Science of the Total Environment 941 (2024) 173623

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Spatially explicit downscaling and projection of population in mainland China

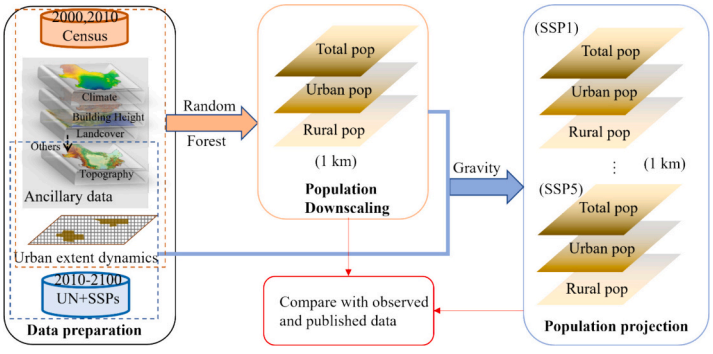
Wenru Xu [a], Yuyu Zhou [b,*], Hannes Taubenböck [c], Eleanor C. Stokes [d], Zhengyuan Zhu [e], Feilin Lai [f], Xuecao Li [g], Xia Zhao [h]

[a] Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China
[b] Department of Geography, The University of Hong Kong, Hong Kong
[c] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, 82234 Weßling, Germany
[d] NASA Headquarters, Washington, USA
[e] Department of Statistics, Iowa State University50011, Ames, IA, USA
[f] Department of Geography and Planning, St. Cloud State University, MN 56301, USA
[g] College of Land Science and Technology, China Agricultural University, Beijing 100083, China
[h] Institute of Land and Urban-Rural Development, Zhejiang University of Finance & Economics, Hangzhou 310018, China

HIGHLIGHTS

- A framework was developed for spatially explicit downscaling and projecting population distributions.
- The accuracy of the downscaled population was improved compared with other datasets.
- The population projections align with the spatial patterns of urbanization.
- The work provides a foundation for studies on long-term socio-environmental interactions.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Kuishuang Feng

Keywords:
Gridded population
Population downscaling
Population projections
Urban sprawl
Shared Socioeconomic Pathways

ABSTRACT

Spatially explicit population data is critical to investigating human-nature interactions, identifying at-risk populations, and informing sustainable management and policy decisions. Most long-term global population data have three main limitations: 1) they were estimated with simple scaling or trend extrapolation methods which are not able to capture detailed population variation spatially and temporally; 2) the rate of urbanization and the spatial patterns of settlement changes were not fully considered; and 3) the spatial resolution is generally coarse. To address these limitations, we proposed a framework for large-scale spatially explicit downscaling of populations from census data and projecting future population distributions under different Shared Socio-economic Pathways (SSP) scenarios with the consideration of distinctive changes in urban extent. We downscaled urban and rural population separately and considered urban spatial sprawl in downscaling and projection. Treating urban and rural populations as distinct but interconnected entities, we constructed a random forest model to

* Corresponding author.
E-mail address: yuyuzhou@hku.hk (Y. Zhou).

https://doi.org/10.1016/j.scitotenv.2024.173623
Received 14 March 2024; Received in revised form 9 May 2024; Accepted 27 May 2024
Available online 28 May 2024

downscale historical populations and designed a gravity-based population potential model to project future population changes at the grid level. This work built a new capacity for understanding spatially explicit demographic change with a combination of temporal, spatial, and SSP scenario dimensions, paving the way for cross-disciplinary studies on long-term socio-environmental interactions.

1. Introduction

The quantitiy, distribution and demographics of the human population are critical components in understanding human-environment connections, identifying vulnerable populations, and guiding sustainable management and policy-making (e.g., Jones et al., 2008; Marzi et al., 2021; Wesolowski et al., 2014). Temporal changes in population size and distribution, especially in regions undergoing rapid expansion and urban development, have notable ecological and socio-economic consequences. For example, the distribution and growth of human populations have played a significant role in generating substantial greenhouse gas emissions, exacerbating air pollution, and increasing resource consumption (Akhmat et al., 2014; Rosa and Dietz, 2012; Müller et al., 2022; Hu et al., 2021). Rapid population increase may threaten protected areas, sensitive habitats, biodiversity, and poses challenges to urban planning and management, as well as the governance of environmental issues within cities (Li et al., 2019; Tatem et al., 2012; Wesolowski et al., 2012; Xu et al., 2020). Therefore, long-term spatially explicit gridded population modeling is in demand and of growing importance in global change and vulnerability assessment, sustainable development, or resource management.

Several methods have been developed and applied to generate long-term gridded population maps by spatially downscaling population projections to the grid cell level using the national-level Shared Socio-economic Pathways (SSP) projections as boundary conditions. These methods include simple proportional scaling techniques (Boke-Olén et al., 2017; Gaffin et al., 2004; Gao, 2017), machine learning methods (Chen et al., 2020; Wang et al., 2022), and gravity-based approaches (Grübler et al., 2007; Jones and O'Neill, 2016; Reimann et al., 2021). The simple proportional scaling is efficient and transparent but inflexible. The approach has been criticized for leading to exaggerated population counts in densely populated regions, since the percent of allocated population change in each cell are held constant according to the base-year population distribution (van Vuuren et al., 2010). Machine learning methods can find important connections between various input datasets, but they often come with rigid assumptions about the stability and covariability of sets of variables that are scaled down at the same time. These methods are not very adaptable in describing the differences in spatial pattern changes among different scenarios in a qualitative way. Neither proportional scaling nor machine learning approaches applied so far have fully considered the dynamic urbanization rate and spatial changes in settlement patterns, making the future population distribution and densities among areas questionable especially when cities outgrow current admin units.

The gravity-based approach used by Jones and O'Neill (2016, National Center for Atmospheric Research, NCAR) (also used in (Zoraghein and O'Neill, 2020a, 2020b)) can effectively capture the urbanization rate change and urban expansion patterns. The model calculates population potentials based on proximity to certain characteristics and allocates urban and rural population changes separately to the responding urban or rural pixels. In this case, accurate information on the dynamics of the urban extent is crucial for the gravity-based population model to better separately simulate the rural and urban population size and distribution. Traditional gravity models determine urban or rural extents based on population density and continuity with completely urban or rural cells to assign projected national total urban/rural population. However, urban extent extracted based on population density is not globally consistent due to the significant variability in urban forms among and within countries (Taubenböck et al., 2022), and fails to capture the urban expansion when population density decrease. In this regard, this study adopted and improved the gravity-based approach from Jones and O'Neill (2013, 2016) to project the distribution of future population by using more accurate urban extents derived from remote sensing data to capture distinct physical changes in spatial development patterns and to better separately simulating the rural and urban population size and distribution.

The base year population distribution is another crucial component for accurately projecting future gridded populations. Great efforts and significant strides have been made to model the spatial distribution of the current population using different approaches at varying levels of complexity. Common methods range from simple areal weighting techniques to more complex data-intensive dasymetric mapping approaches (Leyk et al., 2019; Sapena et al., 2022). Simple areal weighting techniques uniformly redistribute data from the source unit into target grid cells and do not require ancillary data (Goodchild and Lam, 1980; Mennis and Hultgren, 2006). The approach is computationally efficient and simple in generating globally consistent population distribution maps (for example, the Gridded Population of the World (GPW), (Tobler et al., 1997)). However, it cannot capture fine-scale information, particularly in regions with large census units. The data-intensive dasymetric mapping approaches delineate heterogeneous distributions of the population by incorporating ancillary data with finer spatial scale than the input population data, including land use/land cover and nighttime light (NTL), as modeling variables to develop weighting schemes for population redistribution (Boke-Olén et al., 2017; Chen et al., 2020; Weber et al., 2018). Many well-known globally gridded population datasets, for example, LandScan (Dobson et al., 2000), Global Rural-Urban Mapping Project (GRUMP) (Balk et al., 2006), and WorldPop (Stevens et al., 2015), were produced using the dasymetric mapping approaches. Although these datasets can successfully capature spatial details of population, they usually over-allocate population in low-density population areas and under-allocate population in high-density areas (e.g., urban areas). This is particularly the case in regions where the urban and rural populations are greatly different (Cockx and Canters, 2015; Gaughan et al., 2016; Mennis and Hultgren, 2006).

In this study, we incorporate a physical urban boundary dataset (Li et al., 2020) to downscale the historical administrative population (i.e., for the year 2000) by using a random forest-based dasymetric mapping approach. We adopt an SSPs consistent global impervious surface area dataset (He et al., 2023) to allocate the population changes from national-level SSP projections to the grid cell level by using the gravity-based population downscaling model. Furthermore, we aim to produce high-resolution population projections (30 arc *sec*, approximately 1 km) for the years from 2020 to 2100 that are consistent with the Shared Socio-economic Pathways (SSPs). The objective of this study is not solely to generate a reliable and operational dataset but, more importantly, to demonstrate a framework for spatially explicit downscaling and projecting population for over an extended period, which can be utilized in future applications for scenario-driven events.

2. Methodology and data

2.1. General framework

We proposed an explicit spatiotemporal downscaling and projection framework to map population dynamics from the years 2000 to 2100 under different SSPs, including urban and rural populations. Our framework involves four key procedures: 1) *data preparation*, which

included the collection and preprocessing of census and projected population-related data, and spatially ancillary data (e.g., land cover, topography, the world settlement footprint, and urban extent time series data product); 2) *population downscaling*, which fittted a Random Forest (RF) model for producing the dasymetric population maps for years with observations; 3) *projection of population*, which developed a gravity-based population downscaling model to project population maps for five SSPs, and 4) *accuracy assessment,* which evaluated the accuracy of the downscaled and projected population using census population data. The methodological framework of this study is depicted in Fig. 1. Subsequent sections will delve into the specifics of each procedure.

2.2. Data preparation

A total of thirteen types of datasets were employed to fit the RF model for population downscaling, build a gravity-based population downscaling model for future population projection, and assess the accuracy of modeling results (Table 1). The acquisition and preprocessing of these datasets are elaborated upon in the following section.

The population data including total, urban, and rural population for 2000 and 2010 in China were collected from the fifth and Sixth National Population Census of Mainland China (available from http://www.stats.gov.cn/sj/pcsj/, accessed on 20 September 2023). We excluded the provinces (i.e., Hong Kong, Macao, and Taiwan) with unique political and economic status in this study. We utilized the census data at the county level with 2751 units (corresponding to level 3 of the Global Administrative Unit Layer as defined by the Food and Agriculture Organization) to train the RF model, while data at the township level (corresponding to level 4 of the Global Administrative Unit Layer) were used to assess the accuracy of the resulting population map (Gaughan et al., 2016; Ye et al., 2019). For each SSP, the total population data (from SSP Database, https://tntcat.iiasa.ac.at/SspDb/, accessed on 25 September 2023) and the urbanization level (i.e., urban share, the percent of the urban population available from Chen et al., 2022), were used as quantitative constraints for future population distribution projections. The population datasets from UNPD (United Nations, Department of Economic and Social Afairs, Population Division, available from https://population.un.org/wup/, accessed on 25 September 2023) in 2000 and 2010 were adapted to adjust China's population data, making the historical data consistent with the SSPs projections.

Eight types of spatially ancillary data were collected to fit the RF model and build the gravity-based population downscaling model, including land cover, the gridded urban extent, nighttime lights, building height, roads, world database of protected areas (WDPA), topography (elevation and slope), and climate (monthly temperature and precipitation) data. The land cover data, particularly the categories related human activities, had strong relationship with the population distribution (Tatem et al., 2007). We collected land cover products in raster layers from the ESA (European Space Agency) CCI (Climate Change Initiative), including sub-categories of agriculture, forest, grassland, wetland, settlement, and other. The urban extent, nighttime lights, and building height are very useful for accurately estimating population size and distribution, especially in regions with dense population (Jia et al., 2014; Wang et al., 2016). The gridded urban extents (2000, 2010) used here were generated by Li et al. (2020) based on an automatic delineation framework by using 30 m global artificial impervious area data (Gong et al., 2020), providing the physical boundaries of urban regions. Specifically, The inner non-urban areas from the impervious area data were filled using a kernel density estimation approach and cellular-automata-based urban growth model. The urban fringe areas were dilated and eroded using a morphological approach. The urban extent time series data (Li et al., 2020; He et al., 2023) were also used to separate the urban and rural populations for both RF population downscaling and Gravity population projections. The future gridded urban extents (2020−2100) in this study were produced using a global dataset of gridded artificial impervious area under SSPs (He et al., 2023), according to the automatic delineation framework developed by Li et al. (2020). The nighttime light rasters for 2000 and 2010 were collected from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (available from https://ngdc.noaa.gov/eog/dmsp/download_radcal.html, accessed on 26 September 2023). We also collected vector data, the roads, as ancillary data. The distance to roads and the road density were calculated based on the Global Roads Open Access Dataset (available from http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1, accessed on 25 September 2023). Furthermore, the climate and topography also affect human settlement patterns to some extent, which may directly relate to the population distribution (Stevens et al., 2015). We generated the elevation and slope data from the Digital Elevation Model (DEM) dataset (30-m spatial resolution, available from https://gdex.cr.usgs.gov/gdex/, accessed on 26 September 2023). We downloaded the temperature (°C) and precipitation (mm) raster data from WorldClim 2.0 Beta version 1 (Fick and Hijmans, 2017) with a 30 arc-second spatial resolution. These data depict climate data averaged monthly during the period from 1970 to 2000. Additionally, We obtained the settlement footprint data (10-m spatial resolution, Marconcini et al., 2020) to constrain the historical population extent. The class-based data for both vector and raster data (e.g., individual land cover types, water bodies, presence or absence of protected areas, etc.) are transformed into binary masks, resulting in binary covariates. Using these masks, a distance-to-class raster is computed for each dataset. We resampled all the raster layers to 1-km

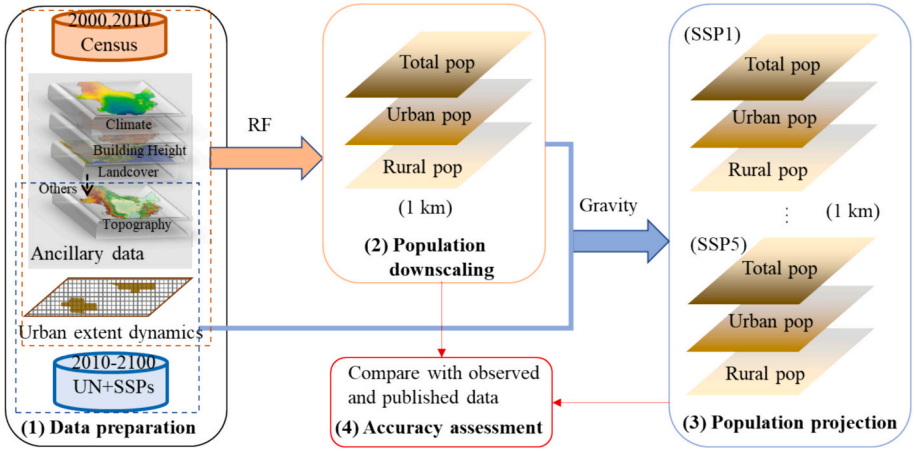


Fig. 1. The framework of downscaling and projection of population, including 1) data preparation, 2) base-year population downscaling, 3) population projection under SSPs, and 4) accuracy assessment. UN: the United Nations datasets; SSPs: Shared Socioeconomic Pathways; pop: population; RF: random forest.

Table 1
Data used for Random Forest-based population downscaling, Gravity-based population projection, and accuracy assessment.

Name	Resolution	Temporal domain	Type	Source	Usage
Census (county-level and township-level urban and rural population)	–	2000, 2010	Text & Vector	The fifth and sixth census (NBSC, 2002, 2012) http://www.stats.gov.cn/sj/pcsj/	Population downscaling and Accuracy assessment
The United Nations country-level urban/rural population	–	1950–2050	Text	UNPD https://population.un.org/wup/Download/Files/WUP2018-F05-Total_Population.xls	Population projection
SSPs population projections	Country-level	2020–2100	Text	SSP Database https://tntcat.iiasa.ac.at/SspDb/	Population projection
SSPs urban share	Country-level	2020–2100	Text	(Chen et al., 2022)	Population projection
Global gridded Urban extent (historical + under 5 SSPs)	~ 1 km	2000–2100	Raster	(Li et al., 2020; He et al., 2023)	Population downscaling (historical) & Population projection (5 SSPs)
The World Settlement Footprint	10 m	2015	Raster	(Marconcini et al., 2020)	Population downscaling
Nightlights	30″ (~900 m)	2000–2011	Raster	DMSP-OLS Stable Nightlights v4, 2015; inter-calibrated, 2016; US NOAA National Geophysical Data Center; https://ngdc.noaa.gov/eog/dmsp/download_radcal.html	Population downscaling
Building height	~ 500 m	2015	Raster	(Zhou et al., 2022)	Population downscaling
Global Roads	–	–	Vector	Global Roads Open Access Data Set, Version 1 (gROADSv1): http://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1	Population downscaling
Global Land Cover	~ 300 m (9 arc-seconds)	2000–2015	Raster	ESA CCI Land Cover https://www.esa-landcover-cci.org/	Population downscaling
World Database of Protected Areas (WDPA)	Comparable to 30″ (~900 m)	1819–2017	Vector	UNEP-WCMC and IUCN	Population downscaling & Population projection
Global Topography (elevation and slope)	3″ (~90 m)	~2000	Raster	http://www.viewfinderpanoramas.org/dem3.html	Population downscaling & Population projection
Monthly temperature and precipitation	30″ (~900 m)	1970–2000	Raster	WorldClim 2.0	Population downscaling

resolution using the bilinear interpolation method.

2.3. Dasymetric urban and rural population mapping

Dasymetric population mapping is one kind of areal interpolation method that seeks to refine population to a finer resolution using auxiliary data with better resolution (Eicher and Brewer, 2001; Mennis, 2009). The fundamental concept involves generating a weight layer (typically gridded), which dictates how census population data can be disaggregated and assigned to finer grid cells within the census unit. A dasymetric population density map can be produced as follows:

$$Pop_i = \frac{Pop}{\sum_{i=1}^{n} W_i} \times W_i \quad (1)$$

where Pop_i is the predicted population of the ith grid cell, W_i is the population-distribution weight for the ith grid cell, n is the number of grid cells within administrative area (i.e., county here), and Pop is total population within the administrative area.

We used a non-parametric method, Random Forest (RF) (Breiman, 2001; Liaw and Wiener, 2002), to generate the population-distribution weight layer and thus downscaled the population at 2000 year to a 1-km pixel level as the base-year population. We also downscaled the 2010 population, working with the 2000 population together, to estimate future projection parameters for next section. We masked uninhabited areas based on the settlement footprint dataset and incorporated a global urban boundary dataset by Li et al. (2020) to assign the urban and rural population separately. Specifically, the 1-km gridded covariates raster layers (e.g., elevation, slope, nighttime light, distance to road, and distance to each landcover type) were first aggregated to administrative regions and subsequently associated with the logarithm of the total administrative census population to train the RF model. Secondly, the well-trained RF model were applied to calculate the population distribution weight for each 1-km gridded cell using the same covariates raster layers. The resulting weight layer was used to determine the regions with the highest probability of experiencing population changes in the future. Then, the weights were converted to urban region and rural region weights according to the urban and rural extent data, where each cell is defined as rural or urban. Finally, the urban and rural weights were used to disaggregate the census urban and rural populations at the administrative level into separate 1-km pixels (Fig. 2).

2.4. Gravity-based population downscaling model and SSP projections

The Shared Socioeconomic Pathways (SSP) project socio-economic global changes up to the year 2100. To do so, they apply five demographic scenarios: 1) sustainability, 2) middle of the road, 3) regional rivalry, 4) inequality, and 5) fossil-fueled development (O'Neill et al., 2017). SSPs have been widely used to project future income inequalities, greenhouse gas concentrations, climate change, and urban expansion.

In this study, we generated scenario-specific spatial projections (at 1-km resolution) for each SSP by refining the changes of national urban and rural population projections. These spatial projections were produced using a gravity-based model to capture the spatial change patterns outlined by each SSP. The process involves (a) estimating the model parameters using historical population distributions to reflect specific patterns of spatial change, (b) choosing parameters that are regionally representative for each SSP, and (c) implementing the allocating procedure of population change for each SSP. Further details regarding the model and parameterization processes are provided below.

To downscale projected changes in national urban and rural population, we used the NCAR gravity-based population downscaling method (Jones and O'Neill, 2016), where population potential is considered as a distance-weighted measure of the population within a specific range surrounding each grid cell. We improved the NCAR approach by using more accurate series of urban extents derived from remote sensing data to capture distinct physical changes in spatial development patterns to better separately simulating the rural and urban population size and distribution. Starting from the gridded base-year population (i.e., the downscaled 2000 population in the previous section), the model includes five basic steps (Fig. 3): (1) Caculation of an urban population potential weight layer; (2) Calculation of a rural population potential weight layer; (3) Allocation of the projected changes of the national

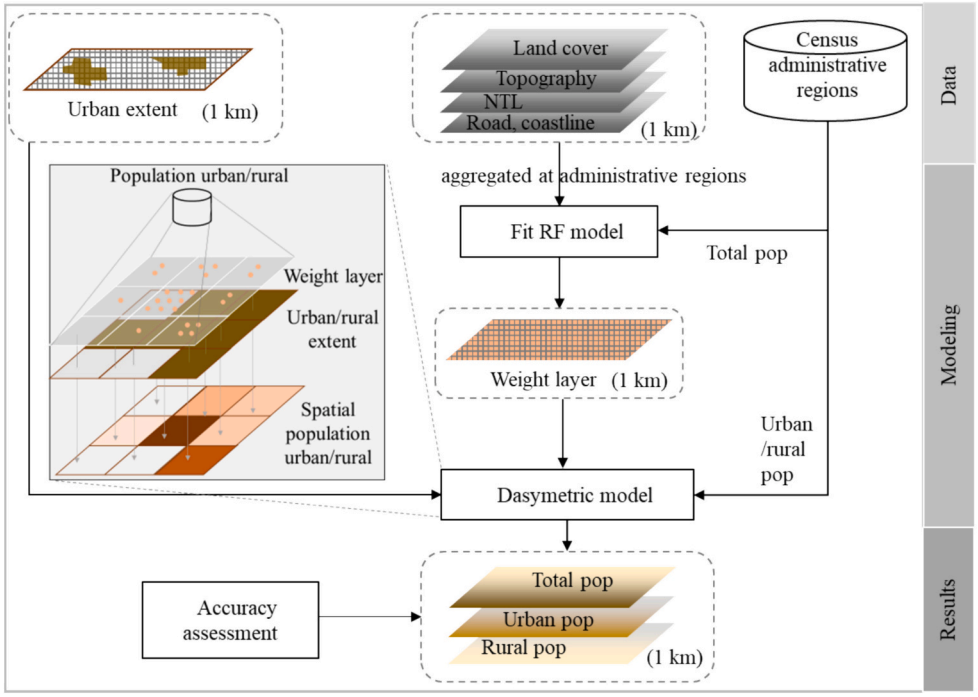


Fig. 2. Flowchart of downscaling urban and rural populations from census units to grid cells using Random Forest and Dasymetric model.

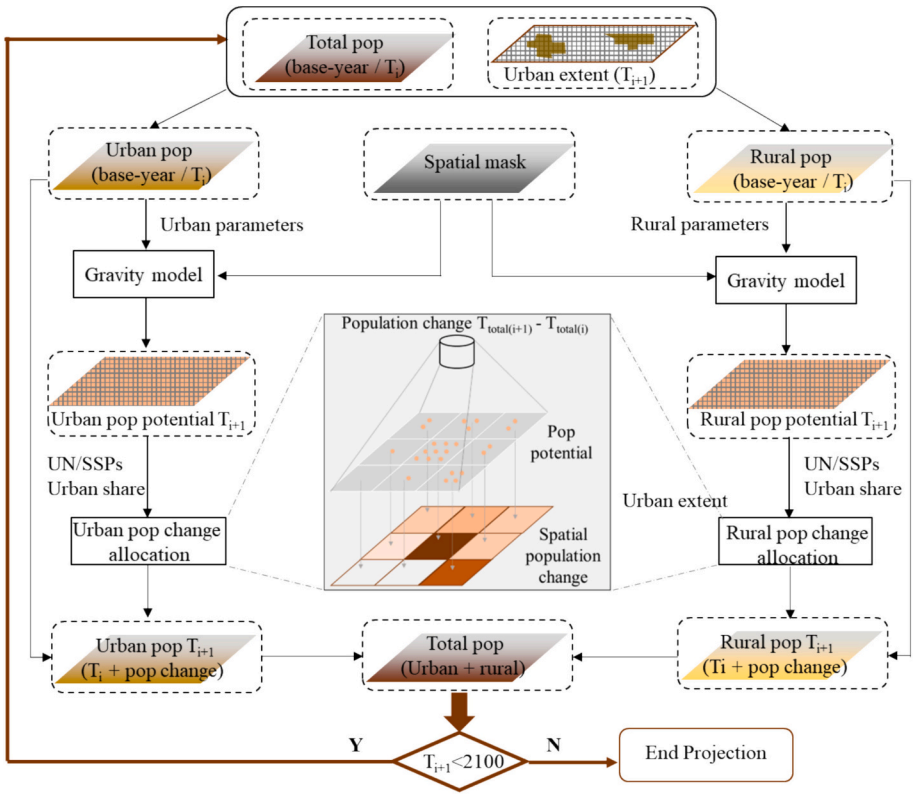


Fig. 3. The flowchart of downscaling the projected changes of national population to grid cells using the gravity-based population downscaling model.

urban population to grid cells in proportion to the corresponding urban potentials; (4) Allocation of the projected changes of the national rural population to grid cells in proportion to the corresponding rural potential; and (5) Redefinition of the population as urban or rural based on the next-decade urban boundary. In this way, the increased urban population is allocated to a cell with a rural population where the cell is redefined as urban at the projection year. These procedures are subsequently reiterated for each 10-year time span.

The population potential weight is developed to distribute projected population changes according to the Eq. (2) (step 4). These data were calculated separately for the urban and rural areas. The population potential weight layers for both the urban and rural are calculated using

the continuous total population distribution, however, the values of a population parameter (α) and a distance-decay parameter (β) used in Eq. (2) for urban and rural populations exhibit variation, capturing the diverse spatial change patterns.

$$v_i = l_i \sum_{j=1}^{m} P_j^{\alpha} e^{-\beta d_{ij}} \quad (2)$$

where v_i represents the potential for cell i, l_i represents the human habitable portion for cell i, j represents an index of the m cells within a 25 km window (according to the Zoraghein and O'Neill (2020a, 2020b)) around cell i, P_j represents the population count within cell j, d represents the geographic distance between cells i and j, and α and β are parameters. l_i is calculated according to topography, protection regions, and land use (details see Jones and O'Neill (2016)).

The α and β parameters are flexible parameters that can be set to reflect the historical observation or future assumptions of population change patterns. Generally, the parameter β influences the extent to which populations spread or sprawl over time, whereas the α determines the degree to which populations consolidate over time (Jones and O'Neill, 2016). To estimate the α and β parameters of the urban and rural for the model that reflect the observed patterns of historical spatial change, we calibrated the model using the observed changes in the urban/rural downscaled population distributions from 2000 to 2010. For each SSP, we adjusted the urban and rural α and β parameters to align with the SSP assumptions regarding spatial development, as outlined in the framework in Zoraghein and O'Neill (2020a, 2020b). The parameters dictate how the population distribution changes over time, influencing whether it tends towards sprawl or concentration; however, the national urbanization rate and total population change are prescribed by the SSPs (Chen et al., 2022; Jiang and O'Neill, 2017). Additionally, we allocated population loss in proportion to the inverse values of the population potentials which are derived based on a period of population gain (Jones and O'Neill, 2016).

2.5. Accuracy assessment

In this study, we took the mainland of China as a case study to demonstrate our newly generated framework of the spatially explicit downscaling and projection of population. We tested the accuracy of the downscaled population (i.e., 2000-year population) using township-level census population of the year 2000. 'Township' is in China the fifth level of the administrative division below the district/county. It includes three types: 1) subdistrict (*Jiedao*), mainly located in the urban centers and the entire population is defined urban. 2) town, where population can be classified urban or rural. 3) country (*Xiang*), where the population is usually classified rural.

We tested the accuracy for the highly urbanized townships (subdistrict), the rural regions (country), and all regions, respectively. We also compared our downscaled dataset with the four most used global datasets (i.e., Worldpop, available from https://www.worldpop.org/datacatalog/; GWP, available from http://sedac.ciesin.columbia.edu/data/collection/gpw-v4; LandScan, available from https://landscan.ornl.gov/; and GRUMP, available from https://sedac.ciesin.columbia.edu/data/collection/grump-v1). We evaluated the accuracy of the projected population using the county-level census population in 2010 and the province-level census population in 2020. Two metrics, specifically root mean square error (RMSE) and mean absolute deviation (MAE), were chosen to assess and contrast the errors present in different population datasets.

2.6. Uncertainty analysis

Future projections naturally contain uncertainties. To account for these, we analyzed the variation in population projections across different SSPs by changing of parameters (α and β). We sampled the parameters from uniform distributions around the reference values specific to each scenario (alpha = [−2.0, 2.0] and beta = [−0.5, 2.0] (Zoraghein and O'Neill, 2020b)), and simultaneously varied instead altering one parameter each time (Saltelli et al., 2019). We calculated the average population density across populated cells. We also calculated the spatial distribution of the coefficient of variation to identify areas that exhibits the highest degree of uncertainty.

3. Results

3.1. Base-year population downscaling

A gridded population dataset at ~1 km (0.00833 decimal degrees) resolution for the year 2000 was constructed for China. Two versions were generated: one where the total population was adjusted to match UNPD national estimates (UNPD 2018), and another where no adjustments were made (Fig. 5a, used for accuracy assessment). Fig. 4a shows the validation results of the RF fit between the modeled total population counts to the original census counts at the township level for the year 2000. The validation analysis shows that our downscaled population, where each data point represents a township, achieves a high accuracy. It correlates to the reported population with an R^2 of 0.83 at a multiplication factor of 1 (Fig. 4a). Furthermore, we compared our downscaled population with four other gridded population datasets (WorldPop, GPW, LandScan, and GRUMP). For those townships, the R^2 of our downscaled dataset is also comparable to or higher than that of other datasets where WorldPop has an R^2 of 0.81, GPW has an R^2 of 0.77, LandScan has an R^2 of 0.73, and GRUMP has an R^2 of 0.63. The Root-mean-square error (RMSE) of our downscaled population (16002) when compared to the observed values is smaller than all of the four other datasets (WorldPop 17994, GWP 31067, LandScan 20926, GRUMP 24940) in 2000. The small RMSE and the large R^2 illustrate that our downscaled population achieved better overall accuracies than the other four datasets.

We evaluated and compared the accuracy of the five datasets in 2000 for the highly urbanized townships and the townships in the rural regions, respectively (Fig. 4a). Among the five datasets, our dataset has values similar to or slightly higher than that of other datasets. Generally, GPW overestimates the population in both, urban and rural areas, whereas LandScan and GRUMP underestimated the population. The WorldPop overestimates the rural population, but still features higher accuracy ($R^2 = 0.81$) than other data sets, but it underestimates the urban population where accuracy levels are lower ($R^2 = 0.68$) (Fig. 4a). We also compared the population density distribution between the prediction and the observation for the five datasets and found a very good performance for all of the datasets at medium population densities. However, this goes with increasing errors at both very low and very high population densities across the majority of the databases (Fig. 4b). At very high population regions, there is a notable tendency for the observed data to be underestimated, suggesting that the modeling process allocated insufficient individuals to densely populated urban centers and instead spread estimations towards less densely populated regions. Compared with the other modeled population datasets, our predictions had less deviation from census data, at both low and high population densities. Our resulting population map depicted higher population densities in urban regions and lower densities in rural regions than other population maps, which is in line with the results in

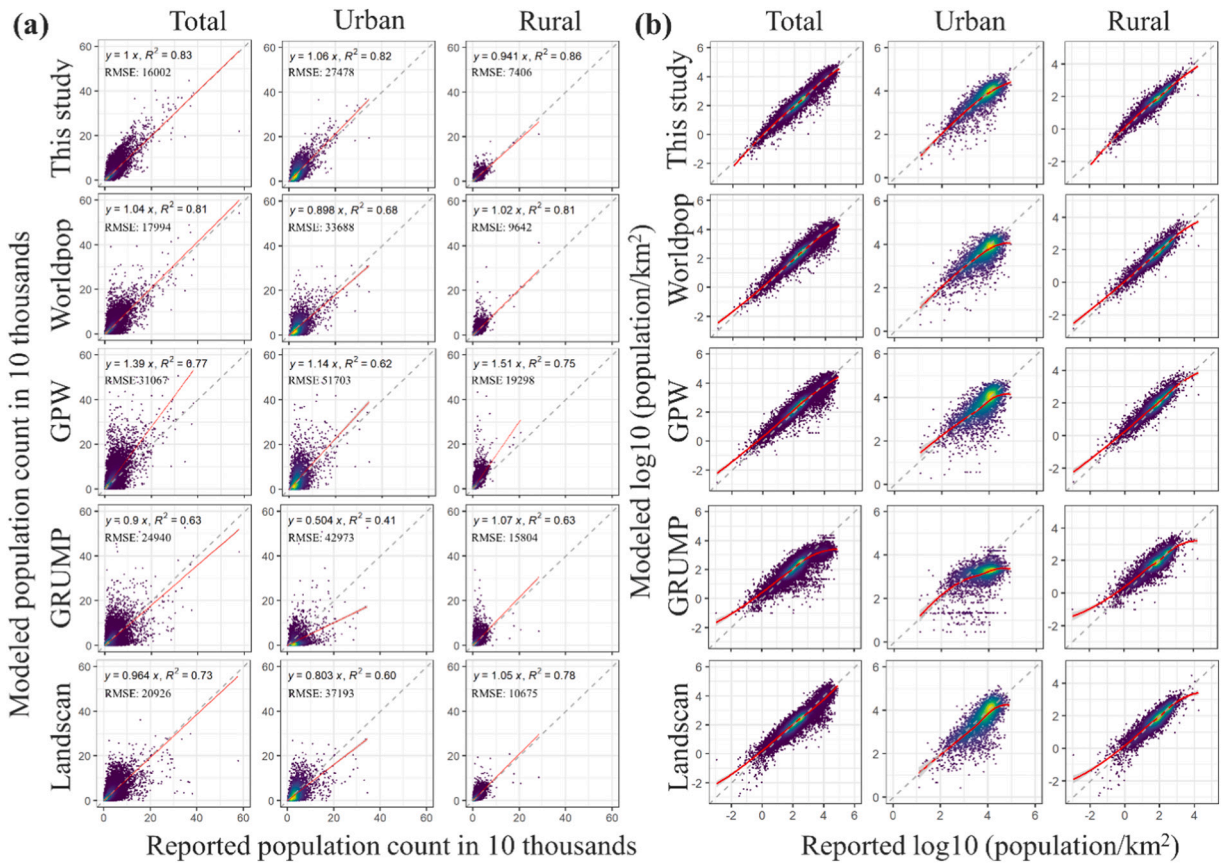


Fig. 4. Accuracy assessment of downscaled population for the year 2000, and its comparison with four well-known global gridded population datasets. (a) Scatter plots between the modeled total, urban, and rural population counts and the corresponding original census counts at township level; (b) scatter plots between the modeled total, urban, and rural population density on a log10-log10 scale and the corresponding original census density at township level. The smoothed trend line, indicating the overall pattern, is derived through LOESS estimation (Cleveland et al., 1992).

Fig. 4. Compared to the GPW and GRUMP, our result is also smoother (Fig. 5).

3.2. SSPs population projection

We calculated the urban and rural parameters (α and β) in the gravity-based population downscaling model using historical populations for the years 2000 and 2010. The populations in 2000 and 2010 were generated using the RF-based dasymetric mapping approach (refer to base-year population downscaling). For urban parameters, both α and β are negative, which implies the population is expanding outward from the exsiting population center (i.e., near the urban boundary development). In contrast, if both rural parameters are positive, reflecting a concentrated population growth pattern where individuals tend to cluster around areas with high population densities.

We conducted model validation by modeling the 2010 and 2020 populations based on the calibrated parameters (α and β) from the historical observation period 2000–2010. Fig. 6 displays validation results by comparing the modeled populations to the observed populations for total, urban, and rural populations in both 2010 and 2020. The results demonstrate that the model performs more effectively for urban populations in contrast to rural populations, with optimal performance observed for total populations, as evidenced by R^2 and the MAE.

We generated a gridded population dataset at 1-km (0.00833 decimal degrees) resolution from 2010 to 2100 for China using an improved gravity model. Fig. 7 (a) and 8 (a) present the future projections for the years 2050 and 2100 for SSP2, respectively. The trends under the other SSPs (supplement Fig. S1) are consistent with SSP2 in that the population would rise until reaching a certain value with considerable urban sprawl and subsequently population would decrease untill 2100. Our projected populations align with both the national population projections (Samir and Lutz, 2017) and urbanization forecasts (Chen et al., 2022), incorporating qualitative assumptions regarding the spatial patterns outlined in each SSP (O'Neill et al., 2017). We compared the differences between five SSPs by subtracting the SSP2 population from other SSPs and normalizing the results for the years 2050 and 2100 (Fig. 7 (b-e), 8 (b-e)). In SSP1, the country experienced a rapid process of urbanization, with urbanization rates of around 85 % (2100), coupled with a decline in population during the latter half of the century, leading to the formation of high-density, compact urban settlements. In SSP2, urbanization proceeds at a less rapid pace, and spatial development exhibits a slightly higher level of concentration compared to historical patterns. SSP3 is characterized by low urbanization rates and less population decrease, resulting in sprawling development. SSP4 has most population decrease in this scenario (as opposed to less in SSP3), and spatial development exhibits a mild degree of concentration. SSP5

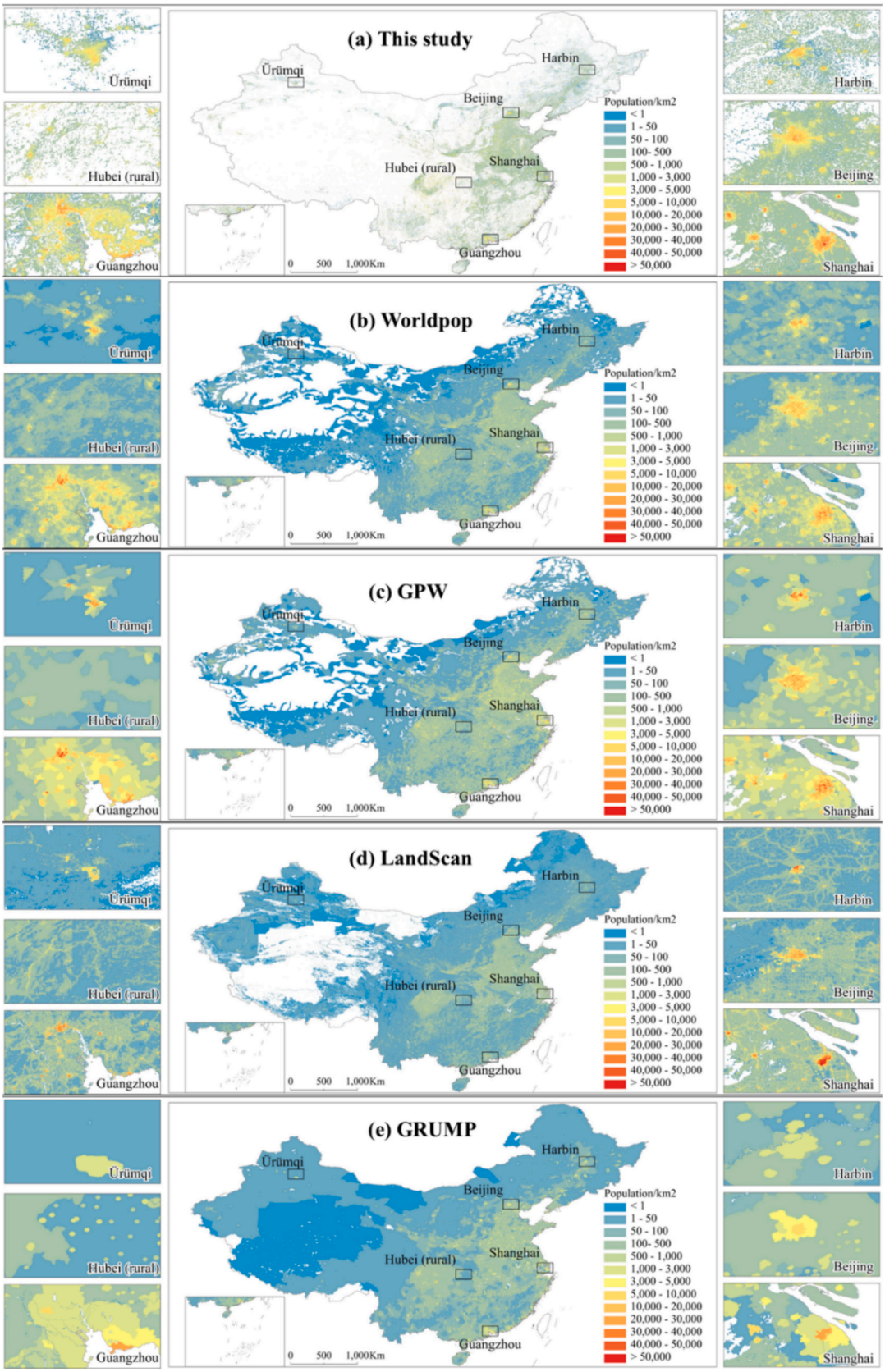


Fig. 5. Downscaled population per cell (1 km) for the year 2000 in mainland China, our result (a) and its comparison with the other four datasets (Worldpop (b), GPW (c), LandScan (d), and GRUMP (e)).

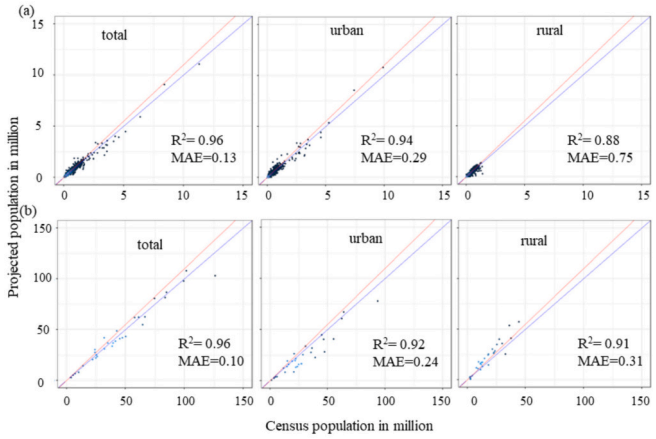


Fig. 6. Scatter plots of the total, urban, and rural population counts between the projected population unit counts at county unit in 2010 (a), and at the province unit in 2020 (b), compared to the corresponding original census counts at the same unit level and time. MAE: Mean Absolute Deviation.

exhibits urbanization levels akin to SSP1, but has considerable urban sprawl.

3.3. Uncertainty analysis

The uncertainties analysis involving variations in input parameters per SSP reveals that the populated cells changes with different input parameters and population density depends on populated cells, which signifies uncertainty within the scenarios (Fig. 9a). The scenarios exhibit significant divergence from each other. The uncertainty ranges regarding the extent of population densities are largest in SSP5 for both temporal projections for the years 2050 and 2100. Uncertainty is positively skewed in SSP1 and SSP5, and negatively skewed in SSP3 and SSP4, with high outliers existing in SSP5. Negative skewness of population density is associated with the unpopulated areas that can newly populated. The increase in average density does not follow a linear trend with population counts, as populated cells that are less desirable are tend to be eliminated.

The spatial distribution of coefficient of variation in population density across the simulation results indicates generally low uncertainty (< 0.1) in urban centers such as Beijing, and in regions where the likelihood of changes is minimal (Fig. 9b). The coefficient is zero in

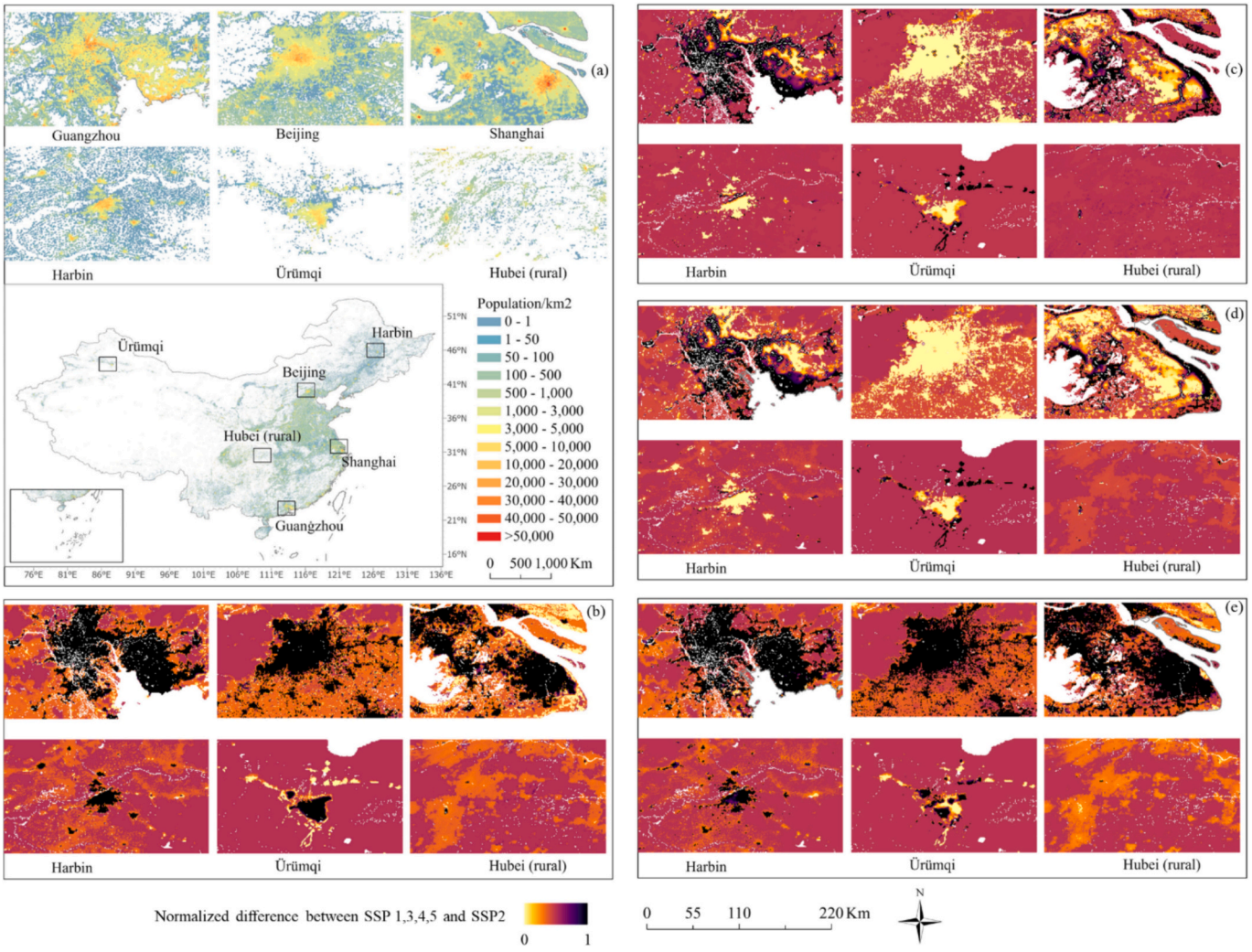


Fig. 7. Projected population per cell (1 km) under SSP 2 (a), and a selected set of the normalized population difference grids SSP 1 - SSP 2 (b), SSP 3 - SSP 2 (c), SSP 4 - SSP 2 (d), SSP 5- SSP 2 (e) in six sample regions (Guangzhou, Beijing, Shanghai, Ürümchi, Harbin, and rural region in Hubei) in 2050. The larger number of the normalized population difference represents higher population than SSP2.

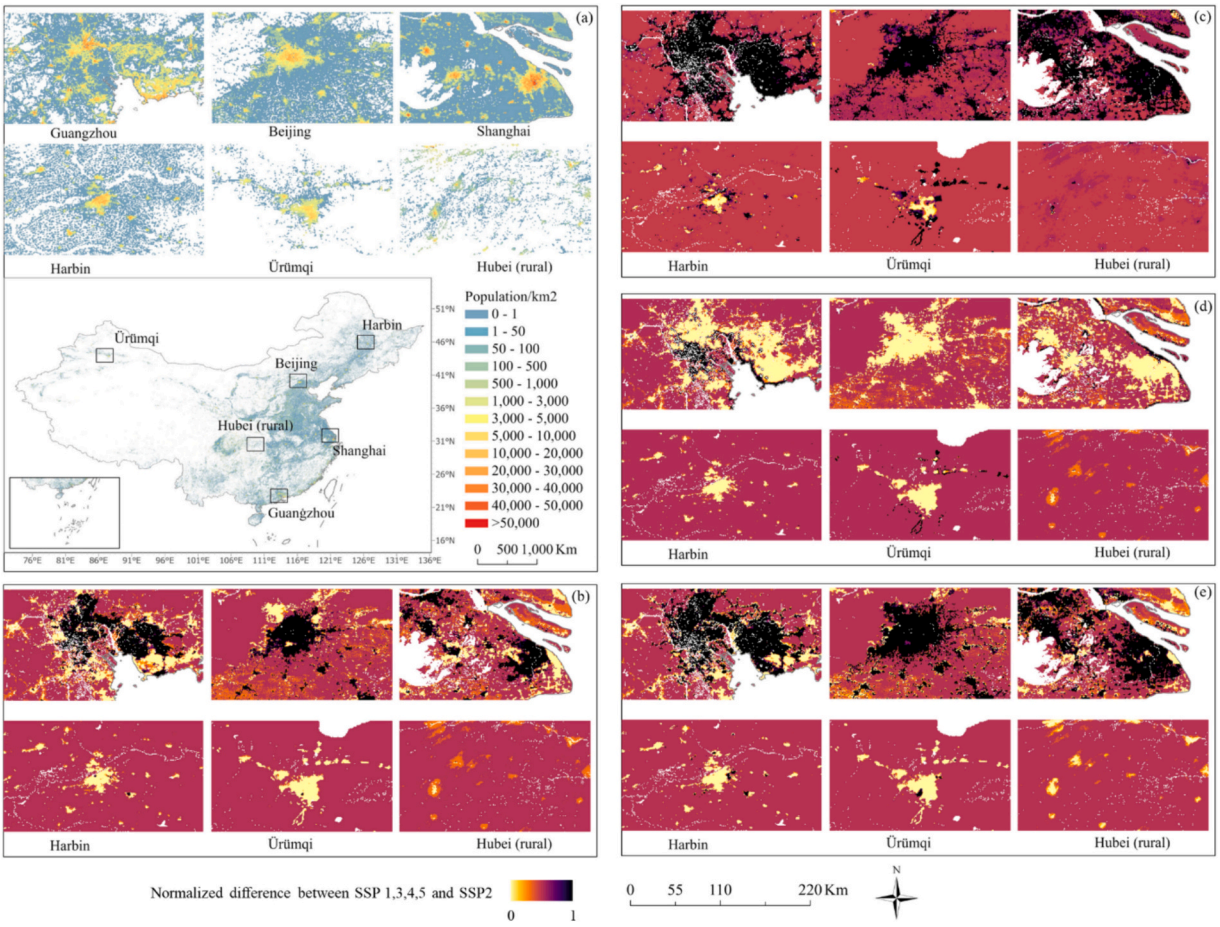


Fig. 8. Projected population per cell (1 km) under SSP 2 (a), and a selected set of the normalized population difference grids SSP 1 - SSP 2 (b), SSP 3 - SSP 2 (c), SSP 4 - SSP 2 (d), SSP 5- SSP 2 (e) in six sample regions (Guangzhou, Beijing, Shanghai, Ürümchi, Harbin, and rural region in Hubei) in 2100. The larger number of the normalized population difference represents higher population than SSP2.

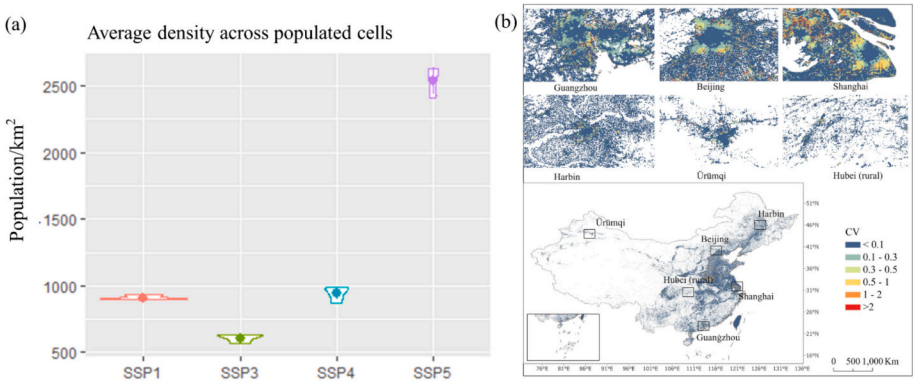


Fig. 9. Uncertainties due to model parameters. (a) Uncertainty analysis of uncertainty distributions on average population density across populated cells for the year 2100, plotted as violin plots. The height of the plots represents the ranges of uncertainty (bandwidths). The violin plots depict the distributions of results, where positive skewness is indicated by fat bottom ends, and negative skewness is indicated by fat upper ends. The horizontal lines within the violin plots display the mean value along with its standard deviation (mean ± sd). The points in each plot provide the mean value of the distribution. (b) Spatial coefficient of variation (CV) on population density in 2100 for SSP5.

unpopulated areas in any simulation. Uncertainty is high at urban boundaries with higher allocation possibilities. The greatest uncertainties are found in SSP5, where population sprawl opens up numerous possibilities. There are clear uncertainty hotspots. The uncertainty hotspots share the characteristic of initially having low population numbers but being more appealing or attractive.

4. Discussion

Dasymetric mapping technique relies on the statistical relationships between population from source units and the aggregated ancillary information to generate a finer-scale predictive weighting layer to redistribute population from coarser source units to finer target units (Eicher and Brewer, 2001; Mennis, 2009). The population density range in the

target unit varies from that of the source due to the changes in scales, and the imposed bimodality in population density between urban and rural regions at the finer scale distribution might not exist at the source scale distribution, resulting in over-allocation of rural populations and under-allocation of urban populations in some regions (Sinha et al., 2019). The statistical approach (Random Forest) assigns non-zero population to all land grid cells leading to the misallocation of the population to uninhabited regions (Leyk et al., 2019). In this study, our approach masked uninhabited areas based on the settlement footprint dataset and incorporated a global urban boundary dataset by Li et al. (2020) to allocate the urban and rural population separately, and thus reduce the misallocation among uninhabited, urban, and rural areas. The under-estimation of the population in urban regions and over-estimation in rural regions in the other datasets were reduced in our downscaled base-year population.

Our future population distribution projection approach incorporates an SSP-consistent global urban extent dataset under the Coupled Model Intercomparison Project 6, to address the misdivision of future urban extent based on population densities. Our projected distribution results show that urban population densities decreased from 2000 to 2020. This suggests that the urban land expansion rate is higher than or equal to that of population growth, which is supported by many studies (Angel et al., 2010; Marshall, 2007; Su et al., 2017). Over the past decades, China has witnessed unparalleled urbanization characterized by a significant surge in urban population and rapid, extensive expansion of urban land (Taubenböck et al., 2012; Shuqing et al., 2015; Zhao et al., 2015). Studies have found that many Chinese cities have shown a trend of urban land expanding at a faster pace than population growth (Su et al., 2017; Zhao et al., 2015). It has also been reported that the average urban population density across 284 Chinese cities decreased from 11,073 to 9381 people per square kilometer over a five-year period (2006–2011) (Su et al., 2017). Our modeling results illustrated that the decline would continue to the end of the 21st century due to the combined effect of urban land expansion and population decline. Declining population densities and urban sprawl have been discussed as an ecological challenge with respect to land consumption, mobility behavior, building energy use, among many other things (e.g. Glaeser, 2010; Güneralp et al., 2017; Taubenböck, 2021), but also in terms of social impacts (e.g. Salvati et al., 2018; Sapena et al., 2021). Accordingly, this spatial population projection is an important anchor point for matching planning visions with the here projected scenarios. It may support better informed or at least more conscious decision-making for balancing settlement demands with social, ecological and economic developments.

Inherent uncertainty exists in all population projections, especially when focusing on smaller, more specific geographic areas. As the geographic unit decreases in size, the task of accurately forecasting the population becomes progressively more difficult. What's more, this uncertainty is propagated further in the projections from the base year. The purpose of developing the SSPs was to organize uncertainty regarding future socio-economic development trends in a systematic manner. Although uncertainty surrounds which SSP trajectory the future might adopt, if any, uncertainty also exists within each SSP and within the methodologies utilized to explore them. Model outputs are contingent upon the model structure, parameters, and inputs, which may not necessarily be logically true (Merkle et al., 2022; Rounsevell et al., 2021). To reduce uncertainty regarding parameter configurations, we calibrate parameters from historic observations. However, it is limited by the fact that the model reproduces spatial patterns in future scenarios mirroring those observed in the past period, in our case this is the decade from 2000 to 2010. In this regard, the parameters used for future scenarios were adjusted based on the SSP assumptions about spatial development instead of the parameters from historic observations. This enables future projections to diverge from historical trends, aligning with the scenarios outlined in the SSP narratives. Our work is also limited to the inconsistency in input data. The remote-sensing-based physical urban boundaries may not be fully consistent with the official urban boundaries. Besides, although our urban extent time series considered climate change, we did not directly account for climate change's impacts on population distribution when implementing the

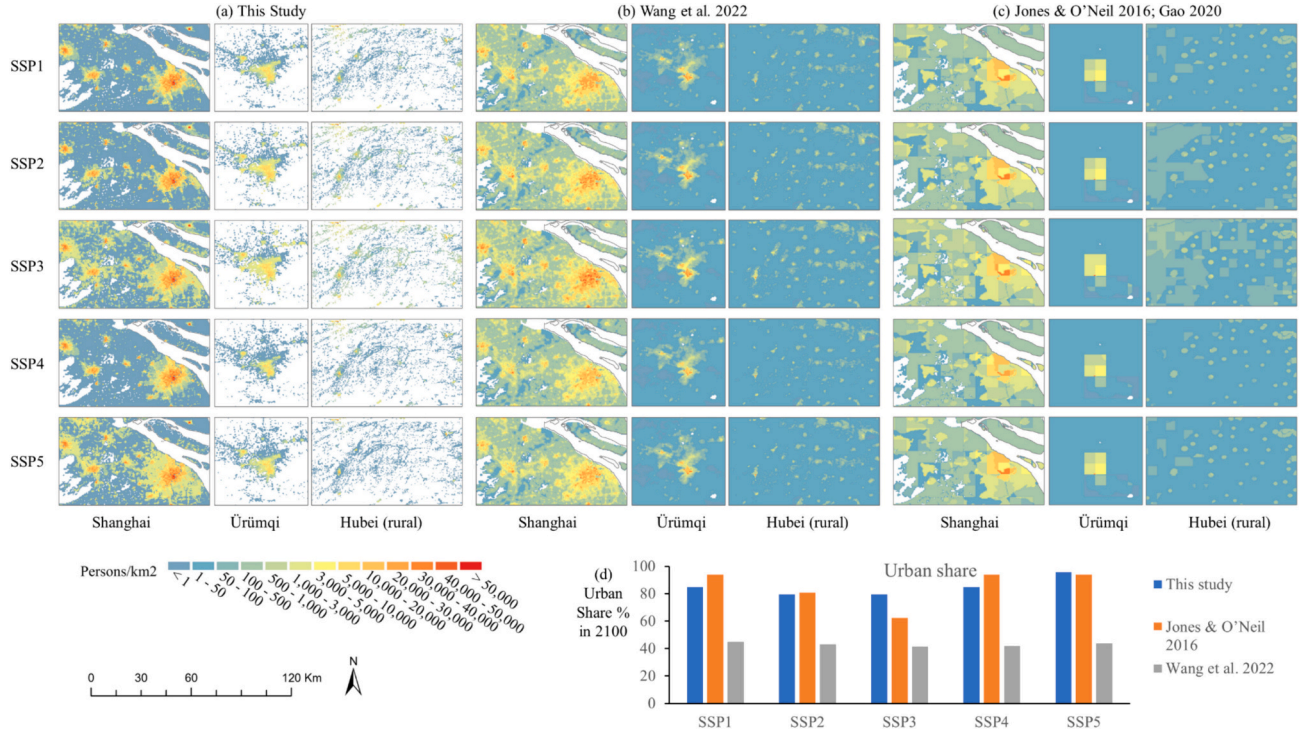


Fig. 10. Comparison of our projections with existing related datasets in three representable sample regions (large city, Shanghai; small city, Ürümqi; and rural region in Hubei). Figures a (this study), b (Wang et al., 2022), and c (Jones and O'Neill, 2016; Gao 2020), are modeled population maps. Figure d is the comparison of urbanization rate among these three datasets in 2100.

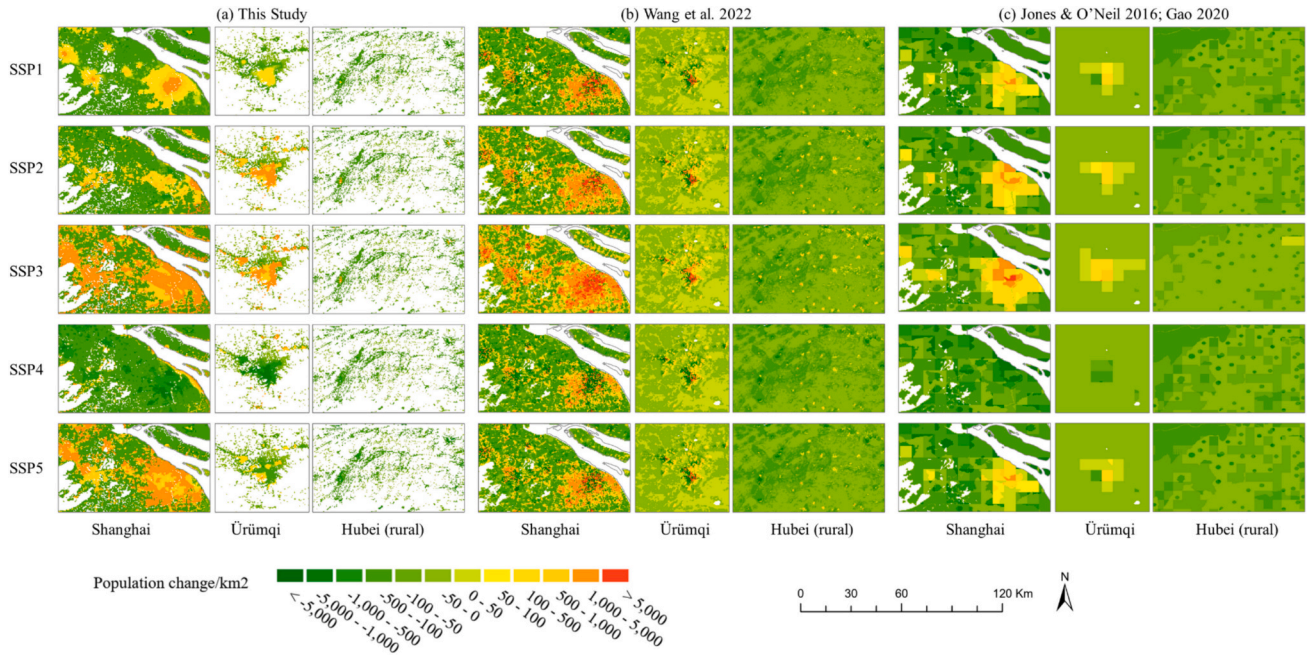


Fig. 11. Comparison with existing related datasets in three sample regions (large city, Shanghai; small city, Ürümqi; and rural region in Hubei). Figures a (this study), e (Jones and O'Neill, 2016; Gao 2020), and f (Wang et al., 2022) population change (2100−2000) maps.

human habitable level in our projection process.

In this study, we compare our analysis with existing related datasets. These include projection datasets from Jones and O'Neill (2016) (Figs. 10 (c) and 11 (c)) and Wang et al. (2022) (Figs. 10 (b) and 11 (b)). The dataset from Jone & O'Neil was produced at a 0.125-degree grid resolution for the globe from 2000 to 2100 by using a gravity-based population downscaling model. Gao (2020) then downscaled the dataset from 0.125 degrees to a 1 km grid using a simple proportional scaling method based on the historical population distribution. Wang et al. (2022) generated a global population distribution dataset from 2020 to 2100 using a random forest model following the static relationships between census data and covariates. However, this method cannot flexibly account for the SSPs assumptions in spatial development patterns. Beyond this, it fails to take the changes in the urbanization rate into account. Figs. 10 and 11 show that our dataset seems to better fit with the urbanization rate and the spatial distribution. This is due to the fact that our results are consistent with the assumptions of the SSPs. And, our results reveal the smoother spatial change in population compared to the dataset that was produced by Jones & O'Neil in 2000 and further downscaled by Gao in 2020.

Gridded population estimates with spatial detail have consistently demonstrated their utility for planning purposes. Our generated long-term population dataset at 1 km resolution can be directly integrated with multiple various scenario-driven events or environmental datasets (e.g., flooding, air pollutants, etc.) at comparable spatial resolutions. Our dataset separates the urban and rural sights and considers the distribution of population changes instead of directly projecting the future population distribution. Thus, the impact of urban environmental change, including e.g. those of habitat loss, human health, environment management, disaster prevention, public area planning, and spatial evaluation of populations at risk can be modeled (Laaidi et al., 2012; Lin et al., 2015). The projected population distribution can be leveraged for long-range urban and infrastructure planning, as well as scientific modeling to anticipate land use changes. Applications of our model encompass integrating various scenario-driven events to generate a spectrum of spatially detailed population forecasts for suitability modeling, urban planning and planning consequence assessment, disaster mitigation and assisstance, and spatially evaluating populations at risk.

5. Conclusions

In this study, we proposed a framework for spatially explicit downscaling and projecting population distributions by accounting for the distinct urban extent dynamics (i.e., urban expansion). We used a random forest-based dasymetric mapping approach to separately disaggregate county-level urban and rural populations at 1 km resolution for historical population distribution (i.e., for the year 2000 in this study) according to a physical urban boundary product. Using the disaggregated historical population distribution as base-year distribution, we used an improved gravity-based population downscaling model that directly incorporates urban expansion patterns to allocate the population changes from national-level SSP projections to the grid cell level. Based on this framework, we produced a gridded population dataset at 1-km resolution for China from the year 2000 to 2100, including the urban and rural populations. The insufficient allocation of the population in urban areas, the excessive allocation in rural areas, and the misallocation to uninhabited areas in other published datasets were mitigated in our historical population map. Our scenario projections are consistent with the urbanization rate and portray the spatial population dynamics indicated by each of the SSP narratives, in line with urban spatial development patterns. This work provides a framework for population downscaling and projection in other countries. The datasets were produced with spatially and temporally consistent methods and inputs, and can be helpful for long-term socioeconomic studies.

CRediT authorship contribution statement

Wenru Xu: Writing – original draft, Methodology, Formal analysis, Data curation. **Yuyu Zhou:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Hannes Taubenböck:** Writing – review & editing. **Eleanor C. Stokes:** Writing – review & editing. **Zhengyuan Zhu:** Writing – review & editing. **Feilin Lai:** Writing – review & editing, Data curation. **Xuecao Li:** Writing –

review & editing. **Xia Zhao:** Writing – review & editing.

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

Data will be made available on request.

Acknowledgments

This research was funded by The University of Hong Kong HKU-100 Scholars Fund, the HKU Social Sciences Internal Seed Grant Scheme, and the National Science Foundation under Grant CBET-2041859.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2024.173623.

References

Akhmat, G., Zaman, K., Shukui, T., Irfan, D., Khan, M.M., 2014. Does energy consumption contribute to environmental pollutants? Evidence from SAARC countries. Environ. Sci. Pollut. Res. 21, 5940–5951.
Angel, S., Parent, J., Civco, D.L., Blei, A., 2010. The Persistent Decline in Urban Densities: Global and Historical Evidence of 'sprawl'. Lincoln Institute of Land Policy Cambridge, In.
Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: methods, applications and data. Adv. Parasitol. 62, 119–156.
Boke-Olén, N., Abdi, A.M., Hall, O., Lehsten, V., 2017. High-resolution African population projections from radiative forcing and socio-economic models, 2000 to 2100. Sci. Data 4 (1), 1–9.
Breiman, L., 2001. Random forests. In: Machine Learning, 45, pp. 5–32.
Chen, Y., Li, X., Huang, K., Luo, M., Gao, M., 2020. High-resolution gridded population projections for China under the shared socioeconomic pathways. *Earth's*. Future 8 (6), e2020EF001491.
Chen, S., Huang, Q., Muttarak, R., Fang, J., Liu, T., He, C., et al., 2022. Updating global urbanization projections under the Shared Socioeconomic Pathways. Sci. Data 9 (1), 137.
Cleveland, W.S., Grosse, E., Shyu, W.M., 1992. Local regression models. In: Chambers, J. M., Hastie, T.J. (Eds.), Statistical Models in S. Wadsworth & Brooks/Cole. (Springer-Verlag), Chapter, New York, p. 8.
Cockx, K., Canters, F., 2015. Incorporating spatial non-stationarity to improve dasymetric mapping of population. Appl. Geogr. 63, 220–230.
Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote. Sens. 66 (7), 849–857.
Eicher, C.L., Brewer, C.A., 2001. Dasymetric mapping and areal interpolation: implementation and evaluation. Cartogr. Geogr. Inf. Sci. 28 (2), 125–138.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.
Gaffin, S.R., Rosenzweig, C., Xing, X., Yetman, G., 2004. Downscaling and geo-spatial gridding of socio-economic projections from the IPCC Special Report on Emissions Scenarios (SRES). Glob. Environ. Chang. 14 (2), 105–123.
Gao, J., 2017. Downscaling Global Spatial Population Projections from 1/8-Degree to 1-km Grid Cells. National Center for Atmospheric Research, Boulder, CO, USA, p. 1105.
Gaughan, A.E., Stevens, F.R., Huang, Z., Nieves, J.J., Sorichetta, A., Lai, S., et al., 2016. Spatiotemporal patterns of population in mainland China, 1990 to 2010. Sci. Data 3 (1), 1–11.
Glaeser, E., 2010. Triumph of the City. Penguin Press HC 352 S.
Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., Zhou, Y., 2020. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. Remote Sens. Environ. 236, 111510.
Goodchild, F.M., Lam, N.S.-N., 1980. Areal interpolation: a variant of the traditional spatial problem. Geo-processing 1 (3), 297–312.
Grübler, A., O'Neill, B., Riahi, K., Chirkov, V., Goujon, A., Kolp, P., et al., 2007. Regional, national, and spatially explicit scenarios of demographic and economic change based on SRES. Technol. Forecast. Soc. Chang. 74 (7), 980–1029.
Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P.L., Fragkias, M., Li, X., Seto, K.C., 2017. Global scenarios of urban density and its impacts on building energy use through 2050. Proc. Natl. Acad. Sci. 114 (34), 8945–8950.
He, W., Li, X., Zhou, Y., Shi, Z., Yu, G., Hu, T., Wang, Y., Huang, J., Bai, T., Sun, Z., Liu, X., 2023. Global urban fractional changes at a 1 km resolution throughout 2100 under eight SSP-RCP scenarios. Earth Syst. Sci. Data Discuss. 1–30.
Hu, J., Wang, Y., Taubenböck, H., Zhu, X.X., 2021. Land consumption in cities: a comparative study across the globe. Cities 113.
Jia, P., Qiu, Y., Gaughan, A.E., 2014. A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida. Appl. Geogr. 50, 99–107.
Jiang, L., O'Neill, B.C., 2017. Global urbanization projections for the Shared Socioeconomic Pathways. Glob. Environ. Chang. 42, 193–199.
Jones, B., O'Neill, B.C., 2013. Historically grounded spatial population projections for the continental United States. Environ. Res. Lett. 8, 044021.
Jones, B., O'Neill, B.C., 2016. Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. Environ. Res. Lett. 11 (8), 084003.
Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451 (7181), 990–993.
Laaidi, K., Zeghnoun, A., Dousset, B., Bretin, P., Vandentorren, S., Giraudet, E., Beaudeau, P., 2012. The impact of heat islands on mortality in Paris during the August 2003 heat wave. Environ. Health Perspect. 120 (2), 254–259.
Leyk, S., Gaughan, A., Adamo, S., de Sherbinin, A., Balk, D., Freire, S., et al., 2019. The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. Earth Syst. Sci. Data 11 (3), 1385–1409 (Review).
Li, K., Fang, L., He, L., 2019. How population and energy price affect China's environmental pollution? Energy Policy 129, 386–396.
Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., et al., 2020. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. Environ. Res. Lett. 15 (9), 094044.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Lin, C., Li, Y., Yuan, Z., Lau, A.K., Li, C., Fung, J.C., 2015. Using satellite remote sensing data to estimate the high-resolution distribution of ground-level PM2. 5. Remote Sens. Environ. 156, 117–128.
Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M., 2020. Outlining where humans live, the World Settlement Footprint 2015. Sci. Data 7 (1), 242.
Marshall, J.D., 2007. Urban land area and population growth: a new scaling relationship for metropolitan expansion. Urban Stud. 44 (10), 1889–1904.
Marzi, S., Mysiak, J., Essenfelder, A.H., Pal, J.S., Vernaccini, L., Mistry, M.N., et al., 2021. Assessing future vulnerability and risk of humanitarian crises using climate change and population projections within the INFORM framework. Glob. Environ. Chang. 71, 102393.
Mennis, J., 2009. Dasymetric mapping for estimating population in small areas. Geogr. Compass 3 (2), 727–745.
Mennis, J., Hultgren, T., 2006. Intelligent dasymetric mapping and its comparison to other areal interpolation techniques. Cartogr. Geogr. Inf. Sci. 33 (3), 179–194.
Merkle, M., Alexander, P., Brown, C., Seo, B., Harrison, P.A., Harmáčková, Z.V., Pedde, S., Rounsevell, M., 2022. Downscaling population and urban land use for socio-economic scenarios in the UK. Reg. Environ. Chang. 22 (3), 106.
Müller, I., Erbertseder, T., Taubenböck, H., 2022. Tropospheric NO2: explorative analyses of spatial variability and impact factors. Remote Sens. Environ. 270, 112839.
O'Neill, B.C., Kriegler, E., Ebi, K.L., Kemp-Benedict, E., Riahi, K., Rothman, D.S., et al., 2017. The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. Glob. Environ. Chang. 42, 169–180.
Reimann, L., Jones, B., Nikoletopoulos, T., Vafeidis, A.T., 2021. Accounting for internal migration in spatial population projections—a gravity-based modeling approach using the Shared Socioeconomic Pathways. Environ. Res. Lett. 16 (7), 074025.
Rosa, E.A., Dietz, T., 2012. Human drivers of national greenhouse-gas emissions. Nat. Clim. Chang. 2 (8), 581–586.
Rounsevell, M.D., Arneth, A., Brown, C., Cheung, W.W., Gimenez, O., Holman, I., et al., 2021. Identifying uncertainties in scenarios and models of socio-ecological systems in support of decision-making. One Earth 4 (7), 967–985.
Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., et al., 2019. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. Environ. Model Softw. 114, 29–39.
Salvati, L., Zambon, I., Chelli, F.M., Serra, P., 2018. Do spatial patterns of urbanization and land consumption reflect different socioeconomic contexts in europe? Sci. Total Environ. 625, 722–730.
Samir, K., Lutz, W., 2017. The human core of the shared socioeconomic pathways: population scenarios by age, sex and level of education for all countries to 2100. Glob. Environ. Chang. 42, 181–192.
Sapena, M., Wurm, M., Taubenböck, H., Tuia, D., Ruiz, L.A., 2021. Estimating quality of life dimensions from urban spatial pattern metrics. Comput. Environ. Urban. Syst. 85, 101549.
Sapena, M., Kühnl, M., Wurm, M., Patino, J.E., Duque, J.C., Taubenböck, H., 2022. Empiric recommendations for population disaggregation under different data scenarios. PLoS One 1–29.
Shuqing, Z., Decheng, Z., Chao, Z., Yan, S., Wenjia, W., Shuguang, L., 2015. Spatial and Temporal Dimensions of Urban Expansion in China.
Sinha, P., Gaughan, A.E., Stevens, F.R., Nieves, J.J., Sorichetta, A., Tatem, A.J., 2019. Assessing the spatial sensitivity of a random forest model: application in gridded population modeling. Comput. Environ. Urban. Syst. 75, 132–145.
Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS One 10 (2), e0107042.

Su, H., Wei, H., Zhao, J., 2017. Density effect and optimum density of the urban population in China. Urban Stud. 54 (7), 1760–1777.
Tatem, A.J., Noor, A.M., Von Hagen, C., Di Gregorio, A., Hay, S.I., 2007. High resolution population maps for low income nations: combining land cover and census in East Africa. PLoS One 2, e1298.
Tatem, A.J., Adamo, S., Bharti, N., Burgert, C.R., Castro, M., Dorelien, A., et al., 2012. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. Popul. Health Metrics 10 (1), 1–14.
Taubenböck, H., 2021. How we live and what that means – a character study with data from space. In: International Geoscience and Remote Sensing Symposium (IGARSS) 12.-16. July 2021, Brüssel, Belgien.
Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., Dech, S., 2012. Monitoring of mega cities from space. Remote Sens. Environ. 117, 162–176.
Taubenböck, H., Droin, A., Standfuß, I., Dosch, F., Sander, N., Milbert, A., Eichfuss, S., Wurm, M., 2022. To be, or not to be 'urban'? A multi-modal method for the differentiated measurement of the degree of urbanization. Comput. Environ. Urban. Syst. 95.
Tobler, W., Deichmann, U., Gottsegen, J., Maloy, K., 1997. World population in a grid of spherical quadrilaterals. Int. J. Popul. Geogr. 3 (3), 203–225.
van Vuuren, D.P., Smith, S.J., Riahi, K., 2010. Downscaling socioeconomic and emissions scenarios for global environmental change research: a review. Wiley Interdiscip. Rev. Clim. Chang. 1 (3), 393–404.
Wang, S., Tian, Y., Zhou, Y., Liu, W., Lin, C., 2016. Fine-scale population estimation by 3D reconstruction of urban residential buildings. Sensors 16, 1755.
Wang, X., Meng, X., Long, Y., 2022. Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. Sci. Data 9 (1), 563.
Weber, E.M., Seaman, V.Y., Stewart, R.N., Bird, T.J., Tatem, A.J., McKee, J.J., et al., 2018. Census-independent population mapping in northern Nigeria. Remote Sens. Environ. 204, 786–798.
Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., Buckee, C. O., 2012. Quantifying the impact of human mobility on malaria. Science 338 (6104), 267–270.
Wesolowski, A., Buckee, C.O., Bengtsson, L., Wetter, E., Lu, X., Tatem, A.J., 2014. Commentary: containing the Ebola outbreak-the potential and challenge of mobile network data. PLoS Curr. 6.
Xu, C., Kohler, T.A., Lenton, T.M., Svenning, J.-C., Scheffer, M., 2020. Future of the human climate niche. Proc. Natl. Acad. Sci. 117 (21), 11350–11355.
Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., et al., 2019. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. Sci. Total Environ. 658, 936–946.
Zhao, S., Zhou, D., Zhu, C., Sun, Y., Wu, W., Liu, S., 2015. Spatial and temporal dimensions of urban expansion in China. Environ. Sci. Technol. 49 (16), 9600–9609.
Zhou, Y., Li, X., Chen, W., Meng, L., Wu, Q., Gong, P., Seto, K.C., 2022. Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the Global South. Proc. Natl. Acad. Sci. 119 (46), e2214813119.
Zoraghein, H., O'Neill, B.C., 2020a. A spatial population downscaling model for integrated human-environment analysis in the United States. Demogr. Res. 43, 1563–1606.
Zoraghein, H., O'Neill, B.C., 2020b. US state-level projections of the spatial distribution of population consistent with shared socioeconomic pathways. Sustainability 12 (8), 3374.