High-Resolution Gridded Population Projections for China Under the Shared Socioeconomic Pathways

Yimin Chen1, Xia Li2, Kangning Huang3, Ming Luo1, and Minyi Gao2

1School of Geography and Planning, and Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, PR China, 2School of Geographic Sciences, and Key Lab of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China, 3Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA

Abstract Gridded population projections consistent with the shared socioeconomic pathways (SSPs) are critical for the studies of climate change impacts and their mitigation. Existing gridded population projections under the SSPs have relatively coarse resolution and issue of overestimation in populated areas, which further bias the analysis of climate change impacts. In this study, we proposed a scheme by integrating high-resolution historical population maps and machine learning models to predict future built-up land and population distributions, which were rendered consistent with the SSPs. Using this proposed method in China, we generated a set of 100-m SSPs population maps for China from 2015 to 2050 at 5-year intervals. Our projections revealed different spatial structures for the population distribution at the grid level and three modes of provincial population change across the five SSPs from 2015 to 2050. By applying the 100-m SSPs population grids, we showed that, from 2015 to 2050, exposure to extreme heat in China will increase by 121–136% and 164–191% under the representative concentration pathways 4.5 and 8.5, respectively. We also found a severe spatial bias in the existing 1/8° SSPs population grids, i.e., 30–43% of the estimated population is wrongly allocated in cropland, forest, and pastureland. This bias results in substantial underestimation of extreme heat exposure in high-density metropolitan areas and overestimation in medium and low-density areas.

Plain Language Summary In this study, we proposed a scheme by integrating high-resolution historical population maps and machine learning models to predict future built-up land and population distributions, which were rendered consistent with the SSPs. Using this proposed method, we generated a set of 100-m SSPs population maps for China from 2015 to 2050 at 5-year intervals. Our projections revealed different spatial structures for the population distribution at the grid level and three modes of provincial population change across the five different SSPs from 2015 to 2050. By applying the 100-m SSPs population grids, we showed that, from 2015 to 2050, exposure to extreme heat in China will increase by 121–136% and 164–191% under the representative concentration pathways 4.5 and 8.5, respectively. We further compared our projections with the existing 1/8° SSPs population grids and we found a severe spatial bias in the 1/8° SSPs population grids: 30–43% of the estimated population is wrongly allocated in cropland, forest, and pastureland. This bias results in substantial underestimation of extreme heat exposure in high-density metropolitan areas and overestimation in medium and low-density areas.

1. Introduction
Spatial projections of future population are important to studies of global change. They are critical elements that can be used to explore future emissions and land cover change scenarios (Seto et al., 2012). Spatial projections of future population are also essential to understand future changes in the exposure and vulnerability of human societies to the impacts induced by climate change, such as sea level rise (Kulp & Strauss, 2019), heat waves (Liu et al., 2017; Luo & Lau, 2018), droughts (Liu et al., 2018), and epidemics (Perkins et al., 2016). Therefore, the development of reliable and fine-resolution spatial projections of future population can help mitigate the impacts and risks caused by global change.

In terms of future human society development trends, the Intergovernmental Panel on Climate Change has developed the shared socioeconomic pathways (SSPs) to describe future socioeconomic developments based on five narratives, which consider possible challenges associated with adaptation to climate change and its
mitigation. A key element of the SSPs is the spatial projections of population distribution, which are highly beneficial to studies of climate change and its impacts.

However, there is only a limited number of gridded SSPs population scenarios. Jones and O’Neill (2016) developed a representative gridded SSPs population data set, which has been applied in several studies of climate change (Chen et al., 2019; Dottori et al., 2018; Liu, Liang, et al., 2017; Mora et al., 2017). However, this data set has two important limitations. First, it has a relatively coarse spatial resolution (0.125°), which is insufficient for investigating the impacts of climate change at regional or local levels. Second, the development of this data set does not use ancillary data, which may cause bias in the predictions of population distribution. As we show later, this data set substantially overestimates the population in populated areas, which may lead to biased analysis of climate change impacts.

Our study develops a new method for high-resolution gridded projections of SSPs population at 100-m resolution. We argue that the construction of high-resolution population grids is essential. On the one hand, high-resolution population grids provide more spatial details and can correct the spatial bias in existing projections. On the other hand, aggregations based on the high-resolution population grids can also provide more reliable inputs for large-scale models.

The method presented in this study has three important features: (1) The proposed method uses contemporary high-resolution population maps (100 m) as the basis for future projections, (2) it adopts sophisticated machine learning algorithms to develop models for the prediction of future built-up land expansion and grid-level population, and (3) to render the results consistent with the SSPs narratives, it uses the SSPs population scenarios as the quantitative constraints. We also apply the anticipated change rates in the built-up land acquired from the Land Use Harmonization Version 2f (LUH v2f) data set.

We applied the proposed method to mainland China and produced a 100-m SSPs gridded population data set, spanning from 2020 to 2050 at 5-year intervals. We further compared our results with the widely used gridded population projections developed by Jones and O’Neill (2016). We also demonstrated an application of our projections with an analysis of exposure to extreme heat events under the representative concentration pathways (RCPs) 4.5 and 8.5. To our knowledge, this is the first 100-m resolution SSPs gridded population data set for a large territory. Nevertheless, as the data we used are open access, the proposed method can be conveniently applied in other regional or global studies.

### 2. Background

There are presently several large-scale gridded population data sets with various spatial resolutions and time spans (Table 1). The Gridded Population of the World data sets, Versions 3 and 4, were developed by allocating populations from census units to grid cells based on an aerial-weighting method (Doxsey-Whitfield et al., 2015). The Global Rural-Urban Mapping Project, Version 1, further uses several sets of ancillary spatial data, such as urban extent and land cover data derived from remotely sensed images, to improve the spatial allocation of populations (Balk et al., 2006). The LandScan data sets also use a series of covariates and dasymetric modeling approaches to estimate the grid-level population (Dobson et al., 2000). All these three data sets consistently have 1-km spatial resolution. Another two recently developed data sets, i.e., the Global Human Settlement (GHS) Population Grid (Melchiotti et al., 2018) and the WorldPop data sets (Lloyd et al., 2017), have finer spatial resolution of 250 and 100 m, respectively. The GHS Population Grid uses the GPWv4 population estimates and the GHS Layer data to generate gridded population distribution.
The WorldPop uses a wide variety of ancillary spatial data and machine learning algorithms to develop global population maps at 100-m resolution, which are perhaps the finest (both spatially and temporally) gridded population data to date. Recent studies have also reported the use of new data sources, such as mobile phone data (Deville et al., 2014) and social media data (Ye et al., 2019), along with machine learning algorithms for large-scale population mapping with fine spatial resolution.

Most of the data products mentioned above focus on population mapping for historical and current situations, whereas spatially explicit projections consistent with the SSPs are less common. There are two available large-scale gridded SSPs population data sets (Table 1). One is the 0.5° Global Population Scenarios data developed by Murakami and Yamagata (2019), which is based on a downscaling approach that considers interactions among cities and an ensemble learning method to address the complex relationships between a number of covariates and population distribution. The other data set, i.e., the spatial population scenarios (0.125°), was developed by Jones and O’Neill (2016) at the National Center for Atmospheric Research (NCAR). They used a parameterized gravity-based model to predict the grid-level population in different countries, constrained by the SSPs population and urbanization projections at country level. Several other studies have developed methods for the regionalization of SSPs population projections Africa (Boke-Olén et al., 2017) (1-km resolution), the Mediterranean coastal zone (Reimann et al., 2018) (1-km resolution), and the United States (Wear & Prestemon, 2019) (county level), etc.

Overall, previous studies have provided important findings and foundations for the development of high-resolution SSPs population grids. First, previous studies have revealed the most important environmental factors relevant to population distribution, such as access to roads and major centers, built-up land distribution, and topographic conditions (Lloyd et al., 2017). Additionally, recent studies have reported on the outstanding performance of machine learning algorithms for the spatial prediction of population distribution (Stevens et al., 2015; Ye et al., 2019), owing to the flexibility and significant capabilities of machine learning algorithms to manage complex nonlinear data relationships (Reichstein et al., 2019). Second, a wealth of high-resolution population maps have been made available. A common feature of these data is that they are multitemporal and, hence, provide a basis to understand how population changes spatially. Moreover, these high-resolution population maps can be used immediately to represent the initial conditions for the projections of SSPs population (Boke-Olén et al., 2017). Third, although current spatial data sets of SSPs population projections are relatively few and coarse, they still can be used to facilitate the regionalization of SSPs population scenarios.

3.Materials and Methods

3.1.Data Sources

We obtained the 2010 and 2015 China population maps from the WorldPop Asia data set (Stevens et al., 2015). The population maps provide the estimated number of persons per pixel with a spatial resolution of 3 arc sec, corresponding to approximately 100 m at the equator. To facilitate the subsequent analysis, we used a natural break to divide the 2015 population map into six strata according to the estimated persons per pixel, including one background stratum with zero-valued pixels, while the other five strata have positive pixel values. The specific intervals are listed in the footnote of Table 2.

We obtained the 2010 and 2015 China land use maps from the Data Center for Resources and Environmental Sciences of the Chinese Academy of Sciences (Ning et al., 2018) (Table 2). The land use maps have six land use categories, namely, cropland, forest, grassland, water, built-up land, and unused land. The spatial resolution is 30 m. We extracted the built-up land from these maps to represent the spatial distribution of human settlements. They were resampled using the nearest approach to have a spatial resolution consistent with that of the population maps.

Additionally, we selected the spatial factors that influence changes in the population and built-up land according to previous studies (Chen et al., 2016; Feng et al., 2020; Gaughan et al., 2016). These factors include the distance to city centers, distance to county centers, distance to town centers, distance to highways, and distance to major roads. To generate these data, we collected the geographic locations of cities, counties, towns, and transportation networks (with highways and major roads) from the Data Center for Resources and Environmental Sciences of the Chinese Academy of Sciences (Table 2). Five distance
variables were produced corresponding to these spatial elements with the same spatial resolution as that of the population maps. These variables were held constant in the projections of built-up land and population distribution.

To predict future growth of built-up land area, a feasible approach is to use models developed based on region-specific indicators such as per capita built-up land area and per capita gross domestic product. By using this approach, a recent study has reported the country-level built-up land areas under the SSPs (Li et al., 2019). In our study, we used an alternative approach to estimate future built-up land area at provincial level. We derived the provincial change rates in built-up land under each SSP from the LUH v2f data set (0.25°) by aggregating the gridded data into provincial level. Then they were used to adjust the predictions of built-up land expansion so that the predictions could align with the SSPs narratives.

As we attempted to produce population grids consistent with the SSPs, we collected the estimated total population of China in different SSPs from 2015 to 2050 (Samir & Lutz, 2017) and used them as quantitative constraints. These data can be accessed from the SSP Database, which is managed by the International Institute for Applied Systems Analysis group. To compare our spatial projections of the population with existing counterparts, we also collected the data from the NCAR Spatial Population Scenarios with a spatial resolution of 0.125°. We refer to this data as the NCAR population grids in the following sections.

To analyze future exposure to extreme heat events, we collected data for the projections of daily maximum temperature ($T_{\text{max}}$) from 2020–2050 from the National Aeronautics and Space Administration Earth Exchange Global Daily Downscaled Projections (NEX-GDDP). This data set was downscaled from 21 Coupled Model Intercomparison Project 5 Global Climate Models using Bias Correction Spatial Disaggregation, with a spatial resolution of $0.25^\circ \times 0.25^\circ$. For each grid in any given year, an extreme heat day is defined as the day when its daily $T_{\text{max}} \geq 35 \, ^\circ\text{C}$, as suggested by the China Meteorology Administration. Based on this definition, in each of the 21 models' $T_{\text{max}}$ projections, we searched for all extreme heat days and counted the yearly frequency for each grid. Finally, for each grid, the yearly frequency of extreme heat days was obtained by averaging the number of extreme heat days derived from each of the 21 models' projections at that grid in the same year.

### 3.2. Methodology

We developed a machine learning based framework to generate the SSPs population grids (Figure 1). This framework has two important components. First, population prediction is based on the complex spatial path dependencies captured by the machine learning algorithms. Spatial path dependencies refer to the

| Data/Maps             | Year       | Resolution | Sources                                                                 |
|-----------------------|------------|------------|-------------------------------------------------------------------------|
| China population maps |
| 2010                  | 100 m      | WorldPop (https://www.worldpop.org)                                         |
| 2015                  |            |            |                                                                         |
| China land use maps   |
| 2010                  | 30 m       | Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (http://www.resdc.cn) |
| 2015                  | —          |            |                                                                         |
| City centers          |
| County centers        |
| Town centers          |
| Highways              |
| Roads                 |
| Population scenarios  |
| 2015–2050             | —          | SSP Database, International Institute for Applied Systems Analysis (https://tntcat.iiasa.ac.at) |
| Built-up land scenarios |
| 2020–2050             | 0.25°      | Land Use Harmonization Version 2f (http://luh.umd.edu/)                     |
| NCAR Spatial Population Scenarios |
| 2020–2050             | 0.125°     | National Center for Atmospheric Research (http://www.cgd.ucar.edu/iam/modeling/spatial-population-scenarios.html) |
| Future daily maximum temperature projections |
| 2020–2050             | 0.25°      | NASA NEX-GDDP (https://cds.ncdc.nas.gov/nex-gdp/)                           |

*Six strata: Background (0), (0, 19], (19, 53], (53, 106], (106, 191], and > 191.
influences that the previous or initial conditions have on the evolution of processes (Brown et al., 2005). Spatial path dependencies are associated with both environmental variability and past behaviors of a system. We followed this logic and used the spatial distribution of the population at time \( T \), as well as a set of environmental factors (e.g., the distance to city centers and distance to county centers), to predict the spatial distribution of population at time \( T + 1 \) (see the upper half of Figure 1). Second, in our framework, the spatial prediction of population change is linked with the prediction of built-up land expansion (as depicted in the lower half of Figure 1). This is to represent the dynamic impacts that the spatial population changes with built-up land (as human settlements). To this end, we also used machine learning algorithms to generate probability maps to determine future built-up land expansion.

We first collected a sample of data using a stratified random sampling scheme (see section 3.2.1). We then used the sampled data to train and test the performance of several machine learning algorithms. We used the trained algorithms to generate the population maps (see section 3.2.2) and the probability maps of built-up land expansion (see section 3.2.3). Finally, we projected the distribution of future population consistent with the SSPs (see section 3.2.4). The following sections provide more details on the methods.

3.2.1. Sampling Scheme
We tessellated the territory of mainland China using 250-km blocks, which result in a total of 279 blocks. We then used the 2015 population map to calculate the block-level density and separated the blocks into four types, namely, high-density, medium-density, low-density, and sparsely populated. Although the first
three types have only 66 blocks, they occupied more than 90% of the total population of mainland China. We then collected sample points in six high-density blocks, six medium-density blocks, and six low-density blocks. In each block, we equally allocated 2,000 points to each of the six population strata (see section 3.1 or the footnote in Table 2) and performed random sampling to collect data samples. Furthermore, we equally split the 18 sample blocks into two sets, one set for training and the other for testing. Each set had three high-density blocks, three medium-density blocks, and three low-density blocks. Figure S1 in Supporting Information (SI) shows the spatial distribution of the training and testing blocks. The specific cities located within each block can be found in SI Figure S2. We also created an extra data set of sample points at local levels to validate our models. We selected 25 representative cities in China (10 of them located in the sample blocks) as validation sites (SI Figure S1), with 1,000 sample points per city.

We collected a separating set of sample points to train the machine learning algorithms to predict built-up land expansion. Sampling was conducted in the same sample blocks mentioned above. In each sample block, we randomly collected 4,000 points, where half of the points were the “changed to built-up” type and the remaining half were the “nonchanged” type. Similar to the population sample data, we used the sample points in the training blocks to train the machine learning algorithms, as well as using the sample points in the testing blocks to assess their performance.

### 3.2.2. Population Prediction

To predict the grid-level population, we selected several machine learning algorithms according to their prevalence and performance in recent studies. They are XGBoost (XGB) (Georganos et al., 2018), random forest (RF) (Belgiu & Drăguț, 2016), and neural network (NN) (Hu et al., 2019).

We trained them individually with the sample data (obtained in the training blocks) to predict the 2015 population at the grid level. The model inputs included a set of environmental factors, that is, distance to city centers, distance to county centers, distance to town centers, distance to highways, and distance to major roads and built-up land in 2015. Additionally, the neighborhood population in 2010 was also used as model input to represent the dependencies of the 2015 population on the previous population conditions in 2010. Three neighborhood sizes were used based on previous studies (Chen et al., 2014; Liu et al., 2017), which were 3 × 3, 5 × 5, and 7 × 7.

We tested all trained machine learning models using the sample data collected from the testing blocks. We used the root-mean-square errors (RMSEs) and percentage term (%RMSE) to evaluate the absolute and relative prediction errors for each trained model:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i} (y_{i,\text{pred}} - y_{i,\text{obs}})^2}
\]  

and

\[
%RMSE = 100\% \times \sqrt{\frac{1}{n} \sum_{i} \left(\frac{y_{i,\text{pred}} - y_{i,\text{obs}}}{\bar{y}_{\text{obs}}}\right)^2}
\]

where \(y_{i,\text{pred}}\) and \(y_{i,\text{obs}}\) are the predicted and observed population at the \(i\)th grid, respectively, and \(\bar{y}_{\text{obs}}\) is the mean of the observed population.

The trained population prediction model can then be applied in the spatial projections of future population distribution, by recursively using the outputs of the population prediction at Time \(T_1\) to generate the model inputs for predictions of the population at time \(T_2\). For instance, we used the predicted 2020 population map to generate the inputs of the neighborhood population for the prediction of the 2025 population map, further using the predicted 2025 population map to generate the inputs for the prediction of the 2030 map, and so on. Additionally, the predicted built-up land distribution was used as input for population prediction (see the next section for more details). Finally, the resulting population predictions were adjusted according to the total population derived from the International Institute for Applied Systems Analysis SSPs database:
where $y_{i,adj}^T$ is the adjusted population at the $i$th grid at time $T$, $y_{i,pred}^T$ is the (immediate) population prediction based on the trained model at the $i$th grid at time $T$, and $Y_{cen}^T$ is the total population in a certain scenario at time $T$.

### 3.2.3. Built-Up Land Expansion Prediction

We individually trained the XGB, RF, and NN algorithms to predict the probability of built-up land expansion at the grid level. We used the variables of distance to city centers, distance to county centers, distance to town centers, distance to highways, distance to major roads, distance to existing built-up land, and slope as the model inputs. Here, the variable of distance to existing built-up land refers to the distance to the built-up land at time $T_1$ if the model is to predict the built-up land expansion probability at Time $T_2$. We tested the performances of the XGB, RF, and NN algorithms using the testing sample points collected from the testing blocks.

We validated the predicted probability maps of built-up land expansion using the receiver operating characteristic curve. This approach plots the true positive rate against the false-positive rate at different values of the predicted probability. The resulting area bounded by the curve and the $x$ and $y$ axes is referred to as the “area-under-curve” (AUC). If the AUC value is greater than 0.5, then this model performs better than a random classification model. Higher AUC values also indicate better model performance.

Based on the predicted probability maps, we projected the future built-up land expansion using an approach similar to that used by previous studies focusing on large-scale built-up land expansion projections (Güneralp et al., 2015; Seto et al., 2012), which simply allocates the expected amount of new built-up land to locations with the greatest predicted probability values. For instance, if the built-up land is expected to increase by $N$ grid cells from $T_1$ to $T_2$, then the $N$ nonbuilt grid cells with the highest probability values are selected as the locations for the new built-up land. We note that there are alternative methods to predict built-up land expansion, such as models based on cellular automata (Santé et al., 2010). However, these models are generally stochastic, implying that model outcomes vary among different runs even though the initial model configurations held unchanged. Therefore, using these models to project future built-up land distribution requires an additional procedure to manage the stochastic processes. In other words, we must determine a single run out of numerous runs that can be used as the final result, or merge the results of multiple runs into one single result by some means. The simple allocation approach avoids these issues.

To validate this approach, we compared the predicted and actual built-up land expansion for 2015 using a modified Kappa statistic, $K_{simulation}$ which adjusts the conventional Kappa to better assess the performance of land change models (van Vliet et al., 2011). This statistic explicitly considers the distribution of land use class transitions when estimating expected agreements. Here, $K_{simulation}$ can be expressed with the following equations:

$$K_{simulation} = \frac{P_o - P_{e(\text{Transition})}}{1 - P_{e(\text{Transition})}}, \quad (4)$$

$$P_o = \sum_i p(a = i \mid s = i), \quad \text{and}$$

$$P_{e(\text{Transition})} = \sum_j p(o = j) \cdot \sum_i p(a = i \mid o = j) \cdot p(s = i \mid o = j), \quad (6)$$

where $P_o$ is the observed fraction of agreement, $P_{e(\text{Transition})}$ is the expected fraction of agreement, given the sizes of the land use transitions, $p(a = i \mid s = i)$ is the fraction of grid cells that have land use type $i$ in the actual map at time $T$ and are predicted as land use type $i$ in the predicted map at time $T$, $p(o = j)$ is the fraction of land use type $j$ in the initial map at time $T - 1$, $p(a = i \mid o = j)$ is the fraction of grid cells that have land use type $i$ in the map at time $T$ while having land use type $j$ in the initial map at time $T - 1$, and, similarly, $p(s = i \mid o = j)$ is the fraction of grid cells that have land use type $i$ in the predicted map at time $T$ while having land use type $j$ in the initial map at time $T - 1$.
3.2.4. Scenario Projections

We projected China’s future population according to the assumptions and narratives of the SSPs. Specifically, SSP1 assumes a more sustainable development path in the future. This scenario also assumes increased investments in education and health that result in a relatively low global population. SSP2 is the “middle of the road” scenario that features medium population growth, urbanization, and development. SSP3 is characterized by regional rivalries, which lead to high population growth in developing countries and low population growth in industrialized countries. SSP4 features high inequalities within and between countries, rapid urbanization, and high fertility and population growth rates. SSP5 describes a future world with substantial technological progress, high education, and low mortality across all countries. More detailed demographic descriptions for each SSP scenario can be found in Jones and O’Neill (2016) and Samir and Lutz (2017).

We used 2015 as the base year and predicted future population distribution at 5-year intervals until 2050. Taking SSP1 as an example, we first produced the 2020 built-up land expansion probability map and then predicted the 2020 built-up land distribution by allocating the new built-up land to grid cells with the greatest probability values. Afterward, we used the predicted 2020 built-up land map, along with other inputs, such as the 2015 neighborhood population and other distance variables, to predict the population map for 2020. Similar procedures were implemented for the next 5-year period, as well as for the subsequent periods and other SSP scenarios.

For each 5-year period, the amount of built-up land change was determined according to the LUH v2f data set. This data set provides the annual gridded fractions of built-up land from 2015 to 2100 for the five SSPs, with a coarse resolution of 0.25°s. Based on this data set, we estimated the provincial built-up land change rates in different scenarios at 5-year intervals using the following equation:

$$r_i^{T+5} = \frac{\sum_j A_j f_j^{T+5}}{\sum_j A_j f_j^T} - 1,$$

where \(r_i^{T+5}\) is the rate of change in built-up land for province \(i\) from year \(T\) to year \(T + 5\), \(A_j\) is the area of grid \(j\) within province \(i\), and \(f_j^T\) and \(f_j^{T+5}\) are the fractions of built-up land in the \(j\)th grid of province \(i\) for years \(T\) and \(T + 5\), respectively. We then applied these provincial change rates to the 2015 built-up land data (Table 2) to obtain the specific amounts of built-up land change for each province:

$$a_i^{T+5} = a_i^T \times r_i^{T+5},$$

where \(a_i^{T+5}\) refers to the amount of built-up land change in province \(i\) from year \(T\) to year \(T + 5\) and \(a_i^T\) is the built-up land area in province \(i\) in year \(T\). Here, we estimated each province’s specific amount of built-up land change from 2015 to 2050 under different SSP scenarios. The SSPs population maps can then be produced by recursively feeding the population prediction model with the results of the predicted built-up land expansion.

3.2.5. Exposure to Heat Extreme

Following Jones et al. (2018) and Liu, Liang, et al. (2017), we calculated the grid-level exposure to extreme heat events by multiplying the yearly frequency of the extreme heat day and total population (unit: person per day). The prefectural yearly exposure to heat extreme was then calculated by summing the exposure of all grid cells within a prefectural unit:

$$e_p^T = \sum_{i,j \in P} e_i^T = \sum_{i,j \in P} y_i^T \times h_i^T,$$

where \(e_p^T\) is the prefectural yearly exposure to extreme heat for prefectural unit \(P\) at year \(T\) and \(e_i^T\) is the grid-level exposure to heat extreme for grid \(i\) at year \(T\), which is calculated as the product of the population \(y_i^T\) and yearly frequency of extreme heat day \(h_i^T\) in this grid at year \(T\). The population \(y_i^T\) can be obtained by aggregating our gridded population into the same resolution as that of the future daily maximum temperature projections (0.25°). To compare the performance between our population grids and the NCAR
Table 3  
The Mean RMSE and %RMSE for the XGBoost, Random Forest, Neural Network, and Support Vector Regression Algorithms (With Standard Deviation Shown in Parentheses)

| Algorithm | Hyperparameter | RMSE (std.) | %RMSE (std.) |
|-----------|----------------|-------------|--------------|
| XGB       | Num. Est. = 500, Booster = “gbtree” | 6.80 (0.00) | 11.24 (0.01) |
| RF        | Num. Trees = 500, Max. Feat. = 3 | 6.93 (0.01) | 11.42 (0.01) |
| NN        | Hid. Lyr. = 1, Max. Iter. = 500, Minibatch = 200 | 6.91 (0.05) | 11.44 (0.11) |
| Ensemble  |                                   |             |              |

Note. XGB = XGBoost; RF = random forest; NN = neural network; Num. Est. = number of estimators; Num. Trees = number of trees; Max. Feat. = the maximum number of features; Hid. Lyr. = the number of hidden layers; Hid. Lyr. Size = the number of neurons in the hidden layer; Max. Iter. = the maximum number of iterations; Minibatch = standard deviation.

Table 4  
The %RMSE of the Predicted Population at the Validation Sites

| City          | XGB   | Guangzhou | Kunming | Shanghai | Shenzhen |
|---------------|-------|-----------|---------|----------|----------|
| XGB           | 8.45  | 12.63     | 14.08   | 9.04     | 14.50    |
| RF            | 8.13  | 12.19     | 12.17   | 8.38     | 13.18    |
| NN            | 8.57  | 12.52     | 14.05   | 9.07     | 17.46    |
| Ensemble      | 8.20  | 12.35     | 13.17   | 8.55     | 14.62    |
| Tianjin       | 9.43  | 10.31     | 12.44   | 9.92     | 7.68     |
| XGB           | 9.40  | 9.05      | 11.67   | 9.76     | 7.78     |
| RF            | 9.08  | 10.55     | 12.29   | 10.06    | 7.71     |
| NN            | 9.18  | 9.61      | 11.94   | 9.74     | 7.56     |
| Ensemble      |       |           |         |          |          |
| Beijing       | 10.80 | 13.63     | 13.37   | 14.94    | 24.94    |
| XGB           | 10.63 | 13.72     | 13.41   | 16.01    | 24.84    |
| RF            | 11.12 | 13.74     | 13.03   | 14.18    | 24.86    |
| NN            | 10.57 | 13.43     | 13.11   | 14.82    | 24.69    |
| Ensemble      |       |           |         |          |          |
| Hangzhou      | 12.57 | 11.97     | 14.25   | 11.34    | 21.39    |
| XGB           | 12.49 | 12.33     | 14.57   | 11.45    | 21.15    |
| RF            | 12.36 | 11.76     | 14.85   | 12.21    | 21.03    |
| NN            | 12.27 | 11.67     | 14.30   | 11.35    | 21.07    |
| Ensemble      |       |           |         |          |          |
| Nanjing       | 20.14 | 12.13     | 12.55   | 15.03    | 11.91    |
| XGB           | 19.51 | 12.09     | 13.03   | 15.09    | 12.02    |
| RF            | 19.80 | 11.87     | 12.66   | 15.23    | 12.50    |
| NN            | 19.65 | 11.85     | 12.41   | 14.92    | 11.77    |

Note. XGB = XGBoost; RF = random forest; NN = neural network; Ensemble = averaging the predictions of the XGBoost, random forest, and neural network.

The differences among these three algorithms were minor, they were all used to predict the grid-level population, and their results were then stacked to form a single output using an averaging approach.

We further evaluated the errors of population prediction using the sampled data of the 25 representative cities (see SI Figure S1 for their geographic locations). Table 4 summarizes these results, while SI Figure S3 shows the scatter plots of population grids, we calculated the differences in the prefectural yearly exposure to extreme heat using these two data sets:

$$e_T^P = e_T^P - e_T^{P, NCAR}$$ and

$$e_T^C = \sum_{P} e_T^P,$$

where $e_T^{P, NCAR}$ is the estimated prefectural yearly exposure to heat extreme using the NCAR population grids and $e_T^C$ is the overall difference at the country level.

4. Implementation and Results

4.1. Population Prediction and Validation

We used the “scikit-learn” and “xgboost” packages in Python to implement model training. For each algorithm, we implemented ten rounds of training, testing their performance using RMSE and %RMSE metrics (Table 3). The results show that the XGB algorithm yielded the smallest mean prediction errors at 6.80 (11.24%). The RF and NN algorithms also had similar performances in terms of the mean prediction errors (6.93% or 11.42% vs. 6.91% or 11.44%). As the differences among these three algorithms were minor, they were all used to predict the grid-level population, and their results were then stacked to form a single output using an averaging approach.

We further evaluated the errors of population prediction using the sampled data of the 25 representative cities. The results also show that the ensemble of individual predictions can maintain a prediction error not worse than the most accurate individual predictions, as well as slightly improving upon the prediction accuracy in several cities, such as Zhengzhou, Baotou, Guiyang, and Hangzhou. Overall, the XGB, RF, and NN algorithms yielded mean prediction errors at the city level of 13.18%, 12.96%, and 13.30%, respectively, while their ensemble further reduced the mean prediction error to 12.91%. This result is consistent with empirical findings that the ensemble approach is an effective technique to improve the prediction accuracies of machine learning algorithms (Su et al., 2017).

Figure 2 illustrates the gridded population predictions generated by the XGB, RF, and NN algorithms, as well as their ensemble. Visually, the resulting patterns are consistent with the observed 2015 population patterns.

4.2. Built-Up Land Expansion Prediction and Validation

We trained the XGB, RF, and NN algorithms to predict built-up land change using the sample points collected from the training blocks and validation sets. The results show that the ensemble of individual predictions can maintain a prediction error not worse than the most accurate individual predictions, as well as slightly improving upon the prediction accuracy in several cities, such as Zhengzhou, Baotou, Guiyang, and Hangzhou. Overall, the XGB, RF, and NN algorithms yielded mean prediction errors at the city level of 13.18%, 12.96%, and 13.30%, respectively, while their ensemble further reduced the mean prediction error to 12.91%. This result is consistent with empirical findings that the ensemble approach is an effective technique to improve the prediction accuracies of machine learning algorithms (Su et al., 2017).

Figure 2 illustrates the gridded population predictions generated by the XGB, RF, and NN algorithms, as well as their ensemble. Visually, the resulting patterns are consistent with the observed 2015 population pattern. The ensemble results further smooth the subtle differences among the individual population patterns generated by the XGB, RF, and NN algorithms. These experimental results confirm that machine learning methods are adequate to use for spatial projections of future population distribution.
evaluated their performances using the sample points collected from the testing blocks. Table 5 summarizes the training and testing results. The performances of these algorithms were similar, with the overall prediction accuracies of between 89% and 90%. By applying these algorithms, the probability maps of the built-up land expansion were predicted and ensembled through averaging. The ensembled probability map was evaluated using the receiver operating characteristic curve approach. SI Figure S4 shows the resulting curves for several representative provincial units. The full results

Figure 2. A comparison of the spatial patterns for the population predicted by the XGBoost, random forest, and neural network algorithms, as well as their ensemble, in Guangzhou, Shenzhen, Chengdu, Wuhan, Shanghai, and Beijing.
for all provincial units can be found in SI Table S1. The AUC values ranged from 0.62 to 0.91 at the provincial level, with an average of 0.8, demonstrating the relatively good performance of the ensembled probability map.

Based on the ensembled probability map, we allocated the new built-up land to grid cells with the greatest probability values. SI Figure S5 shows the agreements/disagreements between the allocated and observed built-up land expansion from 2010 to 2015 for several representative areas. The applied allocation method yielded $K_{simulation}$ values between 0.40 and 0.59, with an average of 0.54 (SI Table S1 summarizes the full results for all provinces). According to the guidelines suggested by Viera and Garrett (2005), the resulting values indicate moderate agreement (0.40–0.60) between the predicted and observed built-up land expansion. Therefore, the proposed method is adequate to use for the prediction of future built-up land expansion under the SSP scenarios.

### 4.3 SSPs Gridded Population and Exposure to Heat Extreme

We determined the aerial changes of built-up land for the five SSPs according to the LUH v2f data set (see section 3.2.4). SI Tables S2–S6 summarize the provincial results, while SI Figure S6a shows the country-level results. Here, SSP5 yielded the largest estimated built-up land area of approximately 674,000 km$^2$. SSP1 had a smaller built-up land growth than SSP5 before 2025 but accelerated rapidly after 2025 to reach 671,000 km$^2$. In contrast, the trend of built-up land growth in SSP4 was nearly identical to that in SSP5 but slowed after 2030, becoming stable at 653,000 km$^2$. SSP2 featured a medium growth trend, reaching 657,000 km$^2$ in 2050. SSP3 had the smallest estimated area of built-up land, which was only 553,000 km$^2$.

Based on the estimated built-up land areas, we predicted the spatial distribution of built-up land in each scenario using the allocation method mentioned in sections 3.2.3 and 4.2. These scenario predictions were then used as model inputs to predict the grid-level population. The resulting population grids were further adjusted according to the total population obtained from the SSP database (Samir & Lutz, 2017) (supporting information Figure S6b). Figure 3 shows the 2050 scenario results for several representative cities while supporting information Figures S7–S11 show the full results. Our results demonstrate evident spatial structures of population distribution, where population density gradients are the highest in city centers and gradually decrease in suburban and rural areas. The spatial extent of high-density areas is the largest in SSP3, due to the highest total population and the smallest built-up land growth in this scenario. In contrast, SSP4 has the smallest total population and a large built-up land growth, which results in a higher proportion of medium or low-density areas. SSP2, however, has a similar built-up land growth to that in SSP4 but also a higher total population, leading to a relatively large proportion of high-density areas. The results for SSP1 and SSP5 are similar as they have nearly identical trends of population and built-up land, as well as a moderate proportion of high-density areas.

Based on the population predictions, we evaluated the rates of change in the provincial population from 2015 to 2050 and compared them with the rates of population change for the entire country. At the country level, all scenarios consistently yield an accelerated decline in China’s population from 2020 to 2050, with SSP4 yielding the fastest acceleration (i.e., from $-0.23\%$ for 2020–2025 to $-4.47\%$ for 2045–2050) (Figure 4a). However, the rates of change in the provincial population had different trends compared with those of the entire country. Figures 4b–4d demonstrate three representative trends. Inner Mongolia represents the first type of changes in the provincial population (Figure 4b). For all scenarios, the population of
Inner Mongolia is expected to increase before 2030 with diminishing growth rates and begin to decline after 2030 with accelerated rates. The second type, represented by Shaanxi, demonstrates consistent declining trends from 2015 to 2050 (Figure 4c). However, the population decrease in all scenarios first shrinks in the periods from 2015–2020 and 2020–2025, even reaching a rate of decline of less than 1% in SSP2 and SSP3, which afterward increased quickly from 2025 to 2050. The trends of population change in Guangxi represent the third type, in which the rates of change fluctuate from 2015 to 2050 (Figure 4d).

**Figure 3.** The 100-m SSPs population grids for Guangzhou, Shenzhen, Chengdu, Wuhan, Shanghai, and Beijing in 2050 under the five SSP scenarios.
We further estimated the trends of population exposure to extreme heat events by 2050 under the RCP4.5 and 8.5. Due to changes in population and climate, both scenarios suggest upward trends of exposure to heat extreme (Figure 5a and 5b) but with different growth rates. Under RCP4.5, the country-level exposure to extreme heat is expected to increase by 121–136% from 2015 to 2050 across the five SSPs. Under RCP8.5, however, due to the assumed greater greenhouse gas concentration levels, the temperature is expected to be substantially warmer than that in the RCP4.5. As a result, exposure to heat extreme in China will increase by between 164% and 191% across the five SSPs. For RCP4.5, increases in the exposure from 2015 to 2050 occur mainly in high-density urbanized areas, such as Beijing-Tianjin-Tangshan, the Yangtze River Delta, the Pearl River Delta, Changsha-Zhuzhou-Xiangtan, and Chengdu-Chongqing (Figure 5e). For RCP8.5, however, nearly the entire region of East China becomes highly exposed to extreme heat (Figure 5f).

4.4. Comparison With the NCAR Population Grids

We compared our results with the NCAR population grids (Jones & O'Neill, 2016). The comparison was performed with two methods, i.e., quantitative agreement evaluation at the provincial level and spatial agreement evaluation at the grid level. Before the comparison, our results were aggregated into the same resolution as that of the NCAR population grids. SI Figure S12 depicts the comparison of the provincial population in 2050 estimated by our population grids (y axis) and the NCAR population grids (x axis). The results yield similar $R^2$ values for SSP1, SSP4, and SSP5, which are approximately 0.64, and slightly lower $R^2$ values for SSP2 and SSP3, which are 0.60 and 0.56, respectively. These results indicate that our results and the NCAR grids have moderate quantitative agreement at the provincial level.

Evident spatial differences can be found between our population grids and the NCAR’s. Figure 6a shows a representative example in the North China Plain. Our results differentiate high- and low-density settlements, while the NCAR population grids have smoother spatial gradients of the changes in population density. Although these two projections consistently highlight high-density settlements, the NCAR population grids have a substantially larger spatial extent of low-density areas (<13 people/ha) than our population grids. This is mainly due to the overestimation of populated areas in the NCAR projections. As demonstrated in Figure 6b, grid cells with population density smaller than 13 people/ha contribute to over 96–97% of the NCAR population grids across the five SSP scenarios for 2050. Most of these grid cells are sparsely populated with density values lower than three persons/ha (86–88%). According to a recent estimation using strict criteria, at least 71% of China’s territory is uninhabited wilderness areas with no human activity (Ma & Long, 2019). Therefore, there are serious overestimations in the populated areas of the NCAR projections.
Then we overlaid the two projections separately with the land cover scenarios of cropland, forest, and pasture, which were obtained from the LUH v2f data set. The results confirm the presence of the overestimation problem in the NCAR projections, in which 30–43% of the estimated 2050 population are in grid cells with pure cropland, forest, or pasture cover types (Table 6). Our population grids, however, effectively mitigate the overestimation problem such that only 11–21% of the total estimated population (for 2050) is located in nonbuilt areas with pure cropland, forest, or pasture cover types.

We further compared the results of exposure analysis with those obtained using the NCAR population grids. For both the RCP4.5 and 8.5, the results of exposure analysis based on the NCAR population grids are substantially lower than those based on our population grids (Figures 5c and 5d). Under RCP4.5, the underestimation of exposure due to the use of the NCAR population grids increases over time from approximately 6% in 2020 to as high as 10% in 2050 (SSP5). The RCP8.5 features an opposite trend. The underestimation of exposure under this scenario tends to decline over time from approximately 8% in 2020 to 7% in 2050.

Figure 5. (a and b) Trends of exposure to extreme heat in China from 2015 to 2050 for the RCP4.5 and 8.5, respectively. (c and d) Country-level underestimation of exposure for the RCP4.5 and 8.5, respectively, due to spatial bias in the NCAR population grids. (e and f) Prefectural-level exposure increase from 2015 to 2050 for the SSP5-RCF4.5 and SSP3-RCP8.5, respectively. (g and h) Prefectural-level differences in the 2050 exposure between the estimations using the population grids developed in this study and the NCAR population grids for the SSP5-RCF4.5 and SSP3-RCP8.5, respectively.
Such an underestimation of exposure may be due to a spatial bias associated with populated areas in the NCAR population grids. As shown in Figure 6, the NCAR population grids overestimate the spatial extent of low-density areas. However, this bias may also cause the underestimation of high-density areas. Figures 5g and 5h depict the difference between the results of exposure analysis at the prefectural level using our population grids and the NCAR population grids, where positive values indicate underestimation and negative values indicate overestimation of exposure due to the use of the NCAR population grids. The results confirm our assumption of the influence of the biased populated areas in the NCAR population grids. Under both the RCP4.5 and 8.5 scenarios, the underestimation of exposure appears mainly in high-density metropolitan areas, such as Beijing-Tianjin-Tangshan, the Yangtze River Delta, and the Pearl River Delta. This is largely due to the underestimation of the population in these high-density areas (Figures 5g and 5h). Similarly, due to the overestimation of the population in low-density areas, the analysis tends to overestimate the exposure in these areas (blue shading in Figures 5g and 5h).
Table 6

|                      | NCAR population grids | Projections in this study |
|----------------------|-----------------------|---------------------------|
|                      | Cropland | Forest | Pasture | Total | Cropland | Forest | Pasture | Total |
| SSP1–1.9             | 7.18%    | 20.10% | 2.65%   | 29.92%| 2.06%    | 8.91%  | 1.36%   | 12.34%|
| SSP1–2.6             | 2.76%    | 21.80% | 2.65%   | 27.21%| 0.99%    | 10.15% | 1.36%   | 12.51%|
| SSP2–4.5             | 22.98%   | 15.82% | 4.10%   | 42.90%| 12.88%   | 5.03%  | 3.02%   | 20.94%|
| SSP3–7.0             | 0.06%    | 26.74% | 2.80%   | 29.60%| 0.00%    | 14.60% | 0.75%   | 15.35%|
| SSP4–3.4             | 19.25%   | 13.66% | 3.09%   | 36.00%| 9.15%    | 4.72%  | 1.45%   | 15.32%|
| SSP4–6.0             | 8.63%    | 16.87% | 3.44%   | 28.94%| 2.97%    | 6.90%  | 1.93%   | 11.79%|
| SSP5–3.4             | 11.84%   | 14.82% | 3.51%   | 30.16%| 4.03%    | 5.33%  | 1.90%   | 11.27%|
| SSP5–8.5             | 12.68%   | 14.59% | 3.52%   | 30.79%| 4.54%    | 5.45%  | 1.90%   | 11.89%|

Note. The projections in this study were resampled to have a spatial resolution consistent with that of the NCAR population grids (i.e., approximately 15 km). The cropland, forest, and pasture cover types in each scenario were acquired from the LUH v2f data set. Grid cells classified as cropland, forest, or pasture cover types correspond to cover fractions greater than 70%.

5. Discussion and Conclusions

Despite the importance of high-resolution gridded population projections, the development of relevant data products is relatively poor. The coarse resolutions of existing gridded SSPs population projections have hindered their use in studies of measurement, adaption, and mitigation of climate change impacts. Here, we proposed a machine learning method for the production of high-resolution gridded population projections consistent with the SSPs, focusing on future changes in China’s population from 2015 to 2050. According to the experiments on predicting historical population change, the proposed method yields an overall testing accuracy of less than 13% at the city level. By applying this method with the quantitative constraints of the SSPs, we produced 100-m gridded population projections for 2015–2050 at 5-year intervals. To our knowledge, this is the first 100-m data set of SSPs gridded population for a large territory.

At the grid level, our projections suggest different spatial structures of population distribution across the SSPs (Figure 3). Specifically, SSP3 had the largest spatial extent of high-density areas, mainly due to the highest total population and smallest expected built-up land area. In contrast, SSP4 features the smallest total population and a relatively large area of built-up land, leading to large extents of medium or low-density areas. At the provincial level, we identified three tendencies of population change: (1) increasing at first and declining thereafter (e.g., Inner Mongolia), (2) accelerated decline (e.g., Shaanxi), and (3) relatively stable rates of decline (e.g., Guangxi) (Figure 4). Based on the results of the gridded population projections, we analyzed future trends of exposure to extreme heat events under the RCP4.5 and 8.5. The results for both scenarios consistently exhibit upward trends of increasing exposure by 121–136% for RCP4.5 and 164–191% for RCP8.5 from 2015 to 2050 (Figure 5a and 5b). The increase in exposure will mainly occur in high-density urbanized areas in east China (Figures 5e and 5f).

Our projections demonstrate a decreasing trend of China population in the next three decades at both national and provincial levels. This trend is in line with that revealed in the recently released United Nations reports (2019), which suggested that population in several developed countries (e.g., Japan, Korea, and Germany) as well as China will start to decline in the near future. On the other hand, built-up land expansion in China is expected to continue with the estimated growth rates of 24–60% by 2050 (SI Figure S6). These trends will result in smaller population size in rural areas and higher population density in urban areas. According to the natural laws of ecology, density is an important limiting factor that constrains the size and growth of population (Cain et al., 2014), mainly because the pressure of competition for food and space, risks of disease infection and so on will increase if density increases (Snider & Brimlow, 2013). Empirical studies also confirmed the negative impacts that density has on human population growth (Lutz & Qiang, 2002). Fertility preferences were found to decline with population density, which might be associated with the increased costs of rearing children (Lutz et al., 2006). Therefore, the increased population density in China may enhance the decreasing trend of total population.

The comparison between the population grids developed in this study and the NCAR population grids revealed an evident spatial bias in the NCAR population grids. The spatial bias manifests as overestimations in the spatial extent of low-density areas and underestimations of population in high-density areas.
Moreover, taking 2050 as an example, approximately 30–43% of the estimated population in the NCAR projections were in uninhabited areas with pure cropland, forest, or pasture cover types (Table 6), which is unreasonable from a population distribution perspective. This bias leads to the underestimation of increased exposure to extreme heat events at the country level by 6–10% (Figures 5c and 5d). Spatially, high-density metropolitan areas are where exposure underestimations are observed while medium and low-density areas have overestimated exposure (Figures 5g and 5h). The results of the comparison demonstrate that the population grids presented in this study and their aggregation can avoid these problems.

Our study has several limitations. First, to obtain projections consistent with the SSPs, we adjusted our results according to the SSP database at the national level rather than explicitly modeling the population dynamics (e.g., fertility, mortality, and urbanization level change, among others). This issue should be settled in future research by incorporating demographic elements to project population scenarios at the provincial or subnational level. Second, we did not account for the impacts that climate change has on population distribution. Nevertheless, our projections can be used as a reference for comparison with population patterns under scenarios representing various climate change events (e.g., droughts, floods, and rising sea level), thereby facilitating policy suggestions for the reallocation of peoples vulnerable to those events. Third, the uncertainty in the data may influence subsequent projections under the SSPs. The calibration period is relatively short using the 2010 and 2015 population maps. We have alleviated the impacts of this problem by collecting sample points in diverse regions for model calibration and testing. In addition, the used built-up land data are binary maps that represent built and nonbuilt areas. However, population distribution may interact differently with the heterogeneous uses of built-up land (e.g., residential, industrial, and commercial). These interactions were not explicitly addressed in our study, but the five distance variables used in our study can somehow represent the heterogeneity of built-up land and its influence on population distribution. The uncertainty and accumulative error of built-up land projections over time also can affect the results of population projections. Therefore, in future studies, it is of great significance to develop more sophisticated methods that can utilize time series maps of population and land use (if available) such that we can more accurately capture the spatial-temporal signatures of population change.

Acknowledgments
This research was supported by the National Key R&D Program of China (2019YFA0607201 and 2017YFA0604042), the National Natural Science Foundation of China (Grants 41871306 and 41871029), and the Key National Natural Science Foundation of China (Grant 41531176). The 100-nm population grids that support the findings of this study are available at figshare website (https://doi.org/10.6084/m9.figshare.11634372). The data of future daily maximum temperature projections are acquired from NASA NEX-GDDP (https://cds.nccs.nasa.gov/nex-gddp/). The historical population maps are available in the WorldPop project (https://www.worldpop.org/). The China land use maps and spatial driving factors are available in Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (http://www.resdc.cn). The shared socioeconomic pathways (SSPs) population scenarios and their spatial data are provided in the websites of SSP database (https://tntcat.iiasa.ac.at) and National Center for Atmospheric Research (http://www.cgd.ucar.edu/iam/modeling/spatial-population-scenarios.html), respectively.

References
Balk, D. L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I., & Nelson, A. (2006). Determining global population distribution: methods, applications and data. Advances in Parasitology, 62, 119–156.
Belgium, M., & Dräger, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.
Boke-Olén, N., Aldí, A. M., Hall, O., & Lehtnen, V. (2017). High-resolution African population projections from radiative forcing and socio-economic models, 2000 to 2100. Scientific Data, 4, 160130.
Brown, D. G., Page, S., Riolo, R., Zellner, M., & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. International Journal of Geographical Information Science, 19(2), 153–174.
Cain, M., Bowman, W., & Hacker, S. (2014). Ecology (3rd ed.). Sunderland, Massachusetts, U.S.A.: Sinauer Associates, Inc.
Chen, Y., Li, X., Liu, X., & Ai, B. (2016). Capturing the varying effects of driving forces over time for the simulation of urban growth by using a spatially explicit cellular automaton with a patch-based simulation strategy. International Journal of Geographical Information Science, 28(2), 234–255.
Chen, Y., Li, X., Liu, X., Ai, B., & Li, S. (2016). Capturing the varying effects of driving forces over time for the simulation of urban growth by using survival analysis and cellular automata. Landscape and Urban Planning, 152, 59–71.
Chen, Y., Li, X., Liu, X., Zhang, Y., & Huang, M. (2019). Tele-connecting China's future urban growth to impacts on ecosystem services under the shared socioeconomic pathways. Science of the Total Environment, 652, 765–779.
Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., et al. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888–15893.
Dobson, J. E., Bright, E. A., Coleman, P. B., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. Photogrammetric engineering and remote sensing, 66(7), 849–857.
Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Aliferi, L., Hirabayashi, Y., et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming. Nature Climate Change, 8(9), 781.
Doxsey-Whitley, E., MacManus, K., Adamo, S. B., Pistolesi, I., Squires, J., Borkovska, O., & Baptista, S. R. (2015). Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. Papers in Applied Geography, 1(3), 226–234.
Feng, Y., Lei, Z., Tong, X., Gao, C., Chen, S., Wang, J., & Wang, S. (2020). Spatially-explicit modeling and intensity analysis of China’s land use change 2000–2050. Journal of Environmental Management, 263, 110407.
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. Scientific Data, 3, 160005.
Georganos, S., Grippa, T., Vanhuyse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E. (2018). Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. GIScience & Remote Sensing, 55(2), 221–242.
Güneralp, B., Güneralp, I., & Liu, Y. (2015). Changing global patterns of urban exposure to flood and drought hazards. Global Environmental Change, 31, 217–225.
Hu, L., He, S., Fan, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. Land Use Policy, 82, 657–673.

Jones, B., & O'Neill, B. (2016). Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. Environmental Research Letters, 11(8), 084003.

Jones, B., Tehbaldi, C., O'Neill, B. C., Oleson, K., & Gao, J. (2018). Avoiding population exposure to heat-related extremes: demographic change vs climate change. Climatic change, 146(3–4), 423–437.

Kulp, S. A., & Strauss, B. H. (2019). New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. Nature Communications, 10(1), 1–12.

Li, X., Zhou, Y., Eom, J., Yu, S., & Asrar, G. R. (2019). Increasing heat stress in urban areas of eastern China: Acceleration by urbanization. Geophysical Research Letters, 46(23), 13,060–13,069. https://doi.org/10.1029/2018GL080306

Lutz, W., & Qiang, R. (2002). Determinants of human population growth. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 357(1425), 1197–1210.

Lutz, W., Testa, M. R., & Penn, D. J. (2006). Population density is a key factor in declining human fertility. Population and Environment, 28(2), 69–81.

Ma, S., & Long, Y. (2019). Mapping potential wilderness in China with location-based services data. Applied spatial analysis and policy, 1–21.

Melchiorri, M., Florczyk, A., Freire, S., Schiavina, M., Pesaresi, M., & Kemper, T. (2018). Unveiling 25 years of planetary urbanization with machine learning algorithms and hedonic modeling to inform equitable housing policies. Applied spatial analysis and policy, 1(1), 1–12.

Mora, C., Douset, B., Caldwell, I. R., Powell, F. E., Geronimo, R. C., Bielecki, C. R., et al. (2017). Global risk of deadly heat. Nature Climate Change, 7(7), 501.

Murakami, D., & Yamagata, Y. (2019). Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling. Sustainability, 11(7), 2106.

Ning, J., Liu, J., Kuang, W., Xu, X., Zhang, S., Yan, C., et al. (2018). Spatiotemporal patterns and characteristics of land-use change in China during 2010–2015. Journal of Geographical Sciences, 28(5), 547–562.

Perkins, T. A., Siraj, A. S., Ruktanonchai, C. W., Kraemer, M. U., & Tatem, A. J. (2016). Model-based projections of Zika virus infections in childbearing women in the Americas. Nature Microbiology, 1(9), 16126.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204.

Reimann, L., Merkens, J.-L., & Valeidus, A. T. (2018). Regionalized shared socioeconomic pathways: Narratives and spatial population projections for the Mediterranean coastal zone. Regional Environmental Change, 18(1), 235–245.

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. Global Environmental Change, 42, 181–192.

Santé, I., Garcia, A., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landscape and Urban Planning, 96(2), 108–122.

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences, 109(40), 16083–16088.

Snider, S., & Brimlow, J. (2013). An introduction to population growth. Nature Education Knowledge, 4(4), 3.

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, S., Sun, Y., Lei, C., Weng, M., & Cai, Z. (2017). Reorienting paradoxical land use policies towards coherence: A self-adaptive ensemble learning geo-simulation of tea expansion under different scenarios in subtropical China. Land Use Policy, 67, 415–425.

United Nations. (2019). World Population Prospects 2019 (United Nations DESS/Population Division Ed.). New York, NY: United Nations. van Vliet, J., Bregt, A. K., & Hagen-Zanker, A. (2011). Revisiting Kappa to account for change in the accuracy assessment of land-use change models. Ecological Modeling, 222(8), 1367–1375.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Family Medicine, 37(5), 360–363.

Wear, D. N., & Prestemon, J. P. (2019). Spatiotemporal downscaling of global population and income scenarios for the United States. PloS one, 14(7), e0219242.

Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., et al. (2019). Improved population mapping for China using remotely sensed and point-of-interest data within a random forests model. Science of The Total Environment, 658, 936–946.