

# SSP-consistent high-resolution maps of disposable income in Europe

## Abstract

The use of high-resolution datasets on socio-economic characteristics of the population is key to assessing climate risk. Income, age and gender are widely recognised as determinants of population vulnerability to extreme climatic events such as heat waves, droughts, heavy rains and storms. Many studies that focus on mapping future climate vulnerability use coarse-resolution income proxies (such as GDP) when assessing climate risk. To address these research gaps, we created datasets for per capita disposable income at 1km resolution between 2015 and 2100 that are consistent with the Shared Socioeconomic Pathways (SSP). First, we collected existing income data (gross income, gross earnings, primary income, etc.) at different geographical levels from national statistical institutes, which we reconciled using statistical operations. We then used a machine learning algorithm trained on publicly available socio-ecological predictors to downscale per capita income. The income projections for the 5 SSP relied mainly on available SSP-based projections of spatial urban areas and population datasets, as well as country-level income growth by decile. These projections provide alternative dynamics for the evolution of spatial inequalities in Europe. This study provides new datasets and opportunities to more accurately assess future climate risks and identify socio-economic vulnerability hotspots.

## Main

Since 1980, within-country inequalities have started to rise following a decrease in the first part of the 20<sup>th</sup> century<sup>1</sup>. These analyses are made possible thanks to the availability of detailed income and wealth datasets. At the global level, the World Inequality Database (<https://wid.world/>) provides time-series of average income for each centile along the income distribution. While very informative, these income distributions do not tell us much about the dynamics of spatial inequalities within countries, as they rather focus on vertical inequalities. Spatial inequalities are much more difficult to capture because of the limitation of fine-resolution datasets. At the global level, Gennaioli et al<sup>2</sup> have constructed a database of subnational regional incomes; and used in Europe, income data published annually by Eurostat (the statistical office of the European Union) at the NUTS2 level (basic regions for the application of regional policies). Although finer than national data, the NUTS2 level is still quite coarse (the average area of NUTS2 of units is approx. 18,000 km<sup>2</sup>), and does not allow for an assessment of local spatial inequalities.

Human-induced greenhouse gas emissions are responsible for the rise in global temperatures and changes in the frequency and intensity of extreme climate events<sup>3</sup>. These events affect human populations unevenly and can be exacerbated by some socio-economic characteristics - gender, ethnicity, race, age and income level – of the population. Climate vulnerability mapping is one way to study the distributional impacts of climate change, that has been increasingly used in recent years. In Europe, such studies are less common and most of them suffer from some methodological limitations<sup>4,5</sup>. Namely, they tend to represent socioeconomic variables at a coarse resolution and use maps of current socioeconomic variables with future maps of climate extreme events. To promote consistency in the climate modelling community, the Shared Socioeconomic Pathways (SSP) were developed and consist of a set of 5 alternative narratives describing plausible developments of society<sup>6</sup>. More recently, various high-resolution datasets for future gross domestic product (GDP) consistent with the SSPs have been developed on the basis of existing regional GDP data<sup>7,8</sup>. While informative on future economic activity, GDP do not properly inform on inequalities as it does not account for redistribution enabled by taxes and transfers. In addition, regional GDP in Europe may be skewed in favour of regions with net

commuter inflows and against regions with net commuter outflows<sup>9</sup>. Hence, GDP is not the most appropriate proxy to represent spatial income inequalities.

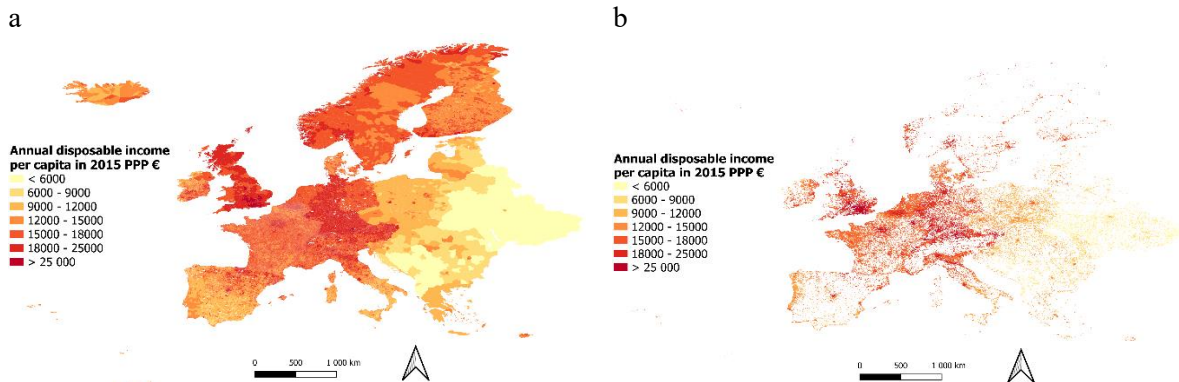
To fill these research gaps, we propose a new dataset of per capita disposable income at the European level at 1km resolution consistent with the SSPs. According to the European System Accounts<sup>9</sup>, disposable income is defined as the sum of net operating surplus/mixed income, compensation of employees, balance of property income, social benefits received including old-age pensions, from which taxes on income/wealth and compulsory social contributions are deducted. This dataset was compiled in two main steps. The first step consists in downscaling income data published by national statistical institutes (NSIs) in Europe. We will then use a machine learning framework to downscale income, following a number of studies that have downscaled socio-economic variables this framework<sup>10–13</sup>. Finally, based on spatially explicit future settlement growth and population dynamics along future country level income growth per decile, we generate new high resolution income maps.

## Results

### The downscaling of income for the year 2015

Countries in Western Europe have a higher within-country income heterogeneity compared to the rest of Europe (Fig.1a). This is explained by the size of the AUs from which income was collected. The downscaling procedure achieves similar trends in the income distribution across Europe (Fig.1b). By using several methods to validate the downscaling, we have consistently obtained high  $R^2$  values and low values for the relative Root Mean Square Error<sup>14</sup> (Supplementary Fig 1). The income distribution in European countries shows a strong longitudinal and latitudinal gradient, with western and northern countries having higher income than southern and eastern countries. Income is also concentrated around the capital cities.

**Fig.1 : Per capita disposable income in 2015**



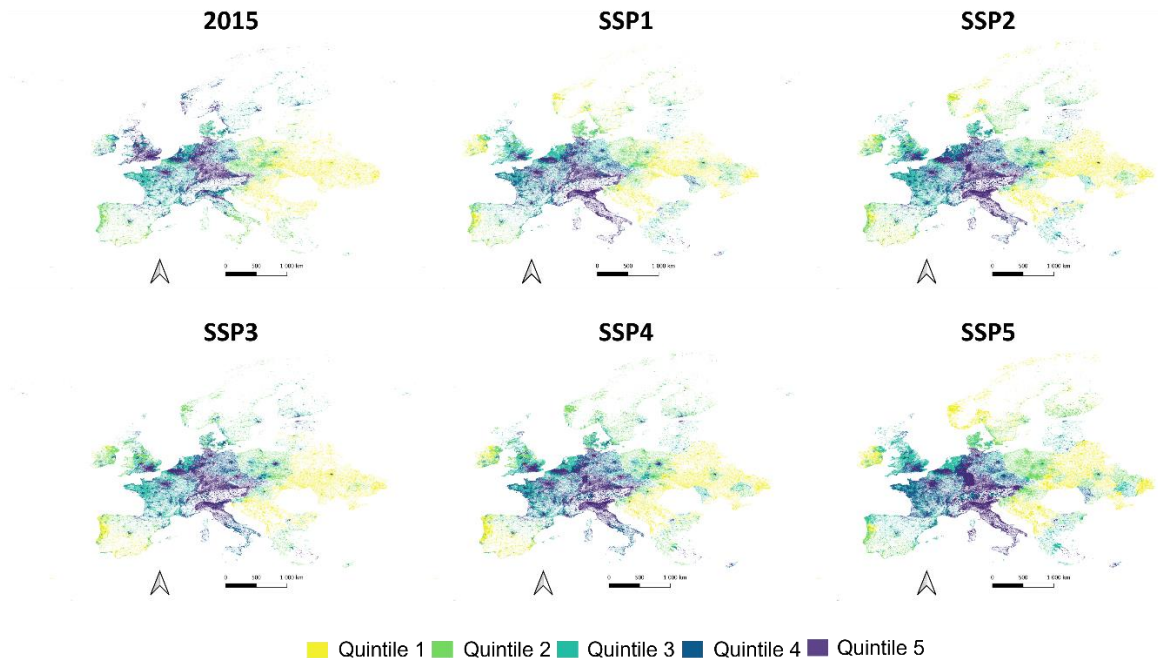
Comparison of per capita disposable income across European countries at the administrative unit level (a) and at the 1km pixel level (b) after the downscaling.

### Static comparison of the spatial distribution of income across different development pathways in 2100

Compared to 2015, the future spatial distribution of income in the different SSPs shows contrasted evolutions. If we look at the relative position of countries in terms of income between 2015 and 2100, 3 main dynamics seem to emerge (Fig.2). The relative position of Northern European countries in the global income distribution decreases in 2100 in most SSPs. In Norway, while more than 90% of the pixels were among the main 2 groups of highest-earners in Europe in 2015, on average more than 80% of the territory falls into lowest 2 groups of income in 2100 (Supplementary Fig 2). On the opposite side, countries like the Czech Republic, Bulgaria and Poland will experience a relative improvement of their position in the European income distribution. Countries in the Balkan Peninsula and some countries

in Western and Central Europe (e.g. France and Germany) will not experience major changes according to our simulations. These results are primarily due to the convergence process in income growth between countries as projected by Dellink et al<sup>15</sup>.

**Fig.2 : Spatial distribution of income in 2100 for different development pathways**



In the different SSPs, income is represented using categories based on quintiles. In each scenario, pixels are equally split in 5 groups of income. Population is not considered to make these splits. Quintile 1 and 5 represent respectively the poorest and richest areas.

Romania, Andorra, Kosovo and Liechtenstein are not represented because one input dataset was missing for these countries

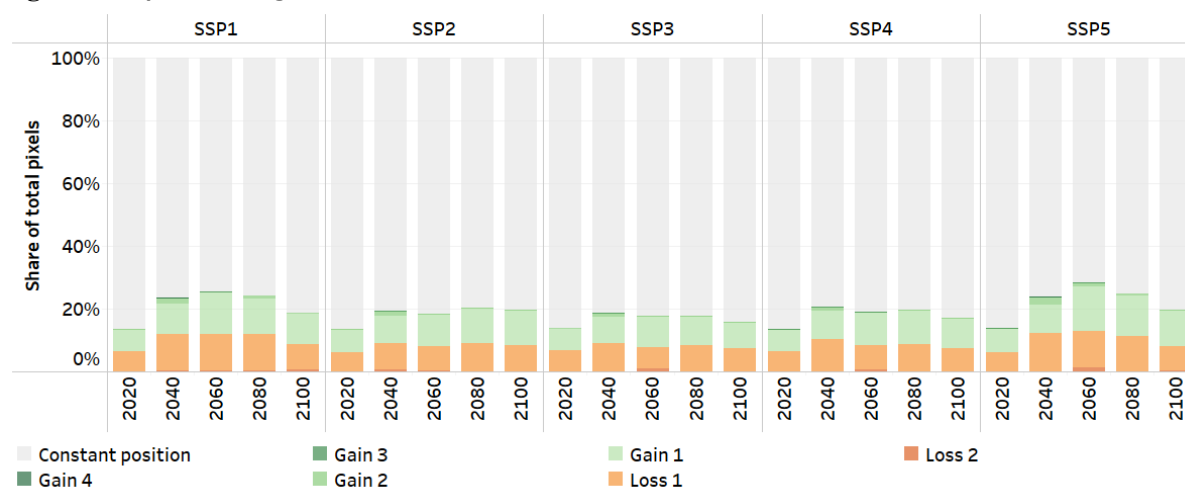
While this process of convergence is inherent (at various degrees) to each future pathway, the differences between SSPs can be explained by alternative evolutions of settlement growth<sup>16</sup>. SSP5 is particularly marked by a high level of settlement expansion compared to the SSP1, where this is much lower. Given the income projection methodology, the number and location of new settlements have implications for the future spatial distribution of income.

These results - by comparing results over only 2 time periods - do not tell us about the evolution of inequalities over time. In the next section, we will examine the temporal dynamics of the evolution of income in the different SSPs.

### **Different dynamics in the evolution of the distribution of income across different development pathways**

With this dataset, we are able to track the temporal evolution of income in different countries. At the European level, about 20% of total pixels – on average - see their position in the European distribution of income change in each 20-year period (Fig.3). These changes – measured by the difference in the quintile position of each pixel in a time span of 20 years - are dominated across all SSPs by pixels gaining or losing one position in the global distribution of income. In SSP 1 and 5, where high levels of growth are expected, the number of pixels affected by a change in income position rises to 25% with the highest changes occurring in the decade 2040-2060.

**Fig.3 : Temporal changes in the distribution of income**



We represent changes in the quintile position of each pixel in different years. The change is calculated using the difference of the quintile group of each pixel between two consecutive periods. Green bars represent the share of pixels that have gained between 1 and 4 positions while oranges ones have lost 1 or 2 positions.

Romania, Andorra, Kosovo and Liechtenstein are not represented because one input dataset was missing for these countries

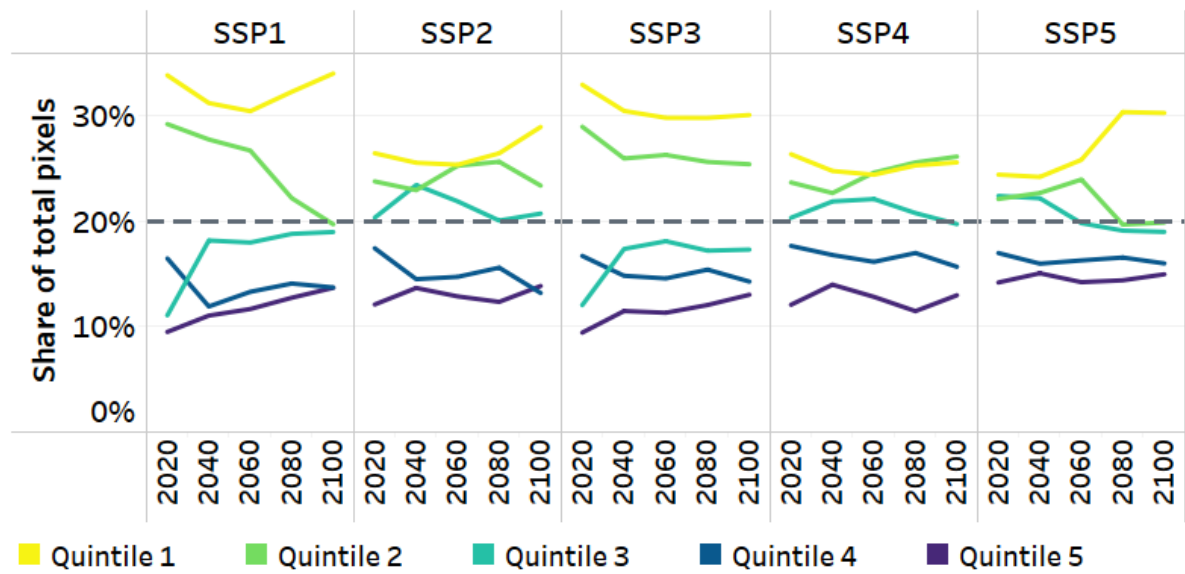
Part of these alternative developments is related to the different levels of expected settlement growth within each SSP, but also to their spatial localisation. As their income is largely determined by the neighbouring settlements, we can understand the characteristics of the areas where these new settlements emerge.

## Understanding the social characteristics of areas next to new settlements over time

When we plot the relative position of new settlements in the European income distribution, we find an overrepresentation of settlements belonging to the first 2 classes of the quintiles of European income and an underrepresentation of settlements belonging to the 2 highest income classes (Fig. 4). This observation remains valid over time (partly due to the path dependency of the projection methodology) and across different SSPs. SSP2 and SSP4, with similar trends in settlement growth, show similar paths in the distribution of new settlements across the 5 European quintiles by 2080. In contrast, SSP1 and SSP5, although differently distributed in 2015, tend to converge to the same state in 2080. The predominance of new settlements in low-income areas is mainly related to the high urbanisation trends projected in the different SSPs in relatively poor countries in Eastern Europe. These results can be extended to understand the situation within individual countries.

If we move from a European perspective to a country-level analysis, we can examine the characteristics - in terms of income - of the surrounding areas of new settlements. In both SSP1 and SSP3, the proportion of new settlements landing in the bottom half of the income distribution within each country is above 55% in most countries in 2100 (Fig.5). This indicates that new settlements are mainly located near areas that belong to the lower half of the income distribution. In SSP5, which is characterised by a high level of settlement expansion, we observe an opposite effect, where a higher proportion of new settlements emerge near relatively richer areas.

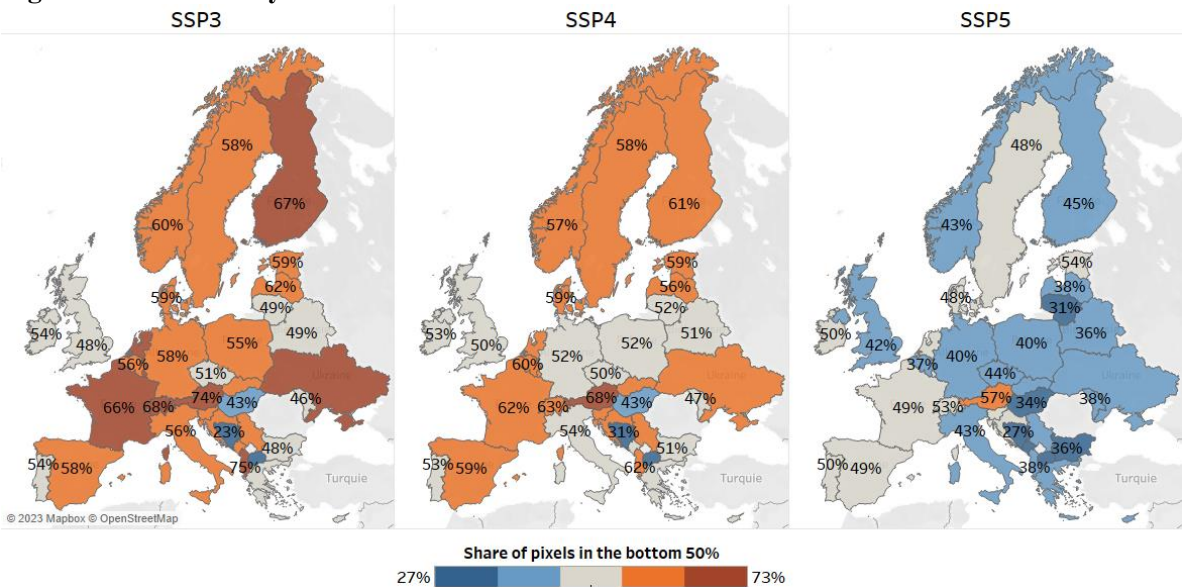
**Fig.4 : Evolution of the relative position of new settlements in the European distribution of income across time**



We only represent here new settlements that have appeared after 2015. In each year, new pixels are added to the pool of new pixels from the previous period. The dotted line represents the value corresponding to the uniform distribution of settlements across the 5 classes of quintile.

Romania, Andorra, Kosovo and Liechtenstein are not represented because one input dataset was missing for these countries

**Fig.5 : Within-country distribution of new settlements**



We only represent here new settlements that have appeared after 2015. We compute for each country the share of new settlements in the bottom 50% category of income in 2100. Results for SSP1 and SSP2 are not represented as they much look like SSP3 and SSP4 respectively.

Romania, Andorra, Kosovo and Liechtenstein are not represented because one input dataset was missing for these countries

## Methods

### Income database

The income database was developed after collecting income data from European NSI at different administrative units (AU). A harmonization procedure was then carried out to convert the income from the NSIs – corresponding to different stages of earnings and income (pre-tax income, net earnings...) – into disposable income. The final dataset covers 42 European countries – distributed over more than 120,000 AU – for the year 2015. When available, post-tax income Gini coefficient was also collected at the same AU than income. If missing, the Gini coefficient was replaced by the regional or national value.

### Collecting socioeconomic and environmental predictors at 1km resolution

To downscale income using a machine-learning algorithm, we compiled a list of 22 indicators that serve as predictors in our training (Table 1). Additional specifications on these indicators can be found in Supplementary Table 1. Most of these predictors have been used by other scholars for the downscaling of other socio-environmental variables. For example, average temperature, average annual precipitation, distance to major roads, distance to different land use types (water, agriculture, barren..) and distance to airports were used to downscale population data<sup>10,17</sup>. To construct high-resolution maps of the secondary and tertiary sectors of GDP in China, scholars have integrated points-of-interest into their predictors<sup>13</sup>. In this study, we collected the geographical locations of several categories of points-of-interest (public buildings, public transportation, factories...) from OpenStreetMap (<https://www.openstreetmap.org/>) using the osmdata (<https://cran.r-project.org/web/packages/osmdata/vignettes/osmdata.html>) package in R software. Nigh-light data has extensively been used in numerous studies to downscale GDP<sup>13,18</sup>. Digital elevation data and the normalised difference vegetation index were used as predictors to determine the location of jobs within cities<sup>11</sup>.

Some data processing operations were conducted to create a harmonized set of predictors. The resolution of these different datasets (between 100 m and 1 km) was transformed to 1km using the aggregate function from the R raster package (<https://cran.r-project.org/web/packages/raster/raster.pdf>). To ensure that these datasets are geographically aligned, we used the *resample* function from the same package to align the location of the pixels in each of these datasets. We also applied a mask of built-up settlements to assign income only to the potentially inhabited pixels<sup>16</sup>.

Table 1 : List of socio-environmental indicators

| Type           | Name                                  | Resolution | Source          |
|----------------|---------------------------------------|------------|-----------------|
| Climate        | Average temperature                   | 1km        | 19              |
|                | Average annual precipitation          | 1km        | 19              |
| Infrastructure | Distance to major roads               | 100m       | 20              |
|                | Share of urban land                   | 1km        | 16              |
|                | Distance to airports                  |            | Ourairports.com |
|                | Distance to waterways                 | 100m       | 20              |
| Land use       | Distance to tree area                 | 100m       | 20              |
|                | Distance to agriculture area          | 100m       | 20              |
|                | Distance to bare area                 | 100m       | 20              |
|                | Distance to aquatic vegetation        | 100m       | 20              |
|                | Distance to vegetation                | 100m       | 20              |
|                | Distance to factories                 |            | OpenStreetMap   |
|                | Distance to tertiary sector buildings |            | OpenStreetMap   |

|                    |                                               |      |               |
|--------------------|-----------------------------------------------|------|---------------|
| Points of interest | Distance to rail public transport             |      | OpenStreetMap |
|                    | Distance to cultural centres                  |      | OpenStreetMap |
|                    | Distance to public buildings                  |      | OpenStreetMap |
| Remote sensing     | Night light data                              | 500m | 21            |
|                    | Normalized difference vegetation index (NDVI) | 1km  | Copernicus    |
| Topographic        | Elevation                                     | 100m | 20            |
|                    | Slope                                         | 100m | 20            |
|                    | Longitude                                     |      |               |
|                    | Latitude                                      |      |               |

## Downscaling framework

Downscaling is not a typical problem of supervised machine learning, since we do not observe data at our target resolution (1km), but only the aggregation of some observations within an AU. Scientists have typically solved this problem by training an AU-level model and applying it at the pixel level<sup>10</sup>. However, this method is not immune to the "ecological fallacy", which states that a correlation between 2 variables does not necessarily exist at different levels of aggregation<sup>22,23</sup>. Compared to other downscaling exercises, we also observed the Gini coefficient, an indicator of inequality. We used this additional information on income to generate multiple samples of a plausible distribution of income within each AU. Each generated sample was used to train a separate machine learning algorithm with the list of socioenvironmental indicators at the 1km resolution. The methodology is summarized in Supplementary Fig 2.

### 1) *Sampling per capita income using a lognormal distribution*

To generate samples of income distributions within each AU, we assumed that individual income was distributed according to a logNormal distribution. This assumption has been shown to be valid for the bottom 97% to 98% of the population in countries like Germany, the United Kingdom or the United States<sup>24</sup>. While Pareto-logNormal or generalized Pareto distributions have demonstrated a better ability to model high incomes<sup>25,26</sup>, we have used the logNormal distribution thanks to its simple characterization by the Gini coefficient. Indeed, we were able to derive the full distribution of individual income from the average income in each AU and the associated Gini coefficient<sup>27</sup>. From the individual income distribution, we generated multiple samples - of the size of the AU population obtained through WorldPop<sup>28</sup> - of income within each AU. These generated incomes were then randomly attributed to a pixel and averaged to obtain pixel-level per capita income.

### 2) *Training and predicting income*

Each sample of pixel-level income was then used to train a Random Forest<sup>29</sup> algorithm at the European level with the list of 22 predictors. This algorithm was chosen because of its extensive use in the literature related to the downscaling of socioeconomic variables<sup>10,11,13</sup>, its relative interpretability<sup>30</sup> and the low number of parameters that are needed to be tuned when compared to other machine learning algorithms. Once trained, each Random Forest model was used to predict pixel-level income on the basis of the same predictors. Finally, the predictions obtained from the different samples were averaged to get the final prediction. Although the randomness associated with the generation of pixel-level income did not properly distribute income within each AU, we believe that the repeated sampling procedure will yield better results.

### 3) *Model validation*

As in every downscaling exercise, the target variable was not observable at the pixel-level resolution. This limitation didn't allow us to follow a *normal* model validation procedure constituted of a training and a test subsets of the data. We have come-up with two alternative ways to validate the procedure:

- i. Averaging income at the administrative units: We averaged the income obtained in each AU after the downscaling and compared it with the per capita income collected from the NSIs.
- ii. 1km gridded income for France: We used an independent gridded dataset of winsorized (it aims at maintaining statistical confidentiality by converting all values above or below a threshold to a given value) income from INSEE, the French NSI. From the income and population at the pixel level, we were able to calculate the average and standard deviation of the per capita income at the AU level. Then the methodology described earlier was applied to this country.

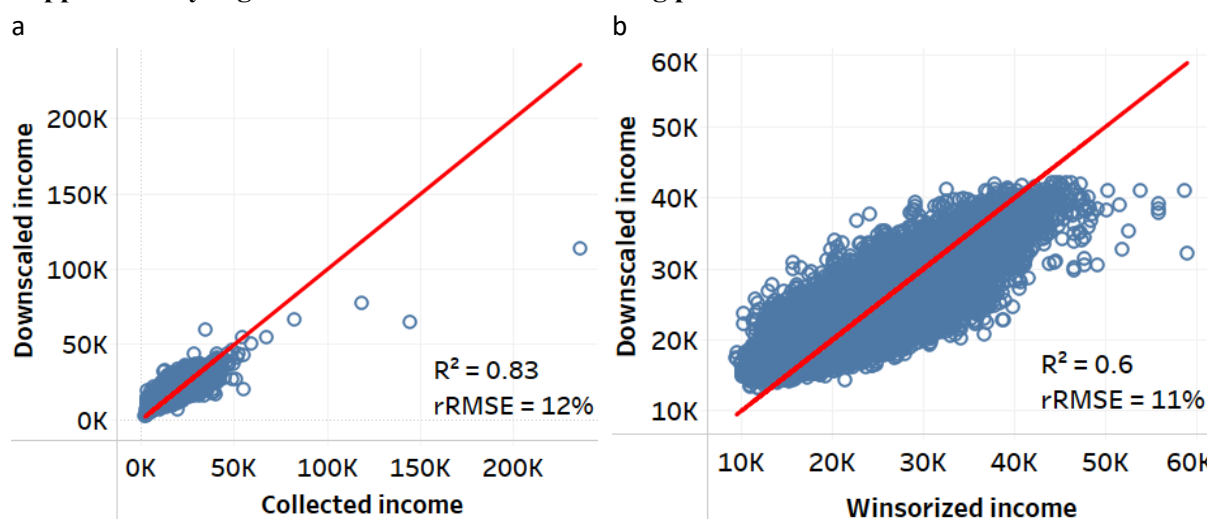
### **SSP consistent income projections**

This step was achieved through the use of multiple – spatially explicit and country-level - global SSP-consistent datasets. Among the spatially explicit datasets, we have used 1km gridded projections of population<sup>31</sup> and future settlement growth<sup>16</sup>. At the country level, we have used projections of per capita GDP<sup>15</sup> and the share of national income captured by each decile<sup>32</sup>. From these 2 country-level datasets and, by assuming that per capita income grows at the same rate than per capita GDP, we derived income growth in each decile. The procedure for projecting income was constructed iteratively and is summarised below for the projection of the income distribution at time t:

1. We began by identifying changes in settlements on the map at time t. Pixels that were no longer considered settlements at time t were left aside, while new settlements were added to the list of pixels that are considered as settlements.
2. Pixels appearing between 2 timestamps were assigned the average income value of their neighbouring pixels. Based on this, we were able to assign each pixel to a decile level using the population dataset.
3. Finally, a constant income growth per decile was applied to pixels belonging to the same income decile.

## Supplementary Information

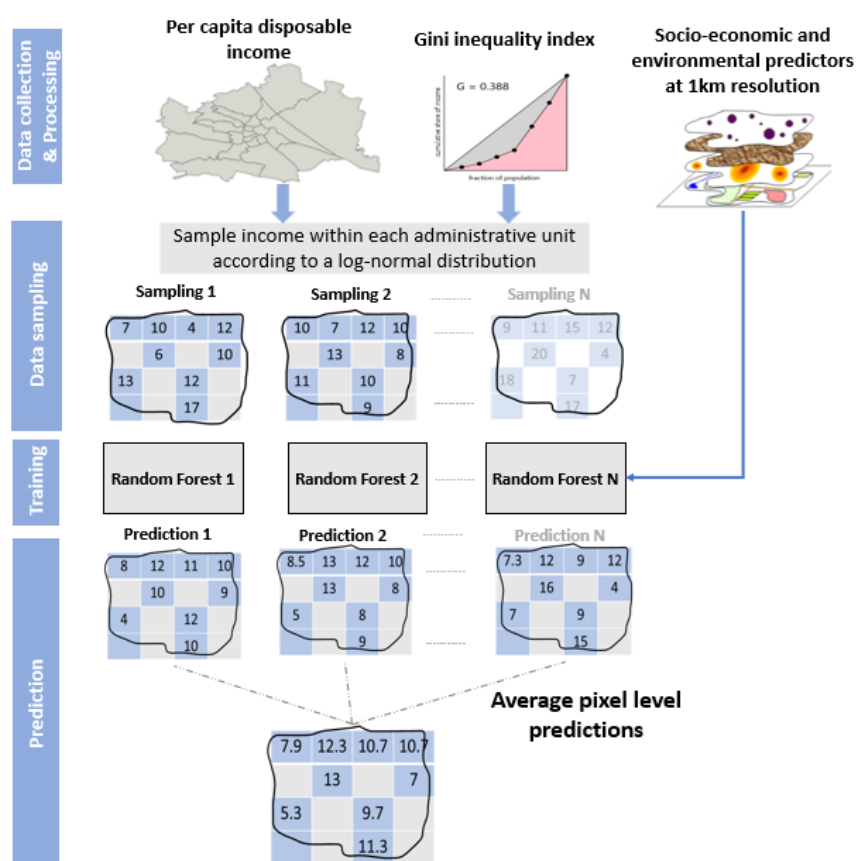
**Supplementary Fig.1 : Validation of the downscaling procedure**



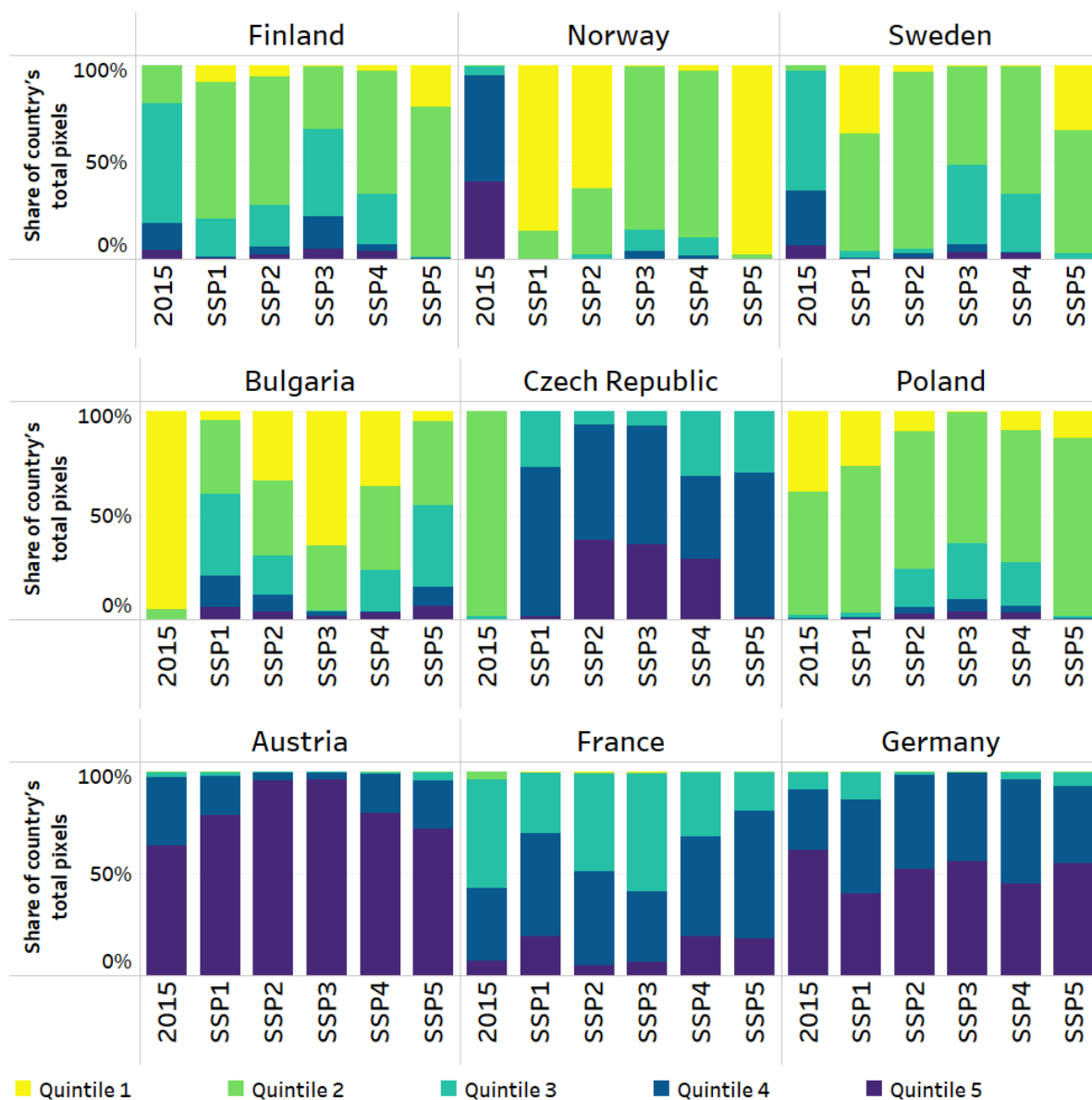
a-b, Income after downscaling plotted against collected income at the administrative unit level (a) at pixel level for France (b). The red line represents the identity line

Note : In (a), the validation is performed at the European level after averaging pixel-level income obtained by the downscaling at the administrative level. In both cases, at least 60% of the variation of the target variable is explained by our modelling. On average, our predictions overestimate or underestimate the target variable by about 10%.

**Supplementary Fig.2 : Framework for downscaling income from the administrative level to the 1km target resolution**



**Supplementary Fig 3 : Comparison between the share of country pixels belonging to different European quintiles between 2015 and 2100 in different SSPs.**



| Type               | Name                                          | Resolution | Source          | Specification                                                                                                                                                           |
|--------------------|-----------------------------------------------|------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Climate            | Average temperature                           | 1km        | 19              |                                                                                                                                                                         |
|                    | Average annual precipitation                  | 1km        | 19              |                                                                                                                                                                         |
| Infrastructure     | Distance to major roads                       | 100m       | 20              |                                                                                                                                                                         |
|                    | Share of urban land                           | 1km        | 16              |                                                                                                                                                                         |
|                    | Distance to airports                          |            | Ourairports.com |                                                                                                                                                                         |
| Land use           | Distance to waterways                         | 100m       | 20              |                                                                                                                                                                         |
|                    | Distance to tree area                         | 100m       | 20              | ESA-CCI-LC 040                                                                                                                                                          |
|                    | Distance to agriculture area                  | 100m       | 20              | ESA-CCI-LC 011                                                                                                                                                          |
|                    | Distance to bare area                         | 100m       | 20              | ESA-CCI-LC 200                                                                                                                                                          |
|                    | Distance to aquatic vegetation                | 100m       | 20              | ESA-CCI-LC 160                                                                                                                                                          |
|                    | Distance to vegetation                        | 100m       | 20              | ESA-CCI-LC 130 140 150                                                                                                                                                  |
| Points of interest | Distance to factories                         |            | OpenStreetMap   | OSM {"landuse";"industrial"}                                                                                                                                            |
|                    | Distance to tertiary sector buildings         |            | OpenStreetMap   | OSM{"landuse";["commercial","retail"]}<br>OSM{"amenity";["restaurant","bar","cafe","biergarten","fast_food","pub"]}<br>OSM{"amenity";["atm","bank","bureau_de_change"]} |
|                    | Distance to rail public transport             |            | OpenStreetMap   | OSM {"railway";["subway_entrance","tram_stop"]}                                                                                                                         |
|                    | Distance to cultural centres                  |            | OpenStreetMap   | OSM{"amenity";["arts_centre","cinema","community_centre","conference_centre","theatre"]}                                                                                |
|                    | Distance to public buildings                  |            | OpenStreetMap   | OSM{"amenity";["college","kindergarten","library","school","university"]}<br>OSM{"landuse";"education"}<br>OSM{"building";"government"}                                 |
| Remote sensing     | Night light data                              | 500m       | 21              |                                                                                                                                                                         |
|                    | Normalized difference vegetation index (NDVI) | 1km        | Copernicus      |                                                                                                                                                                         |
| Topographic        | Elevation                                     | 100m       | 20              |                                                                                                                                                                         |
|                    | Slope                                         | 100m       | 20              |                                                                                                                                                                         |
|                    | Longitude                                     |            |                 |                                                                                                                                                                         |
|                    | Latitude                                      |            |                 |                                                                                                                                                                         |

SI Table 1 : List of predictors used in the downscaling

## References

1. Chancel, L. & Piketty, T. Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality. *J. Eur. Econ. Assoc.* **19**, 3025–3062 (2021).
2. Gennaioli, N., La Porta, R., Lopez-de-Silanes, F. & Shleifer, A. Human Capital and Regional Development \*. *Q. J. Econ.* **128**, 105–164 (2013).
3. Portner, H.-O., Roberts, D., Trisos, C. & Simpson, N. Summary for Policymakers: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. (2022) doi:10.1017/9781009325844.001.
4. de Sherbinin, A. *et al.* Climate vulnerability mapping: A systematic review and future prospects. *WIREs Clim. Change* **10**, e600 (2019).
5. Rohat, G. *et al.* Influence of changes in socioeconomic and climatic conditions on future heat-related health challenges in Europe. *Glob. Planet. Change* **172**, 45–59 (2019).
6. O'Neill, B. C. *et al.* A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400 (2014).
7. Murakami, D., Yoshida, T. & Yamagata, Y. Gridded GDP Projections Compatible With the Five SSPs (Shared Socioeconomic Pathways). *Front. Built Environ.* **7**, (2021).
8. Wang, T. & Sun, F. Global gridded GDP data set consistent with the shared socioeconomic pathways. *Sci. Data* **9**, 221 (2022).
9. Eurostat. *European system of accounts :ESA 2010*. (Publications Office, 2013).
10. Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE* **10**, e0107042 (2015).
11. Barzin, S., Avner, P., Rentschler, J. & O'Clery, N. *Where Are All the Jobs ? : A Machine Learning Approach for High Resolution Urban Employment Prediction in Developing Countries*. <https://openknowledge.worldbank.org/handle/10986/37195> (2022) doi:10.1596/1813-9450-9979.
12. Thomson, D. R., Kools, L. & Jochem, W. C. Linking Synthetic Populations to Household Geolocations: A Demonstration in Namibia. *Data* **3**, 30 (2018).

13. Chen, Q. *et al.* Mapping China's regional economic activity by integrating points-of-interest and remote sensing data with random forest. *Environ. Plan. B Urban Anal. City Sci.* **48**, 1876–1894 (2021).
14. McGough, S. F., Brownstein, J. S., Hawkins, J. B. & Santillana, M. Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data. *PLoS Negl. Trop. Dis.* **11**, e0005295 (2017).
15. Dellink, R., Chateau, J., Lanzi, E. & Magné, B. Long-term economic growth projections in the Shared Socioeconomic Pathways. *Glob. Environ. Change* **42**, 200–214 (2017).
16. Gao, J. & O'Neill, B. C. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. *Nat. Commun.* **11**, 2302 (2020).
17. Ye, T. *et al.* Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *Sci. Total Environ.* **658**, 936–946 (2019).
18. Chen, Y., Wu, G., Ge, Y. & Xu, Z. Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning With Multiple Geospatial Big Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **15**, 1791–1802 (2022).
19. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
20. WorldPop. Global 100m Covariates. (2018) doi:10.5258/SOTON/WP00644.
21. Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C. & Taneja, J. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. *Remote Sens.* **13**, 922 (2021).
22. Hammond, J. L. Two Sources of Error in Ecological Correlations. *Am. Sociol. Rev.* **38**, 764–777 (1973).
23. Robinson, W. S. Ecological Correlations and the Behavior of Individuals. *Am. Sociol. Rev.* **15**, 351–357 (1950).
24. Clementi, F. & Gallegati, M. Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States. *Microeconomics* (2005).

25. Blanchet, T., Fournier, J. & Piketty, T. Generalized Pareto Curves: Theory and Applications. *Rev. Income Wealth* **68**, 263–288 (2022).
26. Colombi, R. A New Model Of Income Distribution: The Pareto-Lognormal Distribution. in *Income and Wealth Distribution, Inequality and Poverty* (eds. Dagum, C. & Zenga, M.) 18–32 (Springer, 1990). doi:10.1007/978-3-642-84250-4\_2.
27. Crow. *Lognormal distribution: Theory and Applications*. (Routledge, 2017).  
doi:10.1201/9780203748664.
28. WorldPop. Global 1km Population Individual countries. (2020)  
doi:10.5258/SOTON/WP00670.
29. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
30. Barredo Arrieta, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
31. Li, M. *et al.* Spatiotemporal dynamics of global population and heat exposure (2020–2100): based on improved SSP-consistent population projections. *Environ. Res. Lett.* **17**, 094007 (2022).
32. Narayan, K. B., O'Neill, B. C., Waldhoff, S. T. & Tebaldi, C. Non-parametric projections of national income distribution consistent with the shared socioeconomic pathways. *Environ. Res. Lett.* (2023) doi:10.1088/1748-9326/acbdb0.