

INSTITUTE  
OF ECONOMICS



Scuola Superiore  
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics  
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy  
ph. +39 050 88.33.43  
institute.economics@sssup.it

# LEM

## WORKING PAPER SERIES

**A unified repository for pre-processed climate  
data weighted by gridded economic activity**

Marco Gortan <sup>1</sup>  
Lorenzo Testa <sup>2,3</sup>  
Giorgio Fagiolo <sup>3</sup>  
Francesco Lamperti <sup>3,4</sup>

<sup>1</sup> School of Finance, University of St. Gallen, Switzerland

<sup>2</sup> Department of Statistics and Data Science, Carnegie Mellon University, USA

<sup>3</sup> Institute of Economics and L'EMbeDS, Sant'Anna School of  
Advanced Studies, Pisa, Italy

<sup>4</sup> RFF-CMCC European Institute on Economics and the Environment, Milan, Italy

**2023/45**

**December 2023**

**ISSN(ONLINE) 2284-0400**

# A unified repository for pre-processed climate data weighted by gridded economic activity

Marco Gortan<sup>1</sup>, Lorenzo Testa<sup>2,3</sup>, Giorgio Fagiolo<sup>3,\*</sup>, and Francesco Lamperti<sup>3,4,\*</sup>

<sup>1</sup>School of Finance, University of St. Gallen, St. Gallen, Switzerland

<sup>2</sup>Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh PA, US

<sup>3</sup>Institute of Economics and LEMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy

<sup>4</sup>RFF-CMCC European Institute on Economics and the Environment, Milan, Italy

\*Co-corresponding authors: Francesco Lamperti (f.lamperti@santannapisa.it); Giorgio Fagiolo (g.fagiolo@santannapisa.it)

## ABSTRACT

Although high-resolution gridded climate variables are provided by multiple sources, the need for country and region-specific climate data *weighted* by indicators of economic activity is becoming increasingly common in environmental and economic research. We process available information from different climate data sources to provide spatially aggregated data with global coverage for both countries (GADM0 resolution) and regions (GADM1 resolution) and for a variety of climate indicators (average precipitations, average temperatures, average SPEI). We weigh gridded climate data by population density or by night light intensity – both proxies of economic activity – before aggregation. Climate variables are measured daily, monthly, and annually, covering (depending on the data source) a time window from 1900 (at the earliest) to 2023. We pipeline all the preprocessing procedures in a unified framework, which we share in the open-access *Weighted Climate Data Repository* web app. Finally, we validate our data through a systematic comparison with those employed in leading climate impact studies.

**Keywords:** climate and weather data; spatial weighting; impact assessment; climate econometrics.

## Background & Summary

Climate change and weather events have been shown to adversely affect a wide spectrum of natural and socio-economic activities<sup>1,2</sup>. A blossoming body of literature reports evidence of significant and non-linear impacts on agricultural<sup>3</sup> and economic production<sup>4-6</sup>, conflict<sup>7</sup>, income inequality<sup>8</sup>, mortality<sup>9</sup>, energy consumption<sup>10</sup>, and the list is far from being conclusive. Most of these studies test the presence of a significant statistical association between climate variables and socioeconomic indicators, adopting either cross-section or panel-data approaches<sup>11,12</sup>.

One common challenge is that weather data are typically available at a much finer spatiotemporal resolution than socioeconomic variables. While indicators such as industrial production, GDP, employment, and fatalities are typically collected annually – at region or country breakdowns – temperatures, precipitations, and other weather variables are instead available at gridded levels and daily frequency. Hence, the common approach requires weather-related variables to be aggregated to match lower temporal frequencies and the geographical boundaries of administrative units.

This process is not straightforward and often requires the use of weights proxying the geographical distribution of economic activities. Indeed, when studying the impact of climatic conditions and weather events on the economy, it is crucial to account for the different exposure of socioeconomic activities within an administrative region. For example, average temperatures in the Mojave Desert (California, US) during the summer may be way higher than in Los Angeles (California, US), but the size of economic activities in the two locations is not even comparable. Indeed, one may easily argue that labor productivity in California is much more affected by temperatures in Los Angeles than in desert areas. Thus, a simple aggregation of climate data that does not account for the geography of socioeconomic activities could introduce a bias in the evaluation of climate impacts, especially when the variability across administrative regions is central to the identification of the effect<sup>11,12</sup>. Further, when a weather-related phenomenon occurs at regional level, in response to averaged weather, the weighting scheme is crucial to reflect the relative overall importance of weather in different regions. For instance, weighting rainfall by the distance from ashore could help to predict the declaration of states of emergency.

Spatially weighted data are increasingly employed in the literature exploring the impacts of climate change and weather events on socio-economic activities. For example, Burke et al.<sup>4</sup>, in a seminal study assessing the effect of global warming

on the dynamics of economic production, employ population-weighted temperatures and precipitations to measure gradual climate change. Accordingly, a number of studies have been relying on Burke et al. dataset to explore the impact of climate on economic inequality and growth<sup>8,13,14</sup>. Furthermore, population weighting is not limited to the case of average temperatures and total precipitations, as it is increasingly employed for a variety of additional climate indicators, e.g., in the evaluation of heating and cooling degree days<sup>15</sup>.

However, replicating published studies using spatially weighted climate data is difficult, as the exact procedure employed to obtain weighted climate variables used for impact assessment is often unclear, under-discussed, or not reported at all in existing contributions. This poses a potential problem, as the way in which weighting is performed may depend on a number of different key factors and choices<sup>16</sup>. Among them, the sources of data used for the construction of weights, the adjustments employed to align gridded information to the borders of administrative regions, and the eventual use of a base year are all elements that can sensibly affect the construction of spatially weighted climate indicators. This also undermines exercises trying to employ existing datasets containing spatially-weighted climate variables (e.g., made available in online repositories as supplementary material of published papers) in further studies or analyses. Indeed, in absence of clear guidelines and documentation, it becomes very hard to build homogenized datasets covering different sets of countries or regions and longer time series (i.e., more recent years).

Here, we argue that the lack of a harmonized, documented, cross-validated, and open-access repository for climate variables that are spatially weighted by economic activity hinders a rigorous and robust estimation of the social and economic impacts of climate change. This may partly explain why unweighted climate indicators are still employed in several studies. For example, Kotz et al.<sup>6</sup> construct a number of indicators proxying the yearly distribution of rainfall within national and subnational regions without accounting for the spatial distribution of economic activities, and use such indicators to show the adverse impact of precipitation extremes on economic growth. Furthermore, spatially unweighted climate data are also employed in the emergent macro-econometric literature on climate impacts<sup>17–20</sup>.

In this paper, we try to close this gap by introducing a unified repository that pipelines the preprocessing and weighting procedures of gridded climate data into a documented, intuitive, and open-access interface. The repository allows researchers to get ready-to-use climate variable datasets aggregated at national and sub-national levels, with global coverage over the period 1900–2023. In particular, the *Weighted Climate Data Repository* provides a user-friendly dashboard to explore and download key climate variables under customizable weighting schemes, temporal frequency, timeframe, administrative level, and file format.

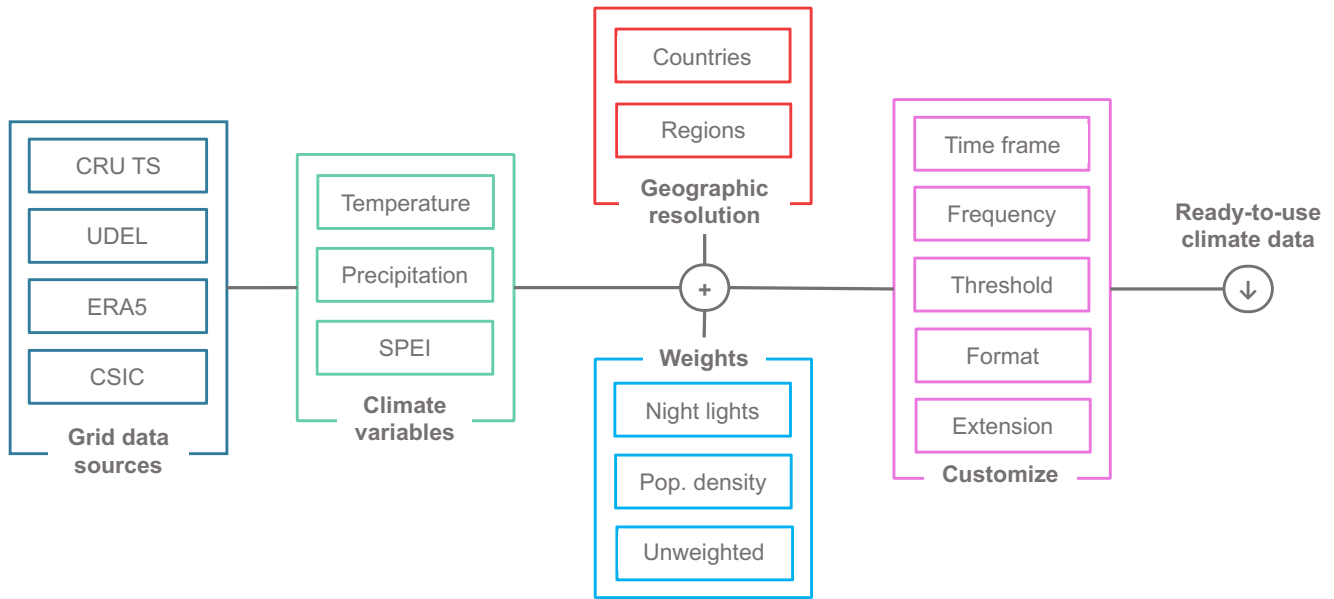
Our repository is intended to support the climate impact assessment community, which is constantly enlarging and increasingly opening to scientists and researchers who aim to work with datasets compiled at the administrative level (e.g., economists and public policy scholars). Indeed, by offering a unified and harmonized access to a wealth of publicly available yet dispersed and unweighted climate and weather indicators, we aim to improve the replicability of impact assessment studies, increase the transparency of data management practices, and incentivize the community to test the robustness of estimates to the choice of data sources and aggregation strategies. The *Weighted Climate Data Repository* is available and maintained at <https://weightedclimatedata.streamlit.app>.

## Methods

The logical steps behind the construction of our repository are illustrated in Figure 1. We combine different datasets, including gridded climate variables from multiple open-access sources, gridded indicators of spatial socio-economic activity, and administrative boundaries at different levels of resolution. The main objective is to obtain climate data that are weighted by socio-economic indicators according to different strategies that are customizable by the user. To achieve this, our procedure follows three key steps:

- **Selection:** In the first step, we choose (i) a specific set of gridded climate variables of interest, (ii) the desired geographical resolution, and (iii) a gridded economic activity indicator for constructing the aggregation weights.
- **Computation of weights:** Next, we integrate the selected information to derive a gridded weighted version of each climate variable. This process ensures that the socio-economic indicators are appropriately considered in the analysis.
- **Aggregation:** Finally, we aggregate the gridded weighted observations across the regions defined by the chosen geographical resolution. This step allows us to obtain a comprehensive view of climate data at the desired level of granularity.

Our repository includes an interactive interface that enables users to customize the aggregation process and the format of downloadable datasets. They can modify parameters such as the base year for constructing weights, the frequency of climate data (i.e., daily, monthly, yearly), and the time span of interest. Additionally, users can query the database to access specific information tailored to their end-use requirements.



**Figure 1.** The *Weighted Climate Data Repository* workflow. Users can combine gridded climate variables, gridded indicators of economic activity, and administrative boundaries to achieve regional climate variables *weighted* by economic activity.

### Gridded variables and administrative boundaries

The core of the *Weighted Climate Data Repository* rests on two groups of gridded variables: climate variables and indicators of economic activity. These variables, together with administrative boundaries, serve as the fundamental components of our repository. Table 1 shows all the sources of data we exploit in our work.

#### Climate data

We leverage raw gridded climate data from four sources that are routinely used in climate impact studies: Climate Research Unit Time-Series<sup>21</sup> (CRU TS v4.07, available from 1901 until 2022), Consejo Superior de Investigaciones Científicas<sup>22</sup> (CSIC v2.7, 1901–2020), ECMWF Reanalysis v5<sup>23</sup> (ERA5, 1940–2023), and University of Delaware<sup>24</sup> (UDEL v5.01, 1900–2017). CRU TS, UDEL, and CSIC provide data at the grid resolution of  $0.5^\circ \times 0.5^\circ$ , while data from ERA5 feature a finer resolution ( $0.25^\circ \times 0.25^\circ$ ). Each source offers *monthly* records for two climate indicators, namely average temperatures (measured in Celsius degrees,  $^\circ\text{C}$ ) and total precipitations (in millimeters, mm), with the exception of CSIC, which provides monthly records for a third climate variable, the Standardized Precipitation-Evapotranspiration Index<sup>25</sup>, also known as SPEI (unit free). In addition to monthly data, ERA5 also provides records at the temporal resolution of *hours*, which we aggregate to obtain *daily* values.

CRU TS employs raw data from an extensive network of weather stations, computes monthly climate anomalies, and interpolates them using angular-distance weighting<sup>21</sup> (ADW). ADW is employed to account for the varying area represented by each grid cell on a spherical Earth, in particular by considering the cosine of the latitude of each grid cell. The cosine of the latitude serves as a measure of the change in grid cell area with respect to latitude. Cells near the equator have larger areas as compared to those near the poles, where cells are smaller.

CSIC leverages CRU TS data to provide the SPEI, a drought index that combines information from both precipitation and evapotranspiration to assess the severity and duration of drought conditions. It is a standardized version of the widely used Palmer Drought Severity Index (PDSI) that takes into account the effects of both precipitation and temperature on water availability. Given its multi-scalar nature, it is able to differentiate among different types of drought; we currently propose the 1-month level of aggregation, focusing on changes in headwater levels.

ERA5 climate data set uses data from radiosondes, which are battery-powered telemetry instruments carried into the atmosphere by weather balloons to measure various atmospheric parameters, including temperature, wind, and humidity profiles. The information collected by radiosondes is transmitted back to the ground via radio signals and is assimilated by ERA5 along with other observations, such as satellite and surface-based measurements, to provide a comprehensive picture of the Earth's climate system<sup>23</sup>.

Finally, UDEL provides gridded estimates mainly based on station records compiled from several publicly available sources (e.g., Global Historical Climatology Network dataset<sup>26</sup>, Global Historical Climatology Network Monthly dataset<sup>27</sup>, the Daily



Global Historical Climatology Network archive<sup>28</sup>). Interpolation is performed with Shepard spatial-interpolation algorithm<sup>29</sup>, modified for use over Earth's near-spherical surface.

**Table 1.** Summary of the main features of the employed data sources

Source	Reanalysis <sup>1</sup>	Variables	Coverage period	Resolution	Version
CRU TS <sup>21</sup>	No	Temperature, precipitation	1901–2022	0.5°	4.07
CSIC <sup>22</sup>	No	SPEI 1-month	1901–2020	0.5°	2.7
ERA5 <sup>23</sup>	Yes	Temperature, precipitation	1940–2023	0.25°	5
UDEL <sup>24</sup>	No	Temperature, precipitation	1900–2017	0.5°	5.01
NASA <sup>30</sup>		Population density	2000, 2005, 2010, 2015	0.25°	4
Li et al. (2020) <sup>31</sup>		Night-light intensity	2000, 2005, 2010, 2015	0.0083°	7
GADM <sup>32</sup>		Administrative boundaries			4.1

### Socio-economic data

We use gridded socio-economic data to gauge information on the spatial distribution of economic and human-based activities. In particular, two distinct indicators are used as weights for the spatial aggregation of climate data into administrative units. The first proxy is population density, available from Columbia University's Gridded Population of the World v4<sup>30</sup>, measured at 0.25° and 0.5° spatial resolutions. The climate econometrics literature has largely employed population density as an indicator of economic activity proxying local exposure to weather conditions<sup>4,11,12</sup>. Note that population density is measured with respect to the land area of each grid. Thus, in our aggregation strategy, we employ the product between the population density and the area of the associated grid to account for population size properly.

A second, alternative indicator of economic activity that we include in our repository is night-time light data<sup>31</sup>, which is originally available at a 30 arc-second spatial resolution (0.0083°). To match this finer resolution with the coarser resolutions of our gridded climate data, we compute the mean of the values of the cells in the 0.25° and 0.5° grids. We aggregate by first taking the mean of 900 (30×30) and 3600 (60×60) most upper-left cells in our coordinate system to produce a single grid at a resolution of, respectively, 0.25° and 0.5°. We then iterate this procedure with the adjacent blocks of cells to obtain all the gridded values of the night-light data for the coarser resolution. We note that the harmonized VIIRS-DMSP `tif` file for the year 2015 presented noise from auroras, especially in the northern hemisphere (see Figure 2, left panel). Therefore, following a standard procedure<sup>31</sup>, we set to 0 the radiance linked to grids above the 45 North and below the 45 South parallels when they had null values in 2000, 2005, and 2010. Figure 2, right panel, shows the result of this correction.

We allow weighting by either population or night-lights using the base years 2000, 2005, 2010, and 2015<sup>2</sup>. Moreover, the repository allows to explore and/or download climate data without weighting them by any spatial economic indicator. This option is referred to as “unweighted”.

### Administrative boundaries

We employ two levels of geographical resolution from the Database of Global Administrative Areas<sup>32</sup> (GADM). While the first level (GADM0) has a coarser resolution and replicates country boundaries, the second level (GADM1) is sub-national and consists of the largest administrative area included within national countries (e.g., states for the US, regions for Italy, etc.). In our work, we used GADM version 4.1 released on July 16, 2022.

### Weighting and aggregation strategy

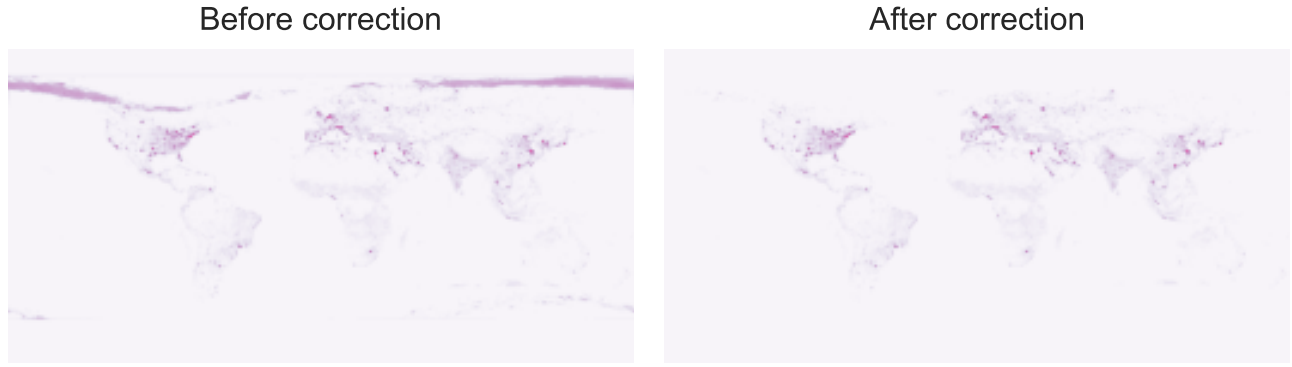
Raw grid data require to be aggregated to match administrative areas, for which many other socioeconomic indicators are usually available. The general weighting scheme is the following:

$$y_{i,t,w,T} = \frac{\sum_{j \in J_i} a_j f_{i,j} w_{j,T} x_{i,t}}{\sum_{j \in J_i} a_j f_{i,j} w_{j,T}} \quad (1)$$

where  $y_{i,t,w,T}$  is the value of the climate variable  $y$  in the geographical unit  $i$  (at a specified GADM resolution) at time  $t$  weighted by proxy  $w$  measured in base year  $T \in \{2000, 2005, 2010, 2015\}$ ;  $J_i$  is the set of grids intersecting the geographic unit  $i$ ;  $f_{i,j}$  is the fraction of grid  $j$  which intersects the geographic unit  $i$ ;  $a_j$  is the area of the grid  $j$ ;  $x_{i,t}$  is the raw grid climate variable. In

<sup>1</sup>Reanalysis refers to the integration of climate models with past observations to provide (i) consistent values over time and (ii) more accurate estimates in the grids not covered by measurement stations.

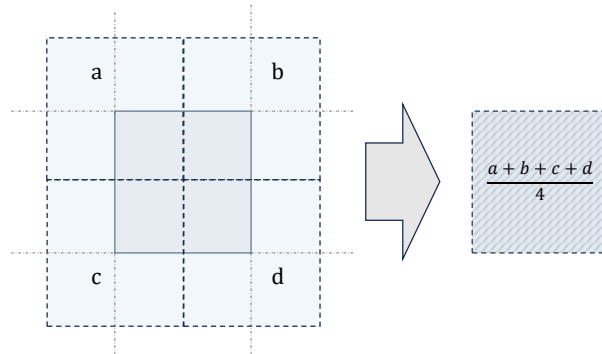
<sup>2</sup>The code needed to replicate analysis with additional base years is available in the repository associated with the paper.



**Figure 2.** Correction of auroras in the night lights data for the year 2015. The left plot shows night light data *before* correction; the right plot shows the same data *after* correction, which consists of setting to 0 the radiance linked to grids above the 45 North and below the 45 South parallels in case the value of the grids were equal to 0 in the night light data of 2000, 2005, and 2010, which were not affected by the auroras issue.

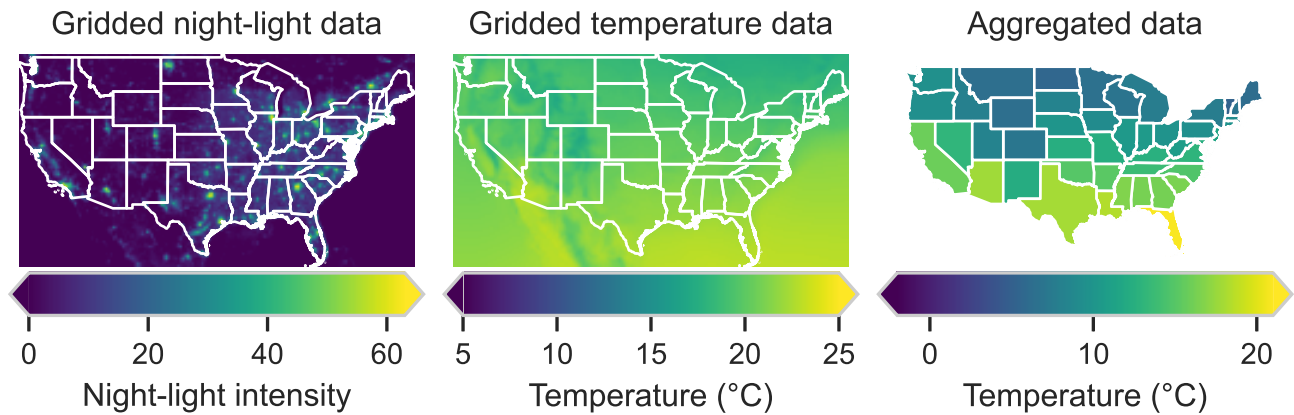
line with the prevailing practice in the literature, the base year  $T$  is fixed ex-ante and does not vary with  $t$ <sup>4,12</sup>. Of course, for the unweighted aggregation,  $w_{j,T} = 1$  for any  $j$  and  $T$ .

We notice that grid resolutions may vary across data sources. The NetCDF files retrievable from ERA5 is made up of a  $721 \times 1440$  grid, with extremities ( $180.125^\circ\text{W}$ ,  $179.875^\circ\text{E}$ ,  $90.125^\circ\text{S}$ ,  $90.125^\circ\text{N}$ ), with a 15 arc-minute spatial resolution. The gridded file of the population density and night-time lights feature instead a  $720 \times 1440$  grid, with extremities ( $180^\circ\text{W}$ ,  $180^\circ\text{E}$ ,  $90^\circ\text{S}$ ,  $90^\circ\text{N}$ ). To make weighting and climate variables from ERA5 consistent with population data, we resampled the values of the weight grids with a simple bilinear interpolation. The logic behind such a procedure is sketched in Figure 3, where the stylized grids of two sources are displayed. This procedure is applied whenever we weigh climate variables from ERA5 with population density and night-time light grid files.



**Figure 3.** Stylized illustration of the bilinear interpolation when the population density and night-time lights grids are used to weigh ERA5 climate variables. The extent of the weighting grids slightly differs from the extent of the ERA5 climate variables grids, both in longitude and latitude, resulting in a difference of  $0.125^\circ$  in both directions, exactly half of the spatial resolution of the ERA5 data. Since our weighting procedure requires weighting and variable grids to overlap, we resample the weighting grids, filling the values with a simple average of the values of the intersecting grids.

As an example of the aggregation strategy, Figure 4 shows three panels. The left and center panels display raw gridded data for night-light intensity in 2015, and ERA5 average annual temperatures in 2015 for the contiguous US, respectively. Nightlights are chosen as the weighting variable in this example. Figure 4, right panel, displays the resulting aggregation at GADM1 resolution and illustrates the output that users can retrieve from our repository.



**Figure 4.** Example of climate data weighting for the US. The left panel shows raw gridded night-light data in 2015. The middle panel displays raw gridded temperature data in 2015. Finally, the right panel shows, for the year 2015, temperatures aggregated at the GADM1 administrative level *weighted* by night lights in 2015.

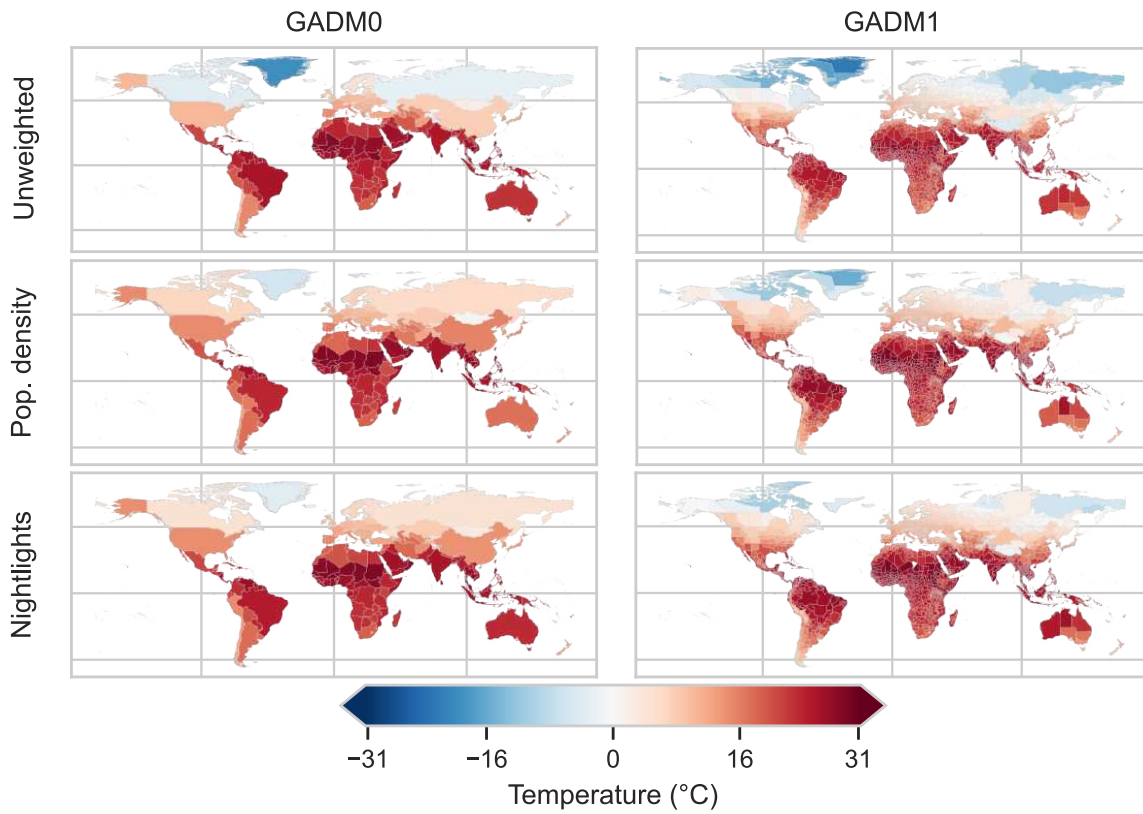
### Customization

Once aggregated, data can be customized in different ways. First, users can select a time interval by defining specific initial and final time boundaries. Second, it is possible to choose among different time frequencies. As mentioned before, daily data are only provided by ERA5. Notice also that weighted values for daily and monthly observations are computed using raw data at the same temporal resolution from the original sources. Annual observations are instead obtained by aggregating monthly to annual observations, via the computation of relevant summary statistics (i.e., average for temperatures, sum for precipitations), without seasonal adjustments. The only exception is the SPEI variable, for which daily data are not available and for which a simple average is not meaningful (i.e., SPEI cannot be linearly aggregated). As a result, we provide SPEI only at a monthly temporal resolution. Third, exploiting daily data provided by ERA5, the repository provides information about the frequency of extreme weather conditions. In particular, users can specify either an absolute or a relative threshold for the climate variable of interest. The repository then returns the number of days for which the climate variable has attained values over that threshold, for each geographical unit, month or year. For example, if the users sets for temperature an absolute threshold of 20°C, the web app returns the number of days for which the average temperature of the geographical unit has exceeded the threshold within the chosen month or year. Similarly, users can set a relative (quantile) threshold of 0.90. In this case, the web app returns the number of days for which the average temperature of the geographical unit has fallen in the top 10% percentile within the chosen month or year. For each geographical unit, the percentile is computed on its historical distribution, i.e. the distributions of temperatures in that region or country *regardless* time measurement.

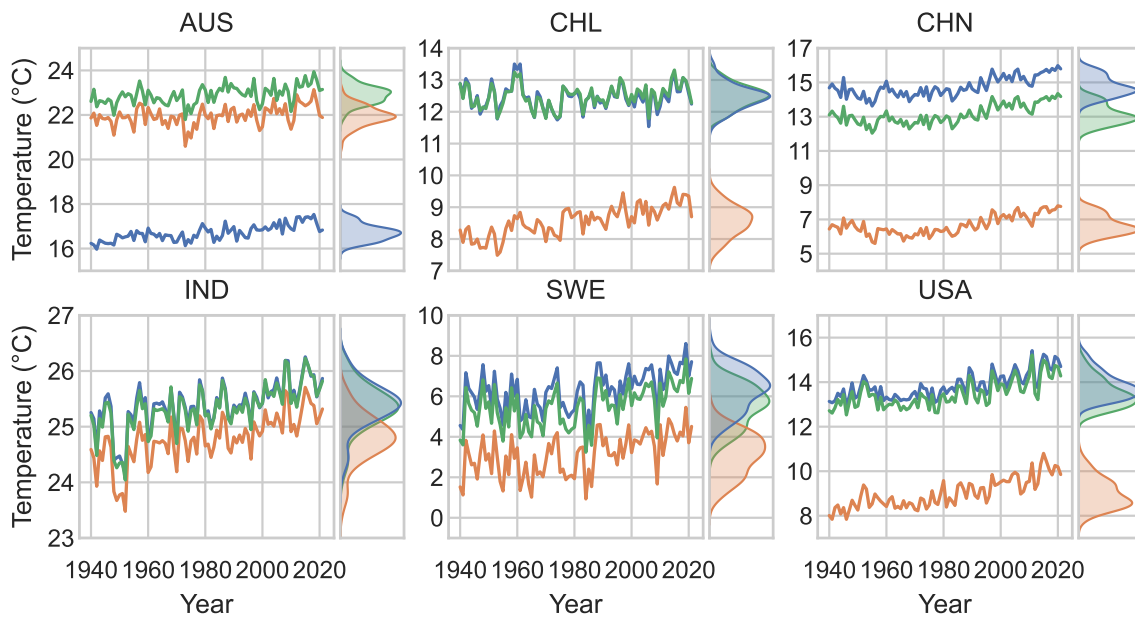
### Data Records

Our repository contains 138 data sets, each referring to a specific combination of geographical resolution (GADM0, GADM1), climate variable (temperature, precipitation, SPEI), climate data source (CRU TS, UDEL, ERA5, CSIC), weighting variable (unweighted, population density, night-lights), time resolution (daily, monthly), and weighting base year (2000, 2005, 2010, 2015). Data can be downloaded from <https://weightedclimatedata.streamlit.app> in two different formats: the wide format has geographical units as keys and values of a climate variable in different years as attributes; the long format has geographical units and years as keys, and the value of a climate variable as only attribute. Data can also be downloaded in three different extensions (csv, json, and parquet).

The results obtained from the pipelines described in the previous section are exemplified and illustrated in Figures 5 and 6. Figure 5 shows World maps at two different geographical resolution levels: GADM0 (country) and GADM1 (sub-national). These maps illustrate the ERA5 temperature levels in year 2015, considering various weighting variables. The maps provide a visual representation of how temperature values vary across countries and regions after being weighted by different indicators of economic activity. It is immediately visible how weighting for proxies of local anthropogenic activity returns a different picture of cross-country differences in experienced average temperature as compared to unweighted data, especially for areas at extreme latitudes. Further, while population and night-light weights provide relatively similar average temperatures in 2015 at country level, remarkable differences are detectable when scaling at the sub-national level, especially in Latin America, Central and Southern Africa, and Australia, which are areas where economies are among the most exposed to climate damages<sup>4,8</sup>.



**Figure 5.** Comparison among different weighting and aggregation schemes; ERA5, temperatures in 2015, weighting base year 2015.



**Figure 6.** Time series of temperatures from ERA5 for selected countries (from top-left to bottom-right: Australia, Chile, China, India, Sweden, USA), weighted by population density (blue), night lights (green) in 2015, or unweighted (orange).

Taking a different perspective, Figure 6 shows the time series of ERA5 temperatures weighted by different indicators of local economic activity for six selected countries from 1940 to 2020. The six panels of Figure 6 demonstrate how temperature values, when aggregated under different weighting schemes, fluctuate over time within each country. While for some countries the choice of the weighting and aggregation scheme is almost irrelevant (e.g., India; bottom-left panel), in others it provides very different average temperature levels (e.g., Australia; top-left panel) and long-run dynamics (e.g., Chile; top-center panel).

## Technical Validation

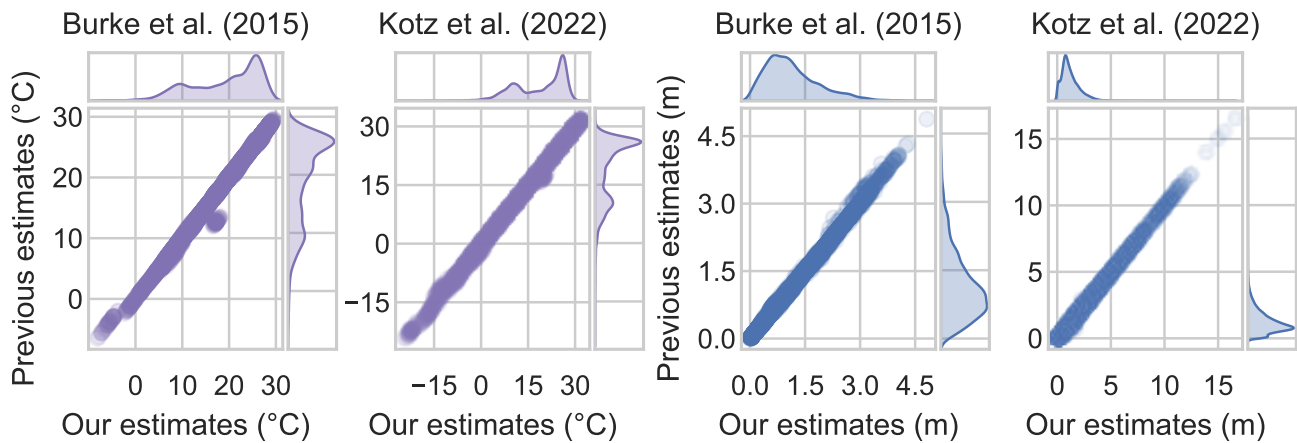
In this section, we validate the datasets constructed in our repository against those employed in two influential climate econometric exercises: Kotz et al.<sup>6</sup> and Burke et al.<sup>4</sup>. We evaluate the agreement between our weighting procedures and those obtained by these two studies, with the aim of supporting the reliability and effectiveness of our approach.

In order to conduct a proper validation exercise, we first align our data sources with the exact versions employed by the two targeted studies, which of course have been employing older versions for both climate and economic activity datasets. This allows us to validate the accuracy and robustness of our data processing pipelines and methods and to ensure a fair and reliable assessment of the quality and consistency of our estimates.

More precisely, Burke et al.<sup>4</sup> exploit UDEL v3.01 for precipitation and temperature data, and v3 of the NASA 0.50° gridded population data in 2000. Population is used as the weighting variable and, although the authors do not specify the source and version of the national administrative boundaries they use, their shape files are publicly available. Conversely, Kotz et al.<sup>6</sup> use 0.25° gridded ERA5 precipitation and temperature data, do not weigh climate data with any indicator of economic activity, and employ GADM1 v3.6 for the spatial aggregation.

Results of our comparative analysis are reported in Figure 7. The figure includes four scatterplots, each representing the relationship between our estimates and those used in the original studies for both temperature and rainfall (SPEI is not used in either of the two). Intuitively, points aligning on the main diagonal of the scatterplots indicate agreement and reflect the similarity between the estimates.

It is important to note that the data shown in Figure 7 encompass all the years analyzed in the original studies. Notably, a substantial majority of our estimates exhibit a high degree of correspondence with the weighted and/or aggregated data employed by previous authors. This indicates a strong level of agreement between our results and those of the earlier studies, corroborating the quality and reliability of the methods employed in our repository. However, there also emerge some minor discrepancies that are worth pointing out. In particular, the first panel on the left highlights two main sources of disagreement between the estimates of Burke et al. and ours. The first one, on the bottom left (where both temperatures are negative), regards Greenland. In this case, the estimates of Burke et al. are less conservative than ours. The second one, where our estimates are instead slightly more generous than the ones by Burke et al., concerns Bhutan. These discrepancies are mainly due to the weighting scheme, and in particular to the fact that population density is highly concentrated in a few regions of Greenland and Bhutan.



**Figure 7.** Comparison of weighted and/or aggregated temperature and precipitation variables in our datasets against data used in Burke et al. (countries) and Kotz et al. (sub-national regions). Average yearly temperature is expressed in degrees Celsius while annual total precipitations are in meters. Values on the main diagonal indicate very similar estimates.



## Usage Notes

The *Weighted Climate Data Repository* dashboard can be accessed at <https://weightedclimatedata.streamlit.app>. The homepage provides an overview of the main features of the web app, introducing users to its visualization and downloading features, which can be experienced by clicking on the respective links on the left-sidebar (see Figure 8, top panel, for an illustration).

The visualization tab, shown in Figure 8 (central panel), provides an easy interface with plotting tools. In particular, users can choose among different options (i.e. climate variables, variable sources, geographical resolutions, weighting indicators, weight base years, threshold options, time frequencies) and filters (i.e. starting and ending years, observations) to produce two kinds of plots in real-time: time series and choropleth maps.

Finally, Figure 8 (bottom panel) shows the download tab. Similarly to the visualization tab, users can customize the files they want to download through several options and filters. In addition to the previous tab, users can also select the data format and extension, as discussed above. Before downloading the data, the dashboard provides a preview option, facilitating the overall user experience. Metadata containing source versions can be downloaded as well.

## Future extensions

We envision several extensions to our current work. By keeping the repository updated to the latest versions of the available data providers, we first aim at covering longer time spans. We also plan to include higher spatial resolution geographical observations, such as GADM2, enhancing the spatial detail of data offered by our repository. Furthermore, we aim to introduce more climate variables and more upscaling summary statistics, enriching the range of possible derived climate indicators generated by the repository. Under this respect, we also plan to extend also the SPEI climate variable to other scales, up until the 48-month variant, which is the longest available.

## Code availability

Python code running the *Weighted Climate Data Repository* dashboard and scripts for aggregating data are available at <https://github.com/CoMoS-SA/climatedepo>. The *Weighted Climate Data Repository* leverages Streamlit. We employed R to process the data, exploiting package `exactextract` (version 0.9.1) for the weighted aggregations.

## References

1. Dell, M., Jones, B. F. & Olken, B. A. What do we learn from the weather? The new climate-economy literature. *J. Econ. literature* **52**, 740–798 (2014).
2. Carleton, T. A. & Hsiang, S. M. Social and economic impacts of climate. *Science* **353**, aad9837 (2016).
3. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proc. Natl. Acad. sciences* **106**, 15594–15598 (2009).
4. Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).
5. Kalkuhl, M. & Wenz, L. The impact of climate conditions on economic production. Evidence from a global panel of regions. *J. Environ. Econ. Manag.* **103**, 102360 (2020).
6. Kotz, M., Levermann, A. & Wenz, L. The effect of rainfall changes on economic production. *Nature* **601**, 223–227 (2022).
7. Abel, G. J., Brottrager, M., Cuaresma, J. C. & Muttarak, R. Climate, conflict and forced migration. *Glob. environmental change* **54**, 239–249 (2019).
8. Palagi, E., Coronese, M., Lamperti, F. & Roventini, A. Climate change and the nonlinear impact of precipitation anomalies on income inequality. *Proc. Natl. Acad. Sci.* **119**, e2203595119 (2022).
9. Carleton, T. *et al.* Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Q. J. Econ.* **137**, 2037–2105 (2022).
10. Auffhammer, M. & Mansur, E. T. Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Econ.* **46**, 522–530 (2014).
11. Hsiang, S. Climate econometrics. *Annu. Rev. Resour. Econ.* **8**, 43–75 (2016).
12. Auffhammer, M. Quantifying economic damages from climate change. *J. Econ. Perspectives* **32**, 33–52 (2018).



**Home**

Explore Data

Download Data

Documentation

Contacts

Institute of Economics and L'EMbeDS Department

Sant'Anna School of Advanced Studies (Pisa, Italy)

## Welcome to the Weighted Climate Data Repository Dashboard!

The Weighted Climate Data Repository (WCDR) project provides a user-friendly dashboard to explore and download climate data weighted by measures of economic activity. The dashboard allows users to:

- Choose to retrieve raw climate data about temperature, precipitation and SPEI (Standardized Precipitation-Evapotranspiration Index) from different sources and geographical/time resolutions;
- Weight raw climate data using alternative gridded measures of economic activity;
- Explore and download the resulting dataset in a single, accessible, and organized repository in a flexible way.

The project is run within the [Institute of Economics](#) and [L'EMbeDS](#) at [Sant'Anna School of Advanced Studies](#) (Pisa, Italy) by [Giorgio Fagiolo](#), [Marco Gortan](#), [Francesco Lamperti](#) and [Lorenzo Testa](#).

## The Dashboard

Within the WCDR dashboard, the user can choose:

- Climate variable:** available climate variables are temperature and precipitation (daily, monthly and annual observations) and SPEI (standardized precipitation-evapotranspiration index; monthly observations);
- Data source:** we currently support data from [Climatic Research Unit \(CRU TS\)](#), [Delaware Climate Office](#), [ECMWF's ERA5](#) and [CSIC](#);
- Geographical resolution:** data can be downloaded at both the [GADM](#) geographical resolution of GADM0 (World countries) and GADM1 administrative areas (i.e. regions within World countries);
- Weighting type:** we provide the possibility of weighting climate data by measures of economic activity linking grid data to administrative units. There are currently three options: (i) **no weights** (i.e.,

### Explore Data

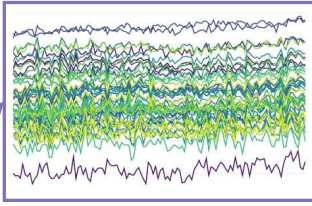
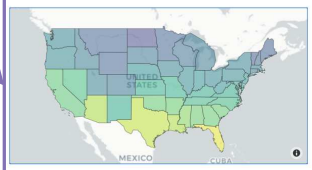
Climate variable: **temper...** Variable source: **CRU TS** Geographical resolution: **gadm0** Weighting variable: **populati...** Weighting year: **2015**

Threshold: **False**

Time frequency: **monthly**

Starting year: **1951** Ending year: **1952**

Countries: **United States**

### Download Data

Climate variable: **temper...** Variable source: **CRU TS** Geographical resolution: **gadm0** Weighting variable: **populati...** Weighting year: **2000**

Threshold: **False**

Time frequency: **yearly**

Starting year: **1951** Ending year: **1952**

Countries: **United States**

Download format: **Wide** Download extension: **csv**

**Download data** **Download metadata**

GID_0	NAME_1	1951	1952	1953	1954	1955	1956	1957	1958
3,411	USA	16.1789	17.383	16.7206	17.062	16.9535	17.3429	17.6351	17.7131
3,412	USA	-0.114	-0.5022	-0.4304	-0.8577	1.324	-0.309	-0.5969	-0.0688
3,413	USA	18.7609	18.276	18.1166	18.7688	18.0182	18.1511	18.3876	18.257
3,414	USA	16.0148	15.8033	15.1913	15.6839	15.3404	15.0896	16.468	16.6769
3,415	USA	15.5685	14.962	15.1336	15.654	15.126	15.3324	14.9854	14.7524
3,416	USA	7.7194	7.5126	6.8035	7.8761	7.0744	7.4647	8.1334	7.5193
3,417	USA	8.3206	8.5654	8.5228	7.0147	8.062	8.8988	7.6926	9.1037
3,418	USA	11.4115	11.8663	11.761	10.2762	11.4062	12.2802	10.9666	12.1682
3,419	USA	District of Col.	12.2741	12.889	12.5139	11.3465	12.3568	13.3217	12.0083
3,420	USA	Florida	20.9542	22.0664	21.399	21.665	22.0757	21.8505	22.3604

Figure 8. Dashboard interface, exemplified

13. Diffenbaugh, N. S. & Burke, M. Global warming has increased global economic inequality. *Proc. Natl. Acad. Sci.* **116**, 9808–9813 (2019).
14. Alessandri, P. & Mumtaz, H. The macroeconomic cost of climate volatility. *arXiv preprint arXiv:2108.01617* (2021).
15. Spinoni, J. *et al.* Global population-weighted degree-day projections for a combination of climate and socio-economic scenarios. *Int. J. Climatol.* **41**, 5447–5464 (2021).
16. Wei, R., Li, Y., Yin, J. & Ma, X. Comparison of weighted/unweighted and interpolated grid data at regional and global scales. *Atmosphere* **13**, 2071 (2022).
17. Ponticelli, J., Xu, Q. & Zeume, S. Temperature and local industry concentration. Tech. Rep., National Bureau of Economic Research (2023).
18. Donadelli, M., Jüppner, M. & Vergalli, S. Temperature variability and the macroeconomy: A world tour. *Environ. Resour. Econ.* **83**, 221–259 (2022).
19. Cipollini, A., Parla, F. *et al.* Temperature and growth: A panel mixed frequency VAR analysis using NUTS2 data. *RECENT WORKING PAPER SERIES* (2023).
20. Donadelli, M., Grüning, P., Jüppner, M. & Kizys, R. Global temperature, R&D expenditure, and growth. *Energy Econ.* **104**, 105608 (2021).
21. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. data* **7**, 109 (2020).
22. Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I., Angulo, M. & El Kenawy, A. A new global 0.5 gridded dataset (1901–2006) of a multiscalar drought index: Comparison with current drought index datasets based on the palmer drought severity index. *J. Hydrometeorol.* **11**, 1033–1043 (2010).
23. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. Royal Meteorol. Soc.* **146**, 1999–2049 (2020).
24. Willmott, C. J. & Matsuura, K. Terrestrial air temperature and precipitation: Monthly and annual time series (1950–1996). *Cent. for Clim. Res. Dep. Geogr. Univ. Del.* (2000).
25. Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Clim.* **23**, 1696–1718 (2010).
26. Peterson, T. C. & Vose, R. S. An overview of the global historical climatology network temperature database. *Bull. Am. Meteorol. Soc.* **78**, 2837–2850 (1997).
27. Lawrimore, J. H. *et al.* An overview of the global historical climatology network monthly mean temperature data set, version 3. *J. Geophys. Res. Atmospheres* **116** (2011).
28. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. atmospheric oceanic technology* **29**, 897–910 (2012).
29. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, 517–524 (1968).
30. Doxsey-Whitfield, E. *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.* **1**, 226–234 (2015).
31. Li, X., Zhou, Y., Zhao, M. & Zhao, X. A harmonized global nighttime light dataset 1992–2018. *Sci. data* **7**, 168 (2020).
32. GADM. <https://gadm.org/>. Accessed: 2022-07-16.

## Acknowledgements

We are grateful to Francesca Chiaromonte, Daniele Colombo and Damiano Di Francesco for useful feedback and suggestions. F.L. acknowledges support from the Italian Ministry of Research, PRIN 2022 project “ECLIPTIC”.

## Author contributions statement

All authors conceived ideas and analysis approaches. M.G. and L.T. retrieved and processed data, implemented pipelines, and performed statistical analyses. All authors wrote the manuscript. G.F. and F.L. supervised the research.

## Competing interests

The authors declare that they have no competing interests.