealEstateData’	[integer]	Number of observations inside the corresponding pixel, for the corresponding transaction, source, and date.

tate website from which the real estate data have been collected (column “Real Estate Website”), and the real estate data scrapping date (column “RE data scrapping date”).

- *EsacciReclassification.xlsx* displays the nomenclature of the ESA CCI land cover data (column “ESA CCI land cover category”) and how these land cover data have been reclassified between “Opened to urbanization” and “Closed to urbanization” (column “Reclassification”).

## 2. Materials and Methods

### 2.1. Methods

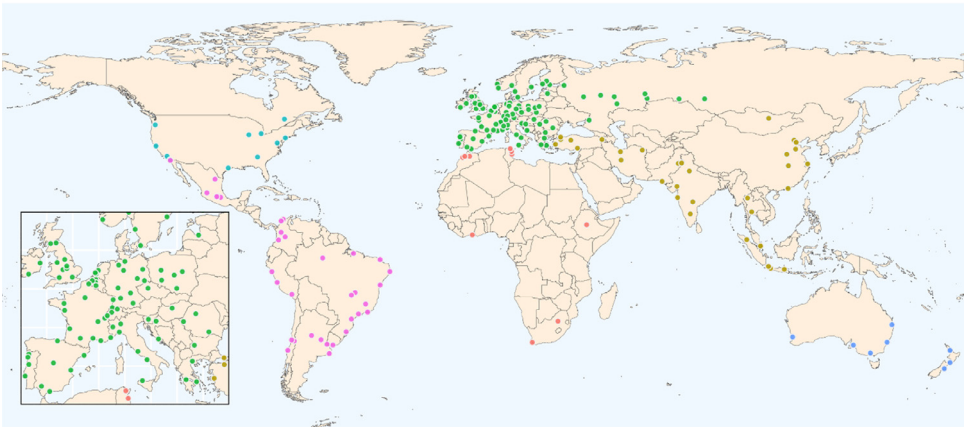
*Selection of the urban areas.* We have chosen the cities of our database following two criteria. First, we selected medium to large cities (with a population of over 300 000) in order to get a large share of the global urban population. Second, we selected cities of different cultural and historical backgrounds and tried to get a wide geographical coverage. In this way, we defined a first sample of 281 cities. However, data collection was possible in 192 cities of this sample (Fig. 1), mainly due to the availability of real estate data, as detailed in the next sections. Figure 4, in the supplementary material, presents the original 281 cities.

*Grid analysis.* For each city, we designed a georeferenced grid of a 1 km<sup>2</sup> resolution, encompassing the whole urban area (Fig. 2). We used these grids to aggregate land cover, population density, real estate, and transportation data at the same resolution.

*Land cover.* We used the European Space Agency land cover data, available worldwide at a 300 m spatial resolution on an annual basis from 1992 to 2015 [2]. These data allow to identify urbanized areas, as well as to distinguish constrained areas, i.e. locations where housing construction is impossible or difficult (water bodies,...) from unconstrained areas (see table *EsacciReclassification.xlsx* in the Zenodo repository).

*Population density.* We used the GHS-POP layers of the open and free GHSL (Global Human Settlement Layer) data of the European Commission, providing population counts per grid cell at a 250 m resolution worldwide and available for 1975, 1990, 2000 and 2015 [1,4].

- Africa    • Europe    • Oceania
- Asia     • North America    • South America



**Fig. 1.** Final sample (192 cities). This figure is taken from the related research article Liotta et al. (2022) [3].

*Real estate.* We have collected real estate data on rents, sale prices, and dwelling sizes by web scrapping real estate websites from 2017 to 2020. We selected these websites following four criteria:

- the website must have a nationwide coverage to ensure consistent results in each country,
- it must geolocalize the dwellings,
- it must have values for both rent or sale prices and dwelling sizes, and
- it has to be written in the local language and to propose prices in local currency to limit real estate ads targeting expatriates.

Real estate websites and scrapping dates can be found in table *DataSources.xlsx*, available in the Zenodo repository. Then, we aggregated the data and displayed them at the pixel level:

- we aggregated dwelling sizes by computing the mean and the median of data per pixel.
- we compute rents or sale prices per sqm by dividing total rents or sale prices by dwelling sizes, and then display the mean and the median per pixel. Alternatively, we regress total rents or sale prices by dwelling sizes for each pixel to find an estimate of rents or sale prices per sqm for each pixel.
- as a robustness check, we also display these data applying three outliers' exclusion methods on prices per sqm before the aggregation: boxplot (with a 1.5 coefficient), percentiles (excluding the top and bottom 2.5% values), and Hampel.

*Transportation.* We collected transport distances and durations to the city centers using Google Maps and Baidu Maps APIs (Application Programming Interfaces). Different methods have been used in the urban economics literature to identify city centers. Most rely on job density data [5–9], which are unfortunately not available on a consistent basis in our sample of cities. Therefore, we defined city centers by a compromise between five qualitative criteria: the geographical center of the data, the historical center of the cities, the location of public transports hubs, the official central business district, and the city hall location.

We collected transport data from the centers defined above to each grid cell at typical afternoon rush hours<sup>1</sup>. We collected, when available, both driving and public transport data. Trans-

<sup>1</sup> We have identified rush hours for each city by extracting a sample of transportation data at 16h, 16h30, 17h, 17h30, 18h, 18h30, 19h, 19h30, and 20h, by comparing the average delay due to congestion, and by defining rush hours as the moment when the average delay is the higher.

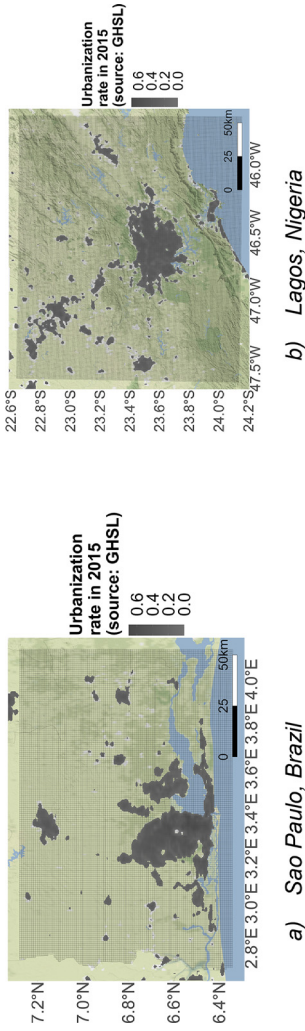


Fig. 2. Example of spatial analysis grids.

port data sources and rush hours can be found in table *DataSources.xlsx*, available in the Zenodo repository. It was not possible to collect transport data from each grid cell, so we collected data from 10% of all cells<sup>2</sup>, and then interpolated them using the *INTERPP* function from R package *AKIMA*.<sup>3</sup>

Code availability. Detailed code generating the database can be accessed from the source code hosted via Github at [https://github.com/CIREd/gridded\\_dataset\\_192\\_cities](https://github.com/CIREd/gridded_dataset_192_cities).

## 2.2. Technical Validation

*Land cover.* For land cover, we rely on the validation performed in the ESA CCI technical report [2]. Validating ESA CCI land cover data against GlobCover 2009 data, the authors found an overall accuracy of 71.45%, with the highest accuracies for rainfed cropland, irrigated cropland, broadleaved evergreen forest, and urban areas. The report mentions that other validation tests are undergoing.

*Population density.* For population density, we rely on the existing validations of the GHSL data. Validation tests of the GHSL built-up data, from which the GHSL population data are directly derived, are presented in the JRC technical report [11]. In this report, the GHSL built-up data are validated using two sources: a systematic field survey managed by EUROSTAT and a set of digital cartographic products with building footprints at a scale of 1:10,000 or better gathered from data portals of official websites of public governmental institutions. From comparing GHSL data with these two sources, total accuracies are 0.9628 and 0.8996 respectively. Other validation tests have been conducted by independent studies; for instance, comparing the GHSL built-up data with Baidu data in 20 Chinese cities, Liu et al. (2020) [12] find an R2 of 0.76.

The GHSL population data are more difficult to validate due to the lack of independent and comparable data. However, benchmarking using the GEOSTAT 2011 resident population data on 18 European countries, Freire et al. (2016) [4] found a correlation of 0.83.

*Real estate.* The quality of real estate prices, rents and dwelling size data differs from one city to another. A first source of error comes from the quantity of data that we have been able to collect. To assess the quality of the real estate data, we computed two variables at the city level: the market data cover, as the total population of the city divided by the number of ads, and the spatial data cover, as the number of pixels for which we have real estate data divided by the number of inhabited pixels. For market data cover, we found that in 95 cities out of 192, we have scrapped more than 1 rent ad per 1000 inhabitants, and that in 174 cities out of 192, we have scrapped more than 1 rent ad per 10,000 inhabitants. For spatial data cover, we found that in 109 cities out of 192, the spatial data cover of rent ads is above 10%, and that in 153 cities out of 192, the spatial data cover is above 5%. We report in our data the number of ads per pixel.

A second potential source of mistake is the systemic bias coming from our data sources. The websites that we scrapped present ads, which may not necessarily reflect the actual rents or prices, if margins of negotiations exist, for instance. They may also be biased and present only dwellings that are not representative of the actual dwelling stock. This may especially be the case if, in a city, online ad websites are not the main way of buying or renting a dwelling. To mitigate these risks, we tried to make sure that the websites we scrapped are actually used by locals.

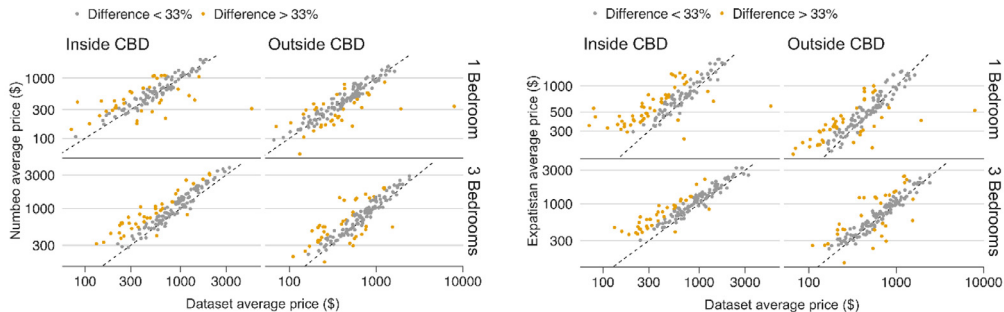
To assess these two risks, we checked the validity of our rents and real estate prices against 4 external databases providing averages per city. We used two crowd-sourced websites aiming at describing the cost of living across the world for expatriates, Numbeo [13] and Expatistan [14], to estimate rents and property prices around the world. We also used two databases built for

---

<sup>2</sup> With a method close to Saiz and Wang (2021) [10], we defined a star shape with 8 branches, centered on the city center, and we collected data from the grid cells at a regular distance on each star branch so that the total number of grid cells for which data are collected sums to 10% of the total number of grid cells.

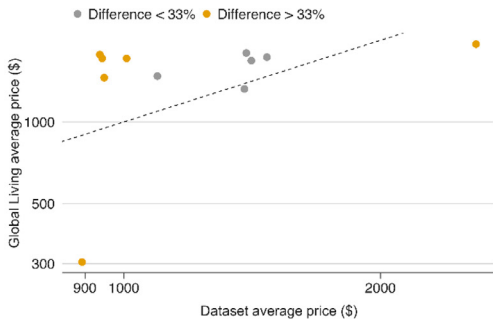
<sup>3</sup> <https://cran.r-project.org/web/packages/akima/index.html>



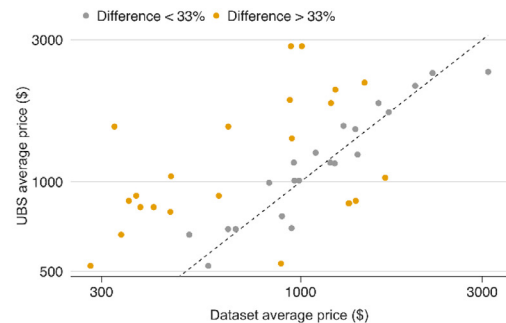


a) Comparison of our rent data with the rent data from the crowdsourced website Numbeo. For 73% of cities, the difference is lower than 33%. The  $R^2$  (on all the rent data, irrespective of the dwelling location or size) is 0.73.

b) Comparison of our rent data with the rent data from the crowdsourced website Expatistan. For 69% of cities, the difference is lower than 33%. The  $R^2$  (on all the rent data, irrespective of the dwelling location or size) is 0.69.



c) Comparison of our rent data with the Global Living (CBRE) database. For 5 out of the 11 cities in common between the databases, the difference is lower than 33%. The  $R^2$  is 0.15.



d) Comparison of our rent data with the UBS database. For 21 of the 44 cities in common between the databases, the difference is lower than 33%. The  $R^2$  is 0.34.

Fig. 3. Comparison of our database on rents with external databases for 2019.

real estate investors: UBS "Prices and Earnings" database [15], and data from CBRE, an expert in real estate and services based in London, whose data were obtained from various local sources [16]. When averaged over the cities, or over the inner core and the outer core of the cities, our data broadly agree with these data (Fig. 3).

*Transportation.* Overall, Google maps and Baidu maps data have several advantages compared with other transport data. Google maps data have a wide coverage compared with other data sources such as OpenStreetMap [19] and are regularly updated. Furthermore, they allow for measuring travel times and distances with and without congestion [20]. For these reasons, they have been used in recent research papers to compute travel times and distances [21].

However, the quality of Baidu and Google maps data is difficult to assess, as these websites use closed algorithms based on users' travel data [17,18]. It should be expected that the quality of these data is a function of the number of users in each city, and should therefore be higher in developed country cities and in large population cities [22–24]. The quality of travel data for public transport depends on whether transport authorities have shared their data with Google or Baidu. Another source of error comes from the number of data points that we use to measure transport times, and from the interpolation process. This source of error is also difficult to assess. We tried to mitigate it by using grid points close to each other near the center of the city, and further apart from each other when moving away from the center.

## Ethics Statement

- Terms of Service (ToS) and Copyright:
  - GHSL data are provided free of charge. Reuse is authorized, provided the source is acknowledged. Copyright notice: © European Union, 2022
  - The ESA CCI land cover data products are made available to the public by ESA and the consortium and can be used for scientific purposes without any fee on the condition that ESA Climate Change Initiative and its Land Cover project are credited. Copyright notice: © ESA Climate Change Initiative - Land Cover led by UCLouvain (2017).
  - At the time of data collection, the Terms of Service of the scrapped real estate websites were unclear regarding scrapping. However, UK,<sup>4</sup> France,<sup>5</sup> and European Union<sup>6</sup> legislation policies allow for data collection through web scrapping and use for non-profit research regardless of the ToS. The only shared data are aggregated data on dwelling sizes and rents or prices per 1km<sup>2</sup> grid cell so that raw data are not shared.
  - Google Maps allow for data collection with an API: <https://cloud.google.com/maps-platform/terms>.
- *Privacy:* All data have been anonymized by a spatial aggregation at a 1 km resolution, preventing from going back to the raw data.
- *Scrapping policy:* Scraping policies of the real estate websites were unclear at the time of the study. However, we limited ourselves to retrieving the dwelling sizes, when available, and rent or prices, and we avoided web scraping of websites employing any measures to limit/block scraping.

<sup>4</sup> <https://blogs.lse.ac.uk/impactofsocialsciences/2014/06/04/the-right-to-read-is-the-right-to-mine-tdm/> and <https://www.gov.uk/guidance/exceptions-to-copyright>, accessed on 29/9/2022.

<sup>5</sup> [<sup>6</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0790&from=FR#d1e961-92-1>, accessed on 29/9/2022.](https://urldefense.com/v3/__https://www.culture.gouv.fr/content/download/281379/file/CSPLA*20-*20Annexes-contributions*20-*20rapport*20exceptions*20fouille*20de*20textes*20et*20de*20donn**Aes.pdf?inLanguage=fre-FR_:_jSUJJSUjJSUjSXdqQ!NlFGqXoFfo8MMQ!79QRDGeuwrwv1ks5QSugxP8OiKbgN537Z3UiNglhhdSaG16ce2XuHejW9QANfAapg$, accessed on 29/9/2022.</a></p>
</div>
<div data-bbox=)

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

A gridded dataset on densities, real estate prices, transport, and land use inside 192 worldwide urban areas (Original data) (Zenodo).

## CRediT Author Statement

**Quentin Lepetit:** Conceptualization, Methodology, Investigation, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization; **Vincent Vigiúé:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition; **Charlotte Liotta:** Formal analysis, Data curation, Writing – review & editing, Visualization.

## Acknowledgments

Funding: This work was supported by the [Agence Nationale de la Recherche](#) through the project DRAGON (ANR-14-ORAR-005).

## References

- [1] M. Schiavina, S. Freire, K. MacManus, GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A, (2019). 10.2905/0C6B9751-A71F-4062-830B-43C9F432370F.
- [2] E.S.A. Land Cover CCI product user guide version 2. Tech. Rep., (2017). [maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2\\_2.0.pdf](https://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf).
- [3] C. Liotta, V. Vigiúé, Q. Lepetit, Testing the monocentric standard urban model in a global sample of cities, *Reg. Sci. Urban Econ.* 97 (2022) 103832, doi:10.1016/j.regsciurbeco.2022.103832.
- [4] S. Freire, E. Droxsey-Whitfield, K. MacManus, J. Mills, M. Pesaresi, Development of new open and free multi-temporal global population grids at 250m resolution, Geospatial Data in a Changing World, Association of Geographic Information Laboratories in Europe (AGILE), AGILE, 2016.
- [5] J.F. McDonald, The identification of urban employment subcenters, *J. Urban Econ.* 21 (1987) 242–258, doi:10.1016/0094-1190(87)90017-9.
- [6] G.M. Ahlfeldt, N. Wendland, How polycentric is a monocentric city? centers, spillovers and hysteresis, *J. Econ. Geogr.* 13 (2013) 53–83, doi:10.1093/jeg/lbs013.
- [7] D.P. McMillen, Nonparametric employment subcenter identification, *J. Urban Econ.* 50 (2001) 448–473, doi:10.1006/juec.2001.2228.
- [8] D.P. McMillen, S.C. Smith, The number of subcenters in large urban areas, *J. Urban Econ.* 53 (2003) 321–338, doi:10.1016/S0094-1190(03)00026-3.
- [9] C.L. Redfearn, The topography of metropolitan employment: identifying centers of employment in a polycentric urban area, *J. Urban Econ.* 61 (2007) 519–541, doi:10.1016/j.jue.2006.08.009.
- [10] A. Saiz, L. Wang, Physical Geography and Traffic Delays: Evidence from a Major Coastal City, Social Science Research Network, Rochester, NY, 2021, doi:10.2139/ssrn.3881405.
- [11] M. Pesaresi, D. Ehrlich, S. Ferri, A.J. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, V. Syrris, Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014, Joint Research Centre (European Commission), LU, 2016.
- [12] F. Liu, S. Wang, Y. Xu, Q. Ying, F. Yang, Y. Qin, Accuracy assessment of Global Human Settlement Layer (GHSL) built-up products over China, *PLoS One* 15 (2020) e0233164, doi:10.1371/journal.pone.0233164.
- [13] Numbeo Cost of Living Database. <https://www.numbeo.com/cost-of-living/> (accessed January 15, 2012).
- [14] Expatistan, cost of living comparisons, expatistan cost living comp. (n.d.). <https://www.expstatistan.com> (accessed January 15, 2021).
- [15] UBS, global cities ranking 2018 - UBS prices and earnings, UBS, 2019. <https://www.ubs.com/minisites/prices-earnings/en/intro/>.
- [16] CBRE Residential, Global Living 2019, CBRE, United Kingdom, 2019. <https://www.cbreresidential.com/uk/en-GB/content/global-living-2019>.
- [17] D. Wang, Stuck in traffic?, Google Off. Blog. (2007). <https://googleblog.blogspot.com/2007/02/stuck-in-traffic.html>.
- [18] D. Barth, The bright side of sitting in traffic: crowdsourcing road congestion data, Google Off. Blog (2009). <https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html>.

- [19] D.J. Weiss, A. Nelson, H.S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T.C.D. Lucas, R.E. Howes, L.S. Tusting, S.Y. Kang, E. Cameron, D. Bisanzio, K.E. Battle, S. Bhatt, P.W. Gething, A global map of travel time to cities to assess inequalities in accessibility in 2015, *Nature* 553 (2018) 333–336, doi:[10.1038/nature25181](https://doi.org/10.1038/nature25181).
- [20] F. Wang, Y. Xu, Estimating O–D travel time matrix by google maps API: implementation, advantages, and implications, *Ann. GIS* 17 (2011) 199–209, doi:[10.1080/19475683.2011.625977](https://doi.org/10.1080/19475683.2011.625977).
- [21] Z. Chang, J. Chen, W. Li, X. Li, Public transportation and the spatial inequality of urban park accessibility: new evidence from Hong Kong, *Transp. Res. Part Transp. Environ.* 76 (2019) 111–122, doi:[10.1016/j.trd.2019.09.012](https://doi.org/10.1016/j.trd.2019.09.012).
- [22] G. Kreindler, Driving Delhi? Behavioural responses to driving restrictions, *SSRN Electron. J.* (2016), doi:[10.2139/ssrn.2966797](https://doi.org/10.2139/ssrn.2966797).
- [23] A. Sudmant, V. Vigi  , Q. Lepetit, L. Oates, A. Datey, A. Gouldson, D. Watling, Fair weather forecasting? The shortcomings of big data for sustainable development, a case study from Hubballi-Dharwad, India, *Sustain. Dev.* 29 (2021) 1237–1248, doi:[10.1002/sd.2221](https://doi.org/10.1002/sd.2221).
- [24] M. Anderson, How to beat the traffic, *Science* 357 (2022) 36–37 (n.d.), doi:[10.1126/science.aan8177](https://doi.org/10.1126/science.aan8177).