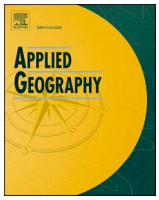
Applied Geography 182 (2025) 103708

Contents lists available at ScienceDirect

Applied Geography

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



High-resolution population density mapping in urban areas using a contextualized geographically weighted neural network (CGWNN) model

Ge Qiu [a,b], Yuchen Li [c,b], Kun Qin [a,b], Chen Li [a,b], Shuhan Yang [a,b], Chun Yin [a,b], Yaolin Liu [a,d,*], Shaoqing Dai [a,b,**], Peng Jia [a,e,f,b,***]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[b] International Institute of Spatial Lifecourse Health (ISLE), Wuhan University, Wuhan, China
[c] School of Geography, University of Leeds, Leeds, UK
[d] Duke Kunshan University, Kunshan, China
[e] School of Public Health, Wuhan University, Wuhan, China
[f] Renmin Hospital (First School of Clinical Medicine), Wuhan University, Wuhan, China

ARTICLE INFO

Handling Editor: Dr. Y.D. Wei

Keywords:
Population mapping
Urban functional zone
Land use
Land cover
Spatial heterogeneity
Artificial neural network
Contextual disparity

ABSTRACT

High-resolution gridded population data are crucial for various fields. Estimating heterogeneous urban populations presents challenges due to nonlinear relationships between influential factors and population density, which vary spatially across grids within different land-use parcels. This study developed a Contextualized Geographically Weighted Neural Network (CGWNN) model to estimate population density on 100 × 100m grid cells in Beijing, China, using multi-source data. This model integrated the artificial neural network with geographically weighted regression to account for nonlinear associations that are similar across proximate grids. By incorporating parcel-level land uses as variable weights, it also considered contextually varying associations across proximate grids located in different land-use parcels. Our CGWNN model achieved superior accuracy (R^2 = 0.85) compared to other models that ignored the aforementioned associations and widely used population datasets. The top three important variables were the distances to the nearest school, restaurant, and auto service, all negatively associated with population density. Additionally, the intensity of artificial light at night (ALAN) exhibited both positive and negative associations with population density in different regions, suggesting that the increased ALAN did not necessarily indicate higher population density in urban areas. Our modeling approach shows promise for accurate population estimation, which could be extended to larger areas, benefiting various fields.

1. Introduction

Urban populations are expected to reach five billion by 2030, accounting for more than half of the global population (Seto et al., 2012). Understanding their spatial distribution is fundamental for various fields, such as urban planning (Mallick et al., 2021), disaster response (Ding et al., 2025), and health management (Jia et al., 2019). Traditional population products (e.g., census data), usually depicted by choropleth maps, illustrate aggregated population counts over administrative units (Mennis, 2003). However, administrative boundaries are often in irregular shapes with a coarse spatial resolution, leading to difficulties in revealing spatial heterogeneities of population distribution within administrative units, and in combining with other high-resolution gridded products (e.g., gridded meteorological datasets) for greater use (Martin, 2011, pp. 655–665). This issue is particularly pronounced in urban areas, where patterns of population distribution are usually more complex and vary more substantially across geography than in rural areas (Jia et al., 2014). However, urban areas have usually been considered as one category (i.e., impervious surface) in land cover products (Chen et al., 2021), the most frequently used ancillary datasets in fine-scale population estimation (Lei et al., 2023), which prevent from capturing intra-urban variations of population density.

* Corresponding author. School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.
** Corresponding author. School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.
*** Corresponding author. School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.
E-mail addresses: yaolin610@yeah.net (Y. Liu), shaoqing.dai@outlook.com (S. Dai), jiapengff@hotmail.com (P. Jia).

https://doi.org/10.1016/j.apgeog.2025.103708
Received 9 December 2024; Received in revised form 24 June 2025; Accepted 26 June 2025
Available online 4 July 2025

High-resolution gridded population products have been a promising solution to address the aforementioned challenges, which could be generated by two major methods: **top-down** and **bottom-up** (Wardrop et al., 2018). Top-down methods usually use the weights, produced from population density-related ancillary data, to assist in redistributing census population counts onto finer-scale grids (Jia & Gaughan, 2016); bottom-up methods usually establish the grid-level associations between influential factors, produced from ancillary data, and population density in small areas (e.g., micro-census) and then extrapolate them to larger areas (Boo et al., 2022; Leasure et al., 2020). The latter has been considered the most feasible approach for population estimation, as the former relies more on the high-quality collection of census data, which are costly, labor-intensive, and time-consuming. Traditional regression (e.g., ordinary least squares regression) was initially used to estimate the associations between influential factors and population density, assuming that these associations were stable among all grids (Song et al., 2019). Nevertheless, such associations may vary spatially across grids, which have been more accurately estimated by geographically weighted regression (GWR) (Roni & Jia, 2020; Song et al., 2019), under the assumption that the associations in nearby grids tend to be more similar than those in distant grids (Fotheringham et al., 1998; Tobler, 1970). However, such spatial similarity may be disrupted by different attributes (e.g., land use) of the parcels in which the grids are located, leading to less similar associations in even nearby grids (Griffith, 2006; Qiu et al., 2020). For example, the association between nightlight intensity and population density in a residential and an industrial grid, although proximate geographically, could be substantially distinct (Song et al., 2019; Wang et al., 2018). Such contextual attributes should thus be considered for more accurate population estimation in urban areas.

Contextualized GWR (CGWR) extends traditional GWR to better capture the varying association across geography by incorporating contextual (parcel in this study) information in the GWR weights matrix, offering a potential solution to the aforementioned issue (Harris et al., 2013). However, when applied for population estimation, CGWR has encountered at least two technical difficulties. One is that CGWR measures contextual differences across parcels by the Euclidean distance, which is applicable for numerical and ordinal data rather than nominal data (Boriah et al., 2008). This makes CGWR inappropriate for population estimation, as contextual attributes of parcels may be nominal, such as land cover and land use categories. The other difficulty is that CGWR uses linear models to estimate the associations between influential factors and population density, which, in the real world, may be more complex, usually nonlinear (Qiu et al., 2020). Previous studies have employed machine learning methods, such as random forest (Chen et al., 2024) and artificial neural networks (ANN) (Cheng et al., 2021), to estimate nonlinear associations between influential factors and population density. Moreover, those machine learning methods have been combined with GWR to model spatially varying nonlinear associations in other research areas than population estimation (Feng et al., 2021; Wei et al., 2020). However, to the best knowledge of the authors, all existing studies of grid-level population estimation have overlooked the impacts of contextual attributes of parcels on the grid-level associations between influential factors and population density.

To fill the aforementioned gaps, this study developed a novel contextualized geographically weighted neural network (CGWNN) method to estimate urban population density in 100 × 100m grids, by considering nonlinear and contextually varying associations between influential factors and population density at the grid level. A dataset named essential urban land use categories in China (EULUC-China) (Gong et al., 2020), including detailed information of urban land use, was used to provide the contextual information of the parcel. To demonstrate the advantages of the CGWNN, the gridded population product produced from this study was compared with several well-known population products, including WorldPop, Global Human Settlement Population Grid (GHS-POP), Gridded Population of the World (GPW), and LandScan Global. The CGWNN method developed in

Table 1
Datasets used in this study.

Data	Format	Description	Year	Source
Census	Table	The 7th national population census data at the township level	2020	National Bureau of Statistics of China
Boundary maps	Polygon	Township-level administrative boundaries	2020	Chinese Academy of Surveying and Mapping
WorldPop	Raster	Gridded population product with a 100-m resolution	2020	University of Southampton, UK
GHS-POP	Raster	Gridded population product with a 100-m resolution	2020	Joint Research Center, European Commission
GPW	Raster	Gridded population product with a 1-km resolution	2020	Center for International Earth Science Information Network, Columbia University, USA
LandScan Global	Raster	Gridded population product with a 1-km resolution	2020	Oak Ridge National Laboratory, USA
NPP-VIIRS ALAN	Raster	Annual composite intensity of ALAN at a 500-m resolution	2020	Colorado School of Mines, USA
ASTER GDEM Version 3	Raster	Elevation dataset at a 30-m resolution	–	NASA, USA
MOD13Q1 NDVI	Raster	NDVI dataset with a 250-m resolution	2020	NASA, USA
Points of interest	Point	18 categories: residential community, restaurant, bank, school, hotel, hospital, toll station, transportation service (e.g., parking lots, and bus stations), tourist attraction, auto service, government building, public facility (e.g., public toilets, kiosks), sport and recreation, company, daily life service (e.g., post offices, barber shops), shopping service, passing facility (e.g., courtyard doors), and other addresses (e.g., village names, address numbers)	2020	AutoNavi Software Co., Ltd., China
EULUC-China	Polygon	5 themes including 12 categories of urban land use: residential, commercial (business office, commercial service), industrial, transportation (road, transportation stations, airport facilities), public management and service (administrative, educational, medical, sport and cultural, park and greenspace)	2018	Tsinghua University, China

ALAN, artificial light at night; ASTER GDEM, Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model; EULUC, essential urban land use categories; GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; MOD13Q1, Moderate-resolution Imaging Spectroradiometer Vegetation Indices; NASA, National

Aeronautics and Space Administration; NDVI, normalized difference vegetation index; NPP-VIIRS, Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite.

this study holds great promise not only for population estimation and the fields relying heavily on such fundamental population products, but also for many other research and application fields where similar methodological bottlenecks (e.g., incapable of considering non-linear associations and contextual information) have limited the modeling accuracy.

2. Methods

2.1. Datasets

The datasets used in this study covered the cities of Beijing and Shanghai, with Beijing serving as the study area for model development and Shanghai used to evaluate the generalizability of the proposed model. These datasets included census, remote sensing, social sensing, and land use data. The latest Census 2020 and administrative boundary data at the township level (equivalent to the Global Administrative Unit Layers at Level 4) were acquired from the National Bureau of Statistics of China and the Chinese Academy of Surveying and Mapping (Table 1). The four well-known gridded population products for the year 2020, including WorldPop (Gaughan et al., 2016), GHS-POP (Carioli et al., 2023), GPW (Center For International Earth Science Information Network-CIESIN-Columbia University, 2018), and LandScan Global (Sims et al., 2023), were used to compare with the population product produced in this study.

Various types of remote sensing data that could reflect the extent of human activities were used as ancillary data: the intensity of artificial light at night (ALAN) was obtained from the Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite (NPP-VIIRS) with a 500-m resolution (Elvidge et al., 2021); elevation and slope were derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) with a 30-m resolution (Jacobsen & Passini, 2010); the normalized difference vegetation index (NDVI) was obtained from the Moderate-resolution Imaging Spectroradiometer Vegetation Indices (MOD13Q1) data with a 250-m resolution (Didan, 2021). Points of interest (POIs), an important type of social sensing data representing the places with (x, y) coordinates on maps, were obtained from the Amap Service, which included 18 categories such as residential community, restaurant, bank, school, and hotel. Additionally, the EULUC-China was produced from dividing impervious surfaces into small parcels by road networks, which included 12 categories such as residential parcel, business office, commercial services, industrial parcel, and road (Gong et al., 2020).

2.2. Data preprocessing and variable selection

This study employed a bottom-up approach for population estimation, which generates population products in the entire region based on the associations between influential factors and population density established in sampling areas at the grid level. As the grid-level population numbers (dependent variable) are not available in census data, they were produced by disaggregating the township-level population numbers (the finest publicly available census data) onto grids using WorldPop population counts (the finest gridded population counts) as weights (Zhuang et al., 2021). Ten grids were randomly selected from each township, with the population counts in selected grids logarithmically transformed to reduce skewness (Figs. 1–2). An uncertainty propagation analysis was conducted to assess the impact of the uncertainties in the approach on the results, stemming primarily from estimation biases/errors in WorldPop grids.

All ancillary data (independent variables) except EULUC-China were standardized by converting grids with different spatial resolutions into 100 × 100m grids: ALAN and NDVI, with the coarser resolution than 100m, were resampled to 100m by a bilinear interpolation method, prior to which NDVI was calculated as the maximum values during July and August due to cloud coverage in single MOD13Q1 NDVI images (Piao et al., 2003); elevation and slope, with the finer resolution than 100m, were aggregated to 100m by calculating the mean values; the density of POIs of each category was calculated as the counts (per unit

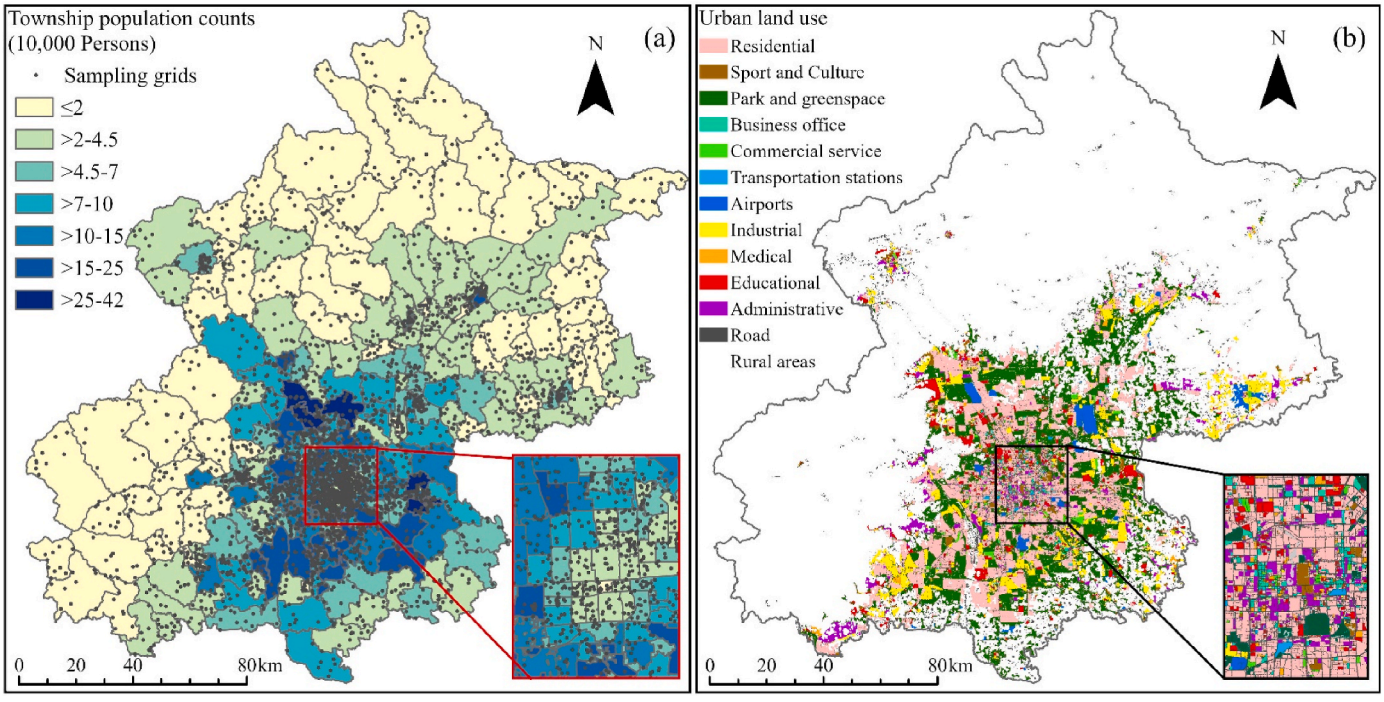


Fig. 1. Population and land use characteristics in Beijing, China: (a) township-level population counts and random sampling areas of grid-level population density; (b) land use map.

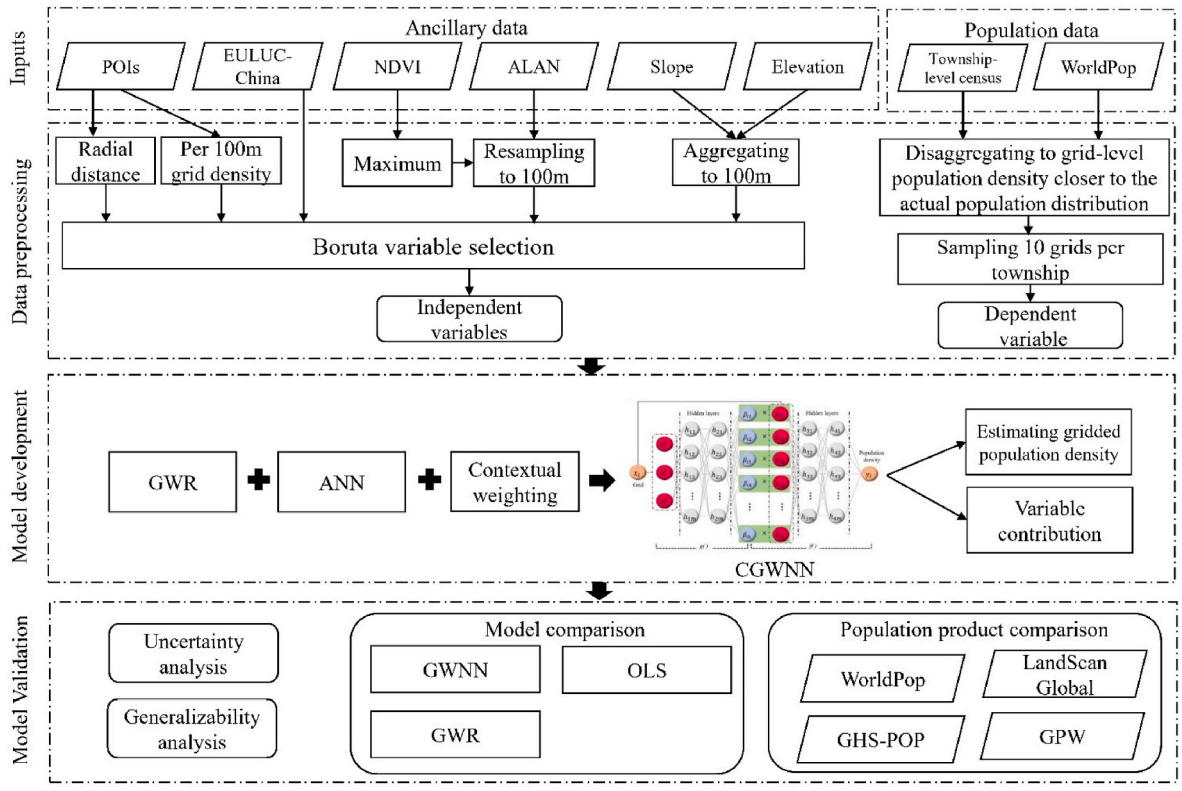


Fig. 2. The flowchart of general procedure in this study. ALAN, artificial light at night; ANN, artificial neural network; CGWNN, contextualized geographically weighted neural network; EULUC, essential urban land use categories; GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; GWNN, geographically weighted neural network; GWR, geographically weighted regression; NDVI, normalized difference vegetation index; POI, point of interest; OLS, ordinary least squares.

area) falling within the grid, while the distance to the nearest POI of each category was calculated as the distance from the center of the grid (equal to 0 if the nearest POI fell within the grid) (Bakillah et al., 2014).

The selection of the aforementioned variables for population estimation was made by a Boruta algorithm (Kursa & Rudnicki, 2010). First, an extended dataset was generated by creating shuffled copies of all independent variables and appending them to the original data. Then, Boruta applied a random forest regressor to the extended dataset, and evaluated the importance of each variable to population estimation by a mean decrease accuracy (MDA) metric, which was calculated by randomly permuting each variable and measuring the resulting drop in model accuracy. This step was repeated 99 times to generate stable distributions of importance scores. The variables with the importance scores significantly and consistently higher than those of the most important shuffled copies were considered necessary and used for population estimation.

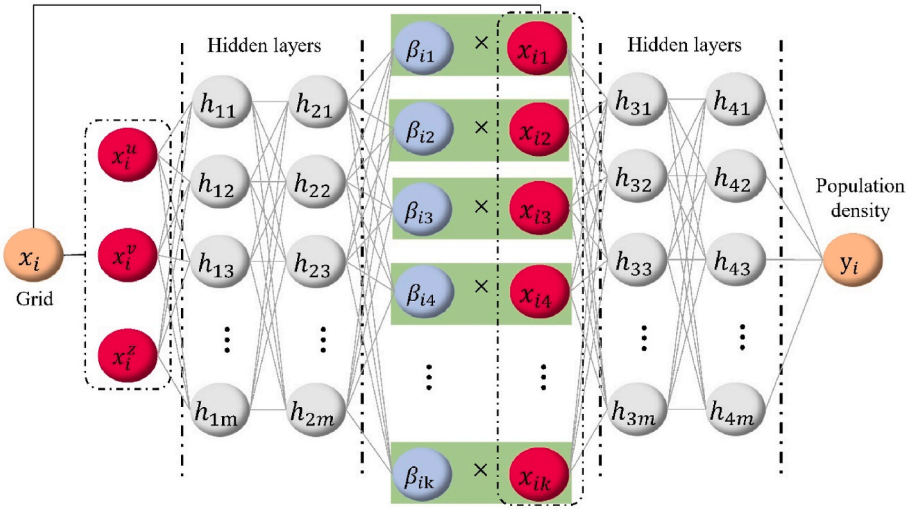


Fig. 3. The architecture of the proposed contextualized geographically weighted neural network. x_i^u, longitude of the grid i; x_i^v, latitude of the grid i; x_i^z, contextual attribute (i.e., land use) of the grid i; β_{ik}, weight of kth independent variable affected by x_i^u, x_i^v, and x_i^z; x_{ik}, kth independent variable.

2.3. CGWNN model development

This study developed a CGWNN to estimate associations between influential factors and population density at the grid level. The model integrated the strength of ANN in capturing nonlinear relationships and the advantage of GWR in capturing spatially varying relationships, enabling it to capture spatially varying nonlinear associations that tend to be more similar among proximate grids. In addition to geographical proximity, the model also incorporates contextual disparity by weighting influential factors with parcel-level land use, explicitly addressing abrupt variations in these associations across nearby parcels. The proposed CGWNN can be mathematically formulated as follows:

$$\begin{aligned} y_i &= \sigma\left(W_4 \cdot \sigma\left(W_3 \cdot \left[\beta_{i1} x_{i1}, \beta_{i2} x_{i2}, \cdots, \beta_{ik} x_{ik}\right]^T + b_3\right) + b_4\right) \\ \beta_i &= \sigma\left(W_2 \cdot \sigma\left(W_1 \cdot \left[u_i, v_i, z_i\right]^T + b_1\right) + b_2\right) \end{aligned} \quad (1)$$

where y_i denotes population density in grid i; $x_i = x_{i1}, x_{i2}, \cdots, x_{ik}$, denoting independent variables in grid i; $\beta_i = \beta_{i1}, \beta_{i2}, \cdots, \beta_{ik}$, denoting the weights of x_i, varies depending on spatial coordinates and context; W_3, W_4 and b_3, b_4 denote the weight matrices and bias terms of two hidden layers in a fully connected neural network, accounting for nonlinear transformations from weighted independent variables $\beta_i x_i$ to y_i; W_1, W_2 and b_1, b_2 denote the weight matrices and bias terms of two hidden layers in another neural network, which considers effects of spatial coordinates (u_i, v_i) and contextual variable z_i (i.e., land use) on the weight β_i; σ denotes the ReLU activation function applied at each hidden layer (Fig. 3).

To establish the CGWNN model, the sampling grids were divided into training, validation, and testing datasets at a proportion of 8:1:1. The model was trained using the Adam optimizer (learning rate = 0.001), with 20 neurons per hidden layer, a batch size of 600, and 200 epochs. Model parameters were updated using the backward propagation of errors calculated by the mean square error (MSE) loss function. SHapley Additive exPlanations (SHAP) were applied to interpret the established associations between population density and variables. Once trained, the model was applied to estimate population density in all 100-m grids covering Beijing. This approach helps avoid potential underestimations, as limiting estimates to residential parcels might overlook populations residing in areas misclassified as non-residential due to inherent uncertainties in land use data.

2.4. Model comparison and accuracy assessment

To demonstrate the advantages of CGWNN, several models were established for comparison, including geographically weighted neural network (GWNN), GWR, and ordinary least squares regression (OLS). The comparison between CGWNN and GWNN aimed to highlight the importance of integrating contextually varying associations between influential factors and population density in improving population estimation accuracy. Although various forms of GWNN have been proposed in previous studies (Du et al., 2020; Zhang et al., 2022), they often differ from the CGWNN in both network architecture and the way input features are treated or spatially processed, making direct comparisons difficult. To ensure a controlled ablation experiment, a GWNN model was implemented using the same architecture and input variables as CGWNN, but using land use as an independent variable rather than a contextual variable, allowing for a fair assessment of the added value of contextual weighting. The mathematical expression of this GWNN is shown as follows:

$$\begin{aligned} y_i &= \sigma\left(W_4 \cdot \sigma\left(W_3 \cdot \left[\beta_{i1} x_{i1}, \beta_{i2} x_{i2}, \cdots, \beta_{ik} x_{ik}, \beta_{i(k+1)} x_{iLU}\right]^T + b_3\right) + b_4\right) \\ \beta_i &= \sigma\left(W_2 \cdot \sigma\left(W_1 \cdot \left[u_i, v_i\right]^T + b_1\right) + b_2\right) \end{aligned} \quad (2)$$

where y_i denotes population density in grid i; $x_i = x_{i1}, x_{i2}, \cdots, x_{ik}, x_{iLU}$, denoting independent variables in grid i; $\beta_i = \beta_{i1}, \beta_{i2}, \cdots, \beta_{i(k+1)}$, denoting the weights of x_i, varies solely depending on spatial coordinates; W_3, W_4 and b_3, b_4 denote the weight matrices and bias terms of two hidden layers in a fully connected neural network, accounting for nonlinear transformations from weighted independent variables $\beta_i x_i$ to y_i; x_{iLU} denotes land uses and used as an independent variable; W_1, W_2 and b_1, b_2 denote the weight matrices and bias terms of two hidden layers in a fully connected neural network, considering only the effect of spatial coordinates (u_i, v_i) on weight β_i; σ denotes the ReLU activation function applied at each hidden layer. The same settings (i.e., activation function, optimizer, learning rate, neuron number in each hidden layer, batch size, epoch, and loss function) used for CGWNN were applied to GWNN to ensure comparability. ANN-based models, such as CGWNN and GWNN, randomly initialize parameters, which may cause slight variations in model performance. As such, to make a more robust comparison between CGWNN and GWNN, this study established 10 models for CGWNN and GWNN, respectively, and compared the average performance between the two sets of models. However, for more effective subsequent visualization and comparison of the resulting population products, the best among the 10 CGWNN models and the optimal among the 10 GWNN models were selected as the main models. To prevent potential overfitting, the difference between training loss and validation loss in the last epoch was monitored for the established CGWNN and GWNN models based on Cohen's d (Prechelt, 1998):

$$\text{Cohen's d} = \frac{\bar{x}_{training} - \bar{x}_{validation}}{\sqrt{SD^2_{training} + SD^2_{validation}/2}} \quad (3)$$

where $\bar{x}_{training}$ and $\bar{x}_{validation}$ denote the mean training loss and mean validation loss, respectively, for the final epoch of the 10 CGWNN or GWNN models; $SD_{training}$ and $SD_{validation}$ denote the standard deviations of the training and validation losses for the final epoch of the CGWNN or GWNN models. Cohen's d values smaller than 0.2 indicate a small and negligible difference. Moreover, GWNN was compared with GWR model to emphasize the expected nonlinear associations between influential factors and population density; two GWR models were compared to evaluate the importance of incorporating land use as an independent variable in population estimation, one of which included land use (GWR-LU) and the other did not; OLS and GWR models were compared to illustrate the existence of spatially varying associations between influential factors and population density.

The OLS, GWR, GWR-LU, GWNN, and CGWNN models were compared at both grid and township levels. The grid-level accuracy assessment was conducted to evaluate the effectiveness of the different models. The aforementioned testing dataset, accounting for 10 % of the sampling grids, was used to assess the performance of each model. Three commonly used metrics — root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2) — were employed. Moreover, to evaluate the relative accuracy of population products generated by different models, each product was aggregated at the township level and compared with census population counts. WorldPop, GHS-POP, GPW, and LandScan Global were also compared with these products at the township level. To enhance the comprehensiveness of the accuracy assessment, mean absolute percentage error (MAPE) was additionally included as a supplementary metric for township-level accuracy assessment, offering insights from the perspective of relative error. However, for grid-level assessment, MAPE was not applied due to the possibility of zero population in certain grids, which would render MAPE undefined:

$$\text{MAE} = \frac{1}{N}\sum \left| P_i^{observed} - P_i^{estimated} \right| \quad (4)$$

$$\text{RMSE} = \sqrt{\sum \left(P_i^{observed} - P_i^{estimated} \right)^2 / N} \quad (5)$$

$$\text{MAPE} = \frac{1}{N}\sum \left| \frac{P_i^{observed} - P_i^{estimated}}{P_i^{observed}} \right| \quad (6)$$

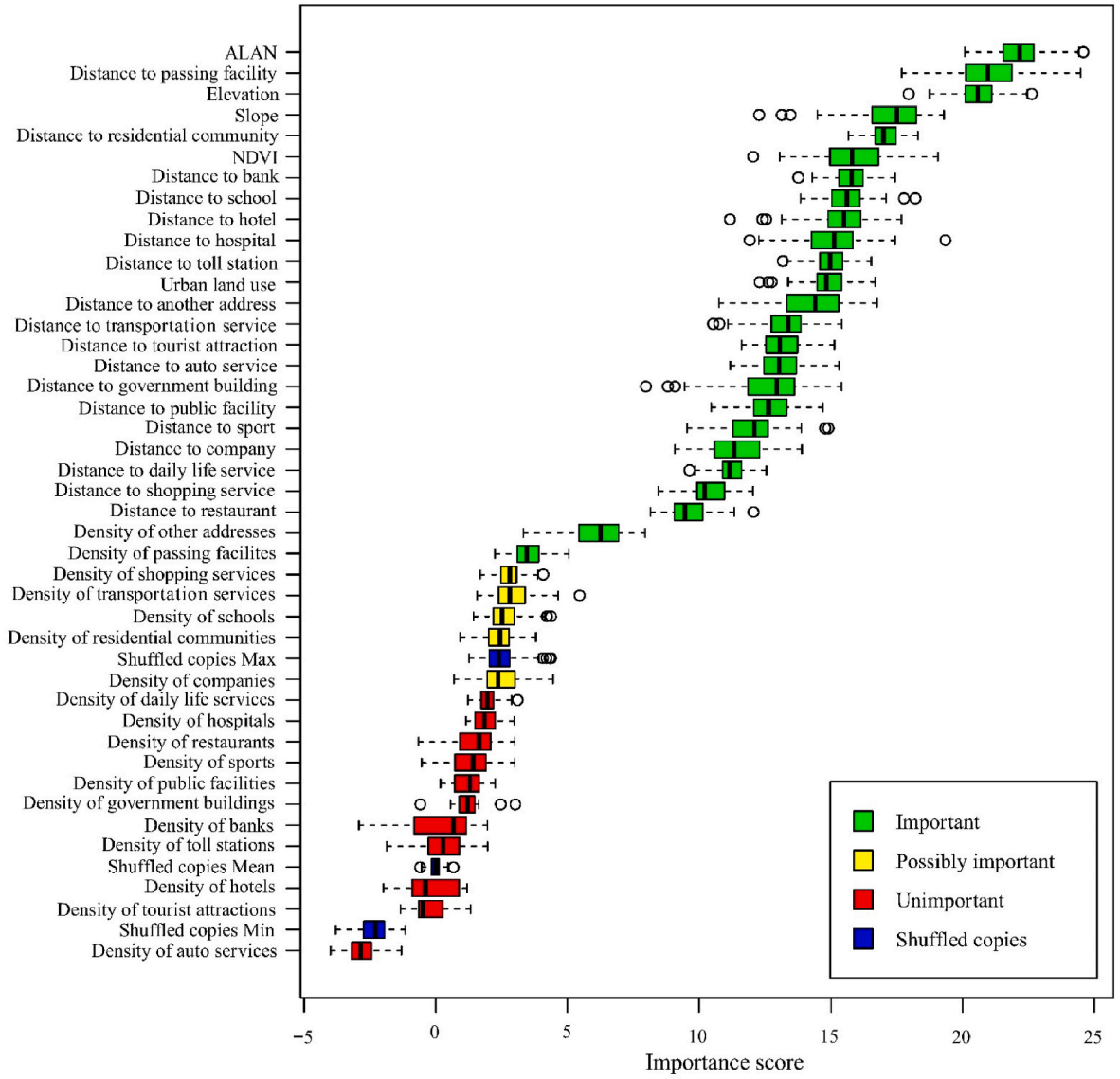


Fig. 4. Results of variable selection based on Boruta. "Density of *x*" denotes the density of the POIs of the category *x*, and "Distance to *x*" denotes the distance to the nearest POI of the category *x*. ALAN, artificial light at night; NDVI, normalized difference vegetation index; POI, point of interest.

where $P_i^{estimated}$ and $P_i^{observed}$ denote estimated and observed population count for each unit *i*, respectively, which can represent either a grid (for grid-level assessments) or a township (for township-level assessments); *N* denotes either the number of grids in the testing dataset (for grid-level assessments) or the number of townships in our study area (for township-level assessments). The expression of R^2 is shown as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{7}$$

where SS_{res} refers to the sum of squared residuals (i.e., differences) between the observed and estimated population counts and SS_{tot} to the total variability of the observed population counts.

2.5. Uncertainty propagation analysis of the CGWNN model

The input population reference for model training was generated using WorldPop-derived weights to redistribute township-level census data, resulting in modified population grids that integrate information from both WorldPop estimates and fine-scale census data (Zhuang et al., 2021). Although these modified population grids are approximate representations of real-world population distributions, uncertainties remain primarily due to potential errors in WorldPop estimates. For instance, previous studies reported that WorldPop misallocated around 20 % of the population onto unpopulated grids in urban areas (Thomson et al., 2022). To quantify the potential uncertainties introduced by WorldPop-derived weights, this study added random errors to these weights for assessment (Dai et al., 2020):

$$Weight_i^{'} = Weight_i + \xi_i, \xi_i \sim N\left(0, \left(\frac{\sigma}{3}\right)^2\right) \tag{8}$$

where $Weight_i^{'}$ denotes the adjusted weight for grid *i*; $Weight_i$ denotes the original weight for grid *i*; ξ_i denotes the normally distributed random error with a mean of zero and a standard deviation set to one-third of the standard deviation (i.e., σ) of the original weights. This approach ensures that the introduced noise reflects moderate variability without excessively distorting the original weight distribution, thus simulating deviations (i.e., errors of WorldPop estimates) observed in previous studies (Thomson et al., 2022). After that, a new CGWNN model was trained using the population reference generated with the adjusted weights. This model was compared to the CGWNN model trained using

Table 2
Characteristics of the variables in the sampling grids.

Variable (N = 3310)	Mean	Standard deviation	Median	Minimum	Maximum
Dependent variable					
Population (persons/ha)[a]	2.999	1.954	3.072	0.000	7.867
Independent variables					
Density of companies (units/ha)[a]	0.174	0.506	0.000	0.000	5.075
Density of passing facilities (units/ha)[a]	0.192	0.420	0.000	0.000	2.639
Density of residential communities (units/ha)[a]	0.100	0.292	0.000	0.000	2.639
Density of schools (units/ha)[a]	0.126	0.392	0.000	0.000	3.664
Density of shopping services (units/ha)[a]	0.239	0.616	0.000	0.000	4.836
Density of transportation services (units/ha)[a]	0.188	0.450	0.000	0.000	2.485
Density of other addresses (units/ha)[a]	0.297	0.571	0.000	0.000	4.094
ALAN (nW/cm^2/sr)	20.500	19.998	16.022	0.256	192.864
Elevation (m)	156.126	245.730	48.000	7.778	1606.670
Slope (°)	8.221	7.975	5.090	0.000	46.546
NDVI	0.602	0.192	0.596	0.089	0.943
Distance to the nearest auto service (m)	1516.141	2512.074	527.340	22.071	24,880.700
Distance to the nearest bank (m)	1797.382	2972.794	614.879	17.500	29,598.400
Distance to the nearest company (m)	889.317	1960.655	209.750	10.000	24,840.900
Distance to the nearest daily life service (m)	921.060	1515.964	272.634	5.000	13,104.000
Distance to the nearest government building (m)	835.198	1648.030	341.009	16.036	25,254.400
Distance to the nearest hospital (m)	1227.613	2058.905	467.592	17.500	25,254.400
Distance to the nearest hotel (m)	1582.387	2468.532	666.559	11.381	24,669.700
Distance to the nearest toll station (m)	1608.179	4124.595	229.788	12.500	49,867.500
Distance to the nearest public facility (m)	1007.826	1803.577	403.567	16.036	24,394.700
Distance to the nearest residential community (m)	1228.403	2166.464	342.930	13.333	24,840.000
Distance to the nearest restaurant (m)	989.884	1537.111	361.902	5.000	13,940.100
Distance to the nearest passing facility (m)	4602.885	7958.118	1450.325	38.262	66,140.400
Distance to the nearest school (m)	1084.298	2081.430	325.423	10.000	24,693.100
Distance to the nearest shopping service (m)	836.242	1380.304	276.466	2.500	13,866.700
Distance to the nearest sport (m)	843.943	1528.712	370.168	2.500	23,220.300
Distance to the nearest transportation service (m)	697.142	1450.350	257.811	13.333	24,303.500
Distance to the nearest tourist attraction (m)	1475.825	2173.421	800.667	26.095	24,669.700
Distance to the nearest another address (m)	322.675	408.143	173.223	0.000	5034.830

ALAN, artificial light at night; NDVI, normalized difference vegetation index.
[a] Being transformed by ln(x+1).

the population reference with original weights. The difference per grid between population products generated by the two models was quantified as follows:

$$\text{Difference} = \left| \frac{P_i' - P_i^{CGWNN}}{P_i^{CGWNN}} \right| \times 100\% \quad (9)$$

where P_i^{CGWNN} denotes the population estimated by the original model; P_i' denotes the population estimated using the model trained with adjusted weights. The accuracy of the CGWNN model trained by adjusted weights was also assessed using metrics mentioned in section 2.4 and compared with other products, to demonstrate the robustness of the findings.

2.6. Generalizability assessment of the CGWNN model

To further assess the generalizability of the proposed CGWNN model, it was applied to estimate gridded population density in Shanghai, a large city with relatively different population distributions and land use patterns compared to Beijing. The accuracy of the resulting population product was assessed by comparing it with census data in Shanghai at the township level. Four well-known population datasets, including WorldPop, GHS-POP, GPW, and LandScan Global, were also included in the comparison to evaluate the relative accuracy and, thereby, the generalizability of the CGWNN model.

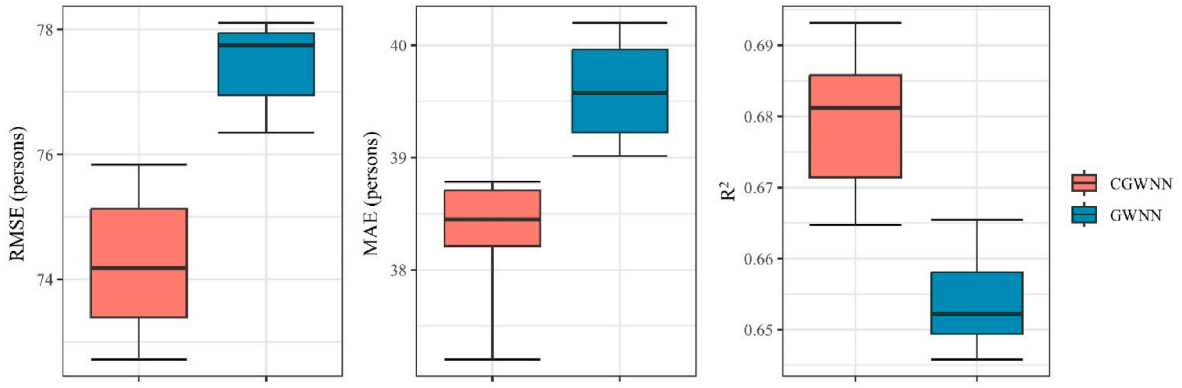


Fig. 5. Boxplots depicting the performance of 10 CGWNN models and 10 GWNN models. CGWNN, contextualized geographically weighted neural network; GWNN, geographically weighted neural network; MAE, mean absolute error; RMSE, root mean square error.

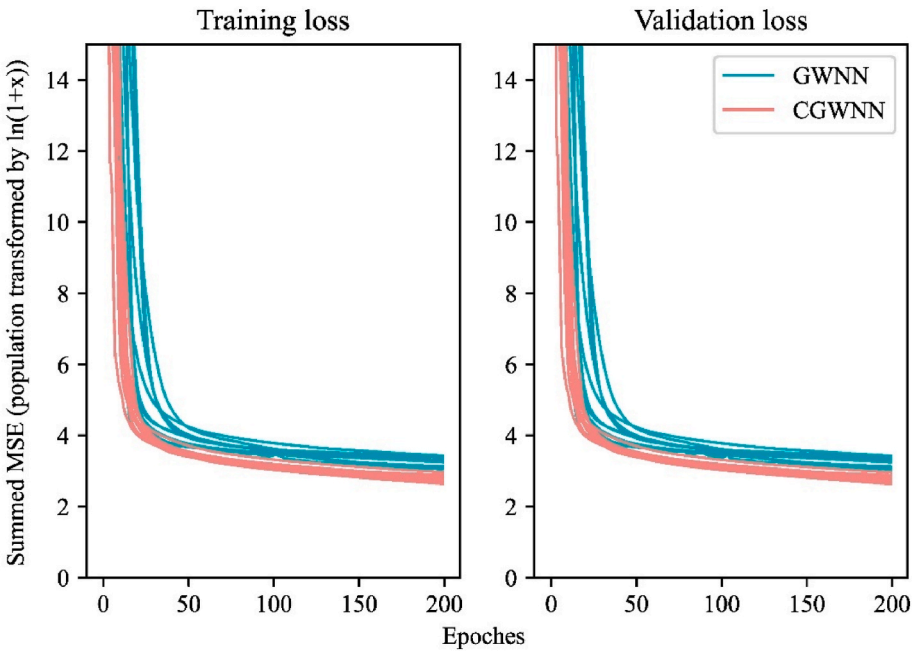


Fig. 6. Training and validation loss of 10 CGWNN and 10 GWNN models across epochs. CGWNN, contextualized geographically weighted neural network; GWNN, geographically weighted neural network; MSE, mean square error.

3. Results

3.1. Characteristics of the selected variables

Based on the Boruta variable selection, unimportant variables were discarded, such as densities of auto services, hotels, tourist attractions, and banks (Fig. 4). Characteristics of the important variables used for model training are shown in Table 2. Land use was also an important variable with 27.34 % of the sampling grids located in residential parcels, 1.30 % in business offices, 0.85 % in commercial services, 4.38 % in industrial parcels, 11.09 % on roads, 0.18 % in transportation stations, 1.45 % in airport facilities, 3.32 % in administrative parcels, 3.33 % in educational parcels, 0.66 % in medical parcels, 1.21 % in sport and cultural parcels, 11.54 % in park and greenspace, and 33.35 % in rural areas.

3.2. Comparisons of different models

The performance of the OLS, GWR-LU, GWR, GWNN, and CGWNN models was compared at the grid level. The GWR model was found to be more accurate than the OLS model ($R^2 = 0.648$ *v.s.* 0.484). The GWR-LU model was found to be more accurate than the GWR model ($R^2 = 0.651$ *v.s.* 0.648). The values of coefficients estimated by OLS, GWR, and GWR-LU models are shown in Table S1–S3.

Average performance of the 10 GWNN models was found to be better than the GWR-LU model ($R^2 = 0.654$ *v.s.* 0.651). Average performance of the 10 CGWNN models using land use as the contextual variable was better than that of the 10 GWNN models using land use as an independent variable ($R^2 = 0.679$ *v.s.* 0.654). Boxplots of the 10 CGWNN and 10 GWNN models revealed similar results of higher accuracy of CGWNN models (Fig. 5). During the model training phase, training loss and validation loss of CGWNN models decreased more rapidly than that of GWNN models as the number of epochs increased (Fig. 6). The Cohen's d values for CGWNN and GWNN models were 0.072 and 0.059, respectively. Given such small and negligible differences between training loss and validation loss, the CGWNN and GWNN models were less likely to experience significant overfitting. The best among the 10 CGWNN models demonstrated higher accuracy compared to the best among the 10 GWNN models ($R^2 = 0.686$ *v.s.* 0.660) (Table 3 and Fig. 7). These two best-performing models were subsequently used for visualization and population product comparison.

Table 3
Performance of different models at the grid level.

Models	MAE	RMSE	R^2
OLS	62.711	125.165	0.484
GWR	40.085	78.428	0.648
GWR-LU	40.294	78.155	0.651
GWNN	39.178	76.757	0.660
CGWNN	38.531	73.297	0.686

CGWNN, contextualized geographically weighted neural network; GWNN, geographically weighted neural network; GWR, geographically weighted regression; GWR-LU, geographically weighted regression with land use as an independent variable; MAE, mean absolute error (persons); OLS, ordinary least squares; RMSE, root mean square error (persons).

3.3. Comparisons of different population products

To assess relative accuracy of the population products generated by different models, each product was aggregated at the township level and compared with census population counts (Table 4 and Fig. 8). The GWR model demonstrated higher accuracy than the OLS model ($R^2 = 0.795$ *v.s.* 0.562). The higher accuracy was found in the GWR-LU model compared to the GWR model ($R^2 = 0.800$ *v.s.* 0.795). The GWNN model outperformed the GWR-LU model ($R^2 = 0.821$ *v.s.* 0.800). The CGWNN model achieved superior results compared to the GWNN model ($R^2 = 0.845$ *v.s.* 0.821). Additionally, the population product generated by the CGWNN model outperformed the other well-known population products, including WorldPop ($R^2 = 0.737$), GHS-POP ($R^2 = 0.802$), GPW ($R^2 = 0.675$), and LandScan Global ($R^2 = 0.664$).

The gridded population maps generated by different models in central Beijing can be found in Fig. 9. We also zoomed in on these maps specifically around Tian'anmen Square, where typical areas exhibit abrupt variations in population density and land use (Fig. 10). Tian'anmen Square located in the center of Beijing, and several landmarks such as the Forbidden City and Tian'anmen Square, have no residential

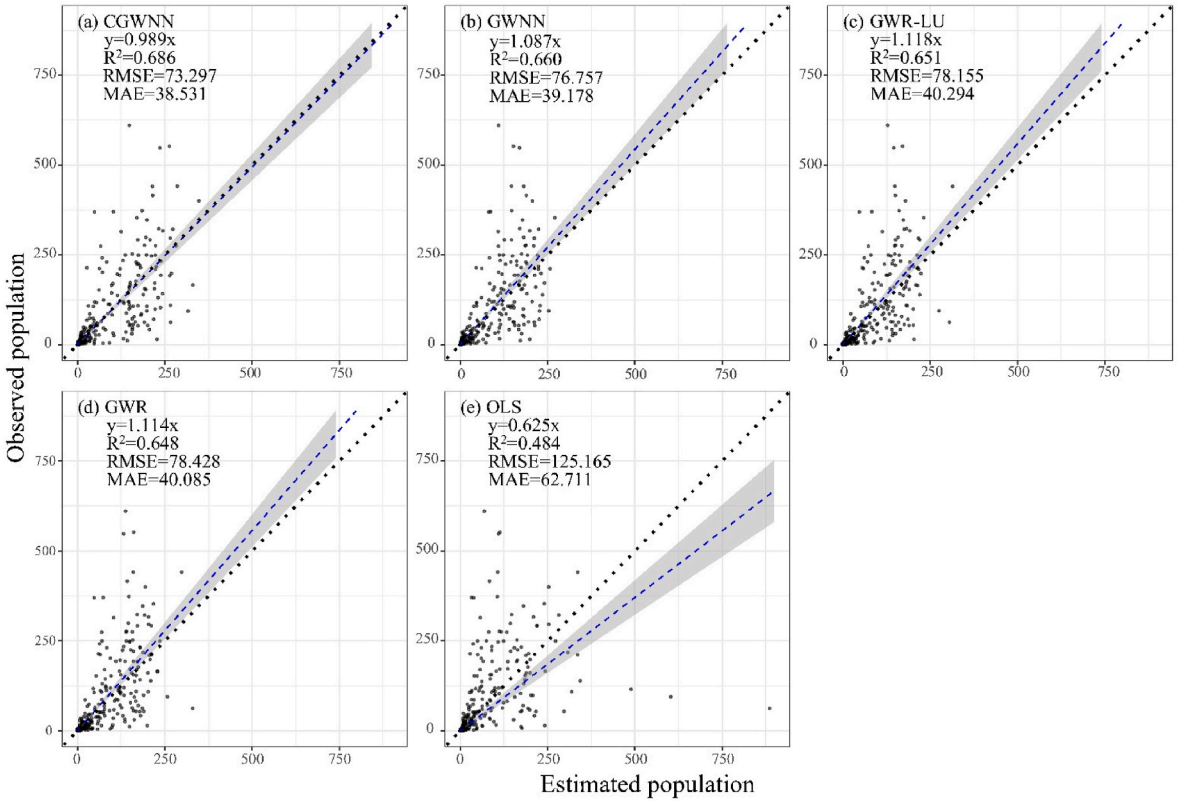


Fig. 7. Scatterplots between the observed population and the estimated population in the testing dataset: (a) the contextualized geographically weighted neural network (CGWNN) model, (b) the geographically weighted neural network (GWNN) model, (c) the geographically weighted regression model with land use as an independent variable (GWR-LU), (d) the geographically weighted regression (GWR) model, (e) the ordinary least squares (OLS) regression model. MAE, mean absolute error; RMSE, root mean square error.

Table 4
Relative accuracy of the population products at the township level.

	MAE	MAPE	RMSE	R^2
Models				
OLS	33,239.898	0.855	61,658.786	0.562
GWR	27,710.900	0.676	40,376.791	0.795
GWR-LU	27,187.691	0.654	39,839.486	0.800
GWNN	22,795.125	0.411	37,676.904	0.821
CGWNN	21,169.884	0.392	35,026.330	0.845
Other population products				
WorldPop	26,748.056	0.481	45,680.829	0.737
GHS-POP	37,729.679	0.506	60,845.741	0.802
GPW	36,318.302	0.925	62,240.487	0.675
LandScan Global	31,934.568	0.835	51,518.883	0.664

CGWNN, contextualized geographically weighted neural network; GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; GWNN, geographically weighted neural network; GWR, geographically weighted regression; GWR-LU, geographically weighted regression with land use as an independent variable; MAE, mean absolute error (persons); MAPE, mean absolute percentage error; OLS, ordinary least squares; RMSE, root mean square error (persons).

population; however, the surrounding residential areas in the center of Beijing are highly populated. CGWNN model estimated lower population density in those landmarks and higher population density in surrounding residential areas, compared to the GWNN, GWR-LU, GWR, and OLS models.

3.4. Evaluation of uncertainty propagation

Comparison between the CGWNN models trained with original and adjusted weights (i.e., original weights with random errors) indicated that more than 60 % of grids exhibited uncertainties below 100 % (Fig. 11), suggesting a minimal impact of potential inaccuracies in WorldPop-derived weights on the estimation results in urban areas (Fig. 12). However, higher uncertainties were observed in rural areas, likely due to the lower population densities, which resulted in smaller denominators and thus amplified the relative uncertainties.

The estimated populations of the CGWNN model trained by the adjusted weights with random errors were compared with observed populations in the testing dataset and township-level census data. Higher accuracy of this model (MAE = 39.091, RMSE = 75.201, R^2 = 0.670) than the GWNN model (MAE = 39.178, RMSE = 76.757, R^2 = 0.660) at the grid level was observed, demonstrating that potential errors of WorldPop-derived weights have minimal impact on the performance of the CGWNN model. At the township level, the CGWNN model trained by the adjusted weights (MAE = 23,583.467, RMSE = 36,268.423, R^2 = 0.837, MAPE = 0.525) also outperformed the GWNN model (MAE = 22,795.125, RMSE = 37,676.904, R^2 = 0.821, MAPE = 0.411) in terms of RMSE and R^2 but showed worse MAE and MAPE. This discrepancy arises because RMSE and R^2 are more sensitive to larger errors in highly populated urban areas, whereas MAE and MAPE are more influenced by the errors in rural areas with smaller populations. As this study focused on urban population mapping, RMSE and R^2 could be more suitable indicators.

3.5. Variable contributions in the CGWNN model

SHAP values demonstrated the associations between population density and variables in the CGWNN model (Fig. 13). Among the weighting variables, the contextual characteristics of grids (i.e., land use) were found to be more influential than their spatial coordinates. Most land use categories were generally associated with higher

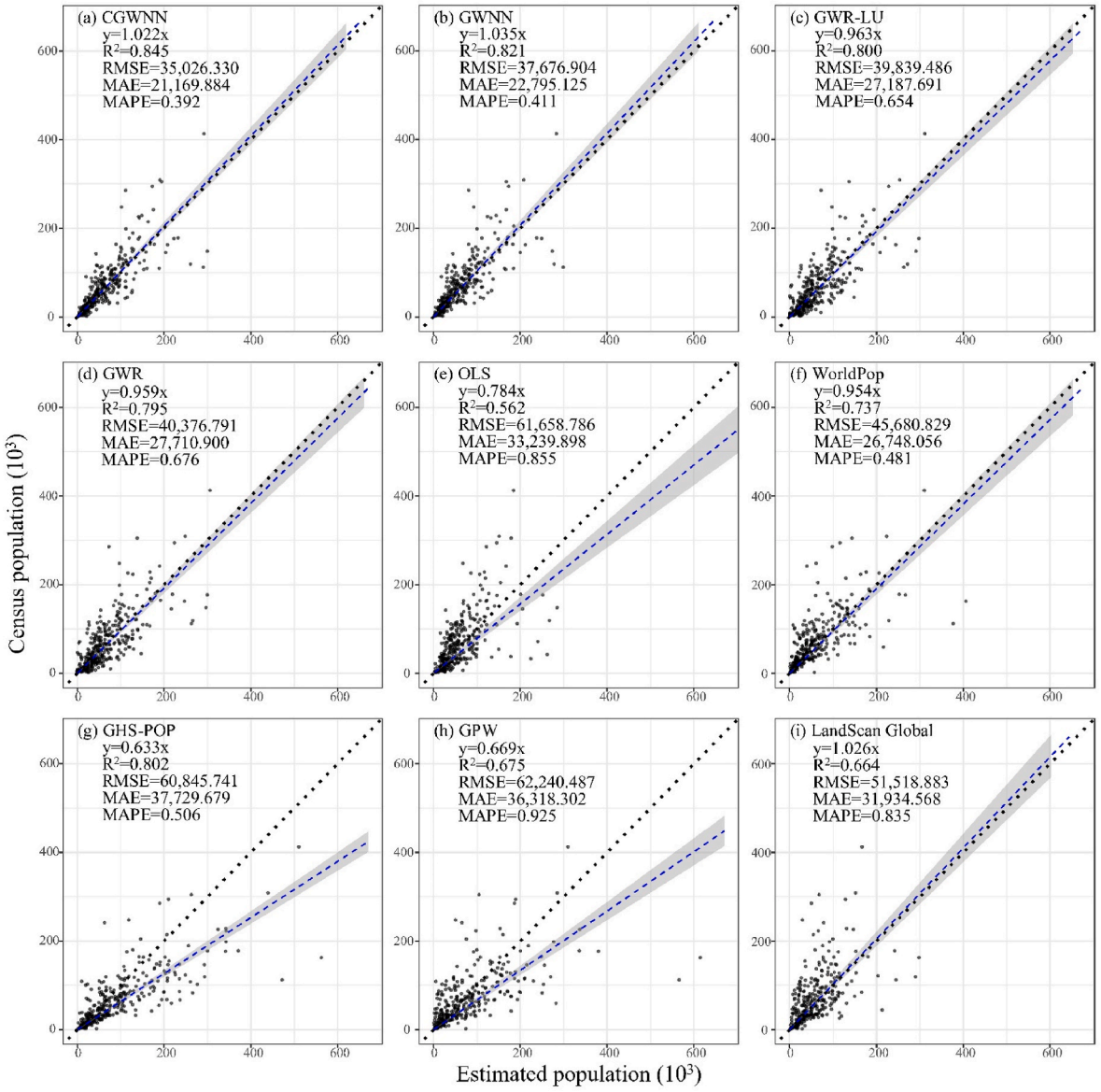


Fig. 8. Scatterplots between the census population and the estimated population generated by (a) the contextualized geographically weighted neural network (CGWNN) model, (b) the geographically weighted neural network (GWNN) model, (c) the geographically weighted regression model with land use as an independent variable (GWR-LU), (d) the geographically weighted regression (GWR) model, (e) the ordinary least squares (OLS) regression model, (f) WorldPop, (g) GHS-POP, (h) GPW, (i) LandScan Global. GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean square error.

population density, whereas roads and rural areas were generally associated with lower population density. For the independent variables, the three most important ones, were the distances to the nearest school, restaurant, and auto service. These variables tended to be negatively associated with population density, indicating that grids closer to these amenities tend to have higher population densities. The distances to the nearest POIs tended to be more influential than the densities of POIs. Moreover, the association between ALAN and population density was complex, as both positive and negative associations between ALAN and population density were found. Elevation and slope also exhibited non-monotonic relationships with population density. A negative association was observed between NDVI and population density.

3.6. Generalizability of the CGWNN model

When applied to Shanghai, the CGWNN model achieved the highest accuracy among all evaluated products with the lowest MAE of 33,015.658 and RMSE of 44,784.591, as well as the highest R^2 of 0.903, indicating strong generalizability (Tables 5 and S4). In contrast, WorldPop, GHS-POP, GPW, and LandScan Global exhibited lower accuracies in terms of these metrics. However, regarding MAPE, the CGWNN exhibited lower errors than GPW and LandScan Global, but higher errors than WorldPop and GHS-POP (Fig. 14). This discrepancy likely arises from the use of MSE as the loss function in the CGWNN model, which places greater emphasis on minimizing errors in high-population areas, whereas MAPE tends to highlight errors in low-population regions. Collectively, the established CGWNN model achieved the highest overall performance across multiple metrics in Shanghai, demonstrating relatively strong generalizability to new geographic areas. The CGWNN model generated population maps and the other population datasets in Shanghai are shown in Fig. 15.

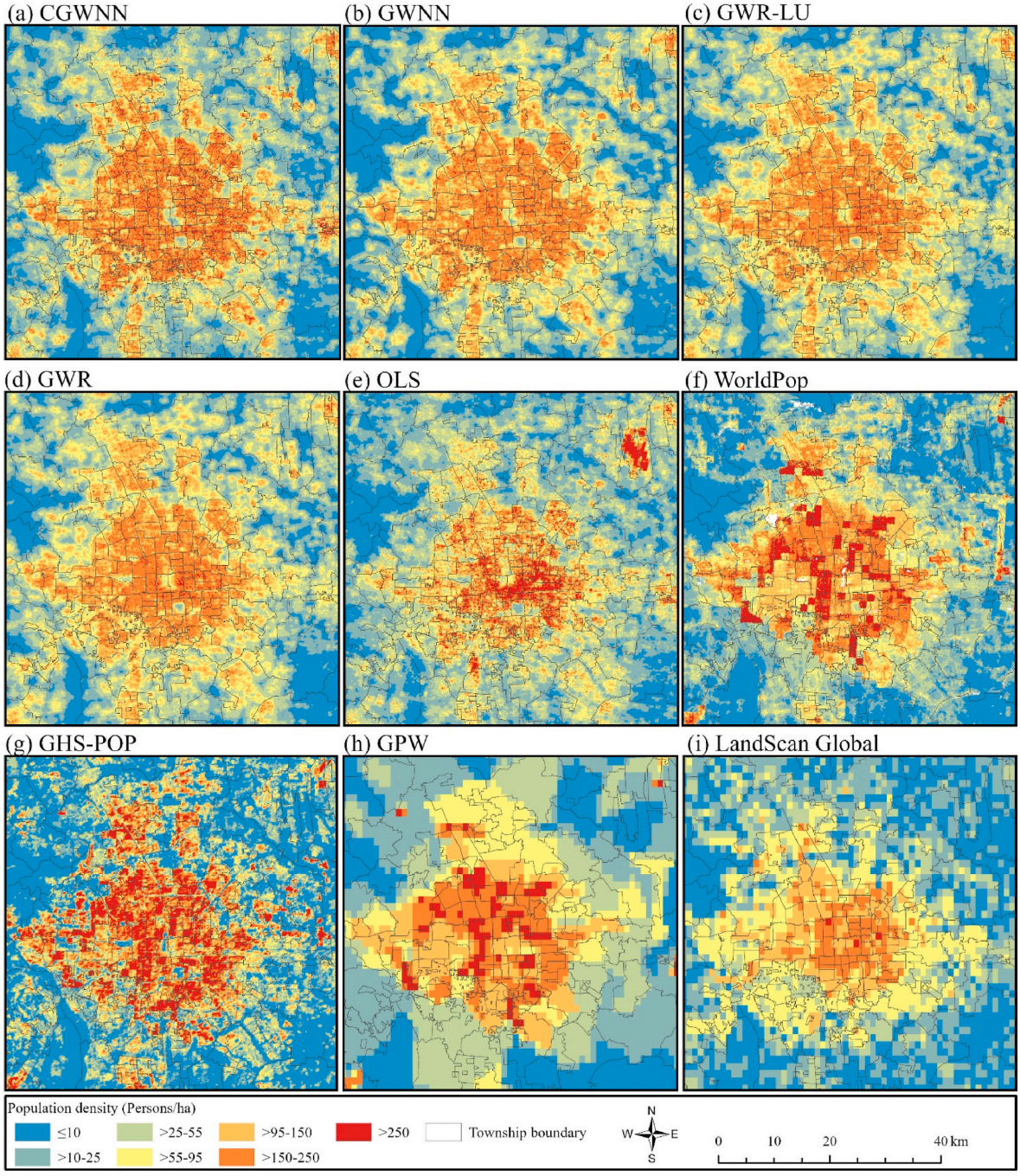


Fig. 9. Overview of gridded population maps generated by different models and the WorldPop, GHS-POP, GPW, and LandScan Global datasets in central Beijing, China in 2020. CGWNN, contextualized geographically weighted neural network; GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; GWNN, geographically weighted neural network; GWR, geographically weighted regression; GWR-LU, geographically weighted regression with land use as an independent variable; OLS, ordinary least squares.

4. Discussion

4.1. Summary of the key findings

This study developed a novel CGWNN model for gridded population estimation in complex urban environments based on remote sensing, social sensing, and land use data. CGWNN integrated ANN into GWR to account for nonlinear and spatially varying associations between influential factors and population density. It also considered abrupt variations in these associations across proximate grids situated within differing land-use parcels. Results showed that the population product generated by our CGWNN model achieved higher accuracy than that of GWNN, GWR-LU, GWR, and OLS models as well as the other well-known population datasets, including WorldPop, GHS-POP, GPW, and LandScan Global. The SHAP values indicated that the three most important variables for population estimation were the distances to the nearest school, restaurant, and auto service, all of which were negatively associated with population density. The contextual characteristic (i.e., land use) of grids was a more important weighting variable than their spatial coordinates.

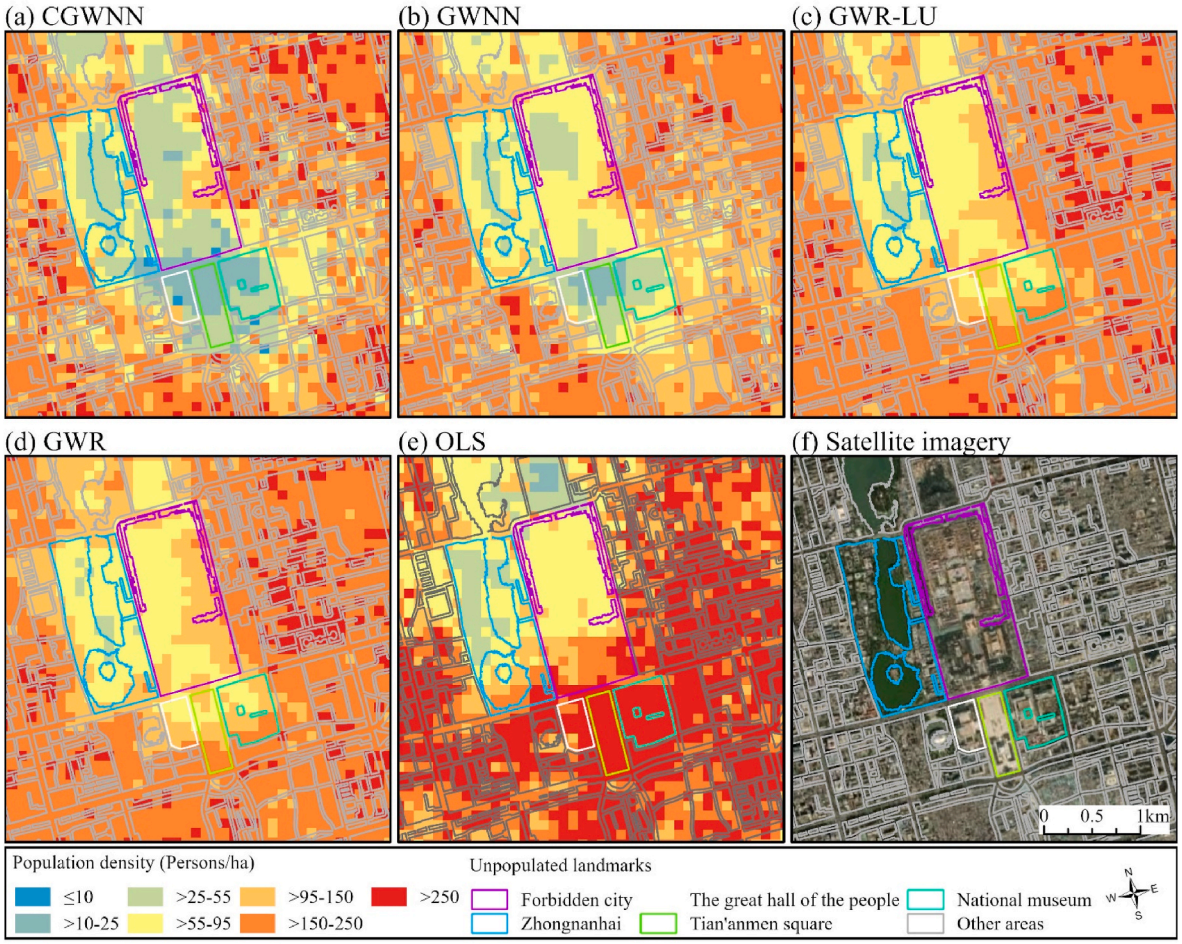


Fig. 10. Gridded population maps generated by different models around Tian'anmen Square, Beijing, China in 2020. CGWNN, contextualized geographically weighted neural network; GWNN, geographically weighted neural network; GWR, geographically weighted regression; GWR-LU, geographically weighted regression with land use as an independent variable; OLS, ordinary least squares.

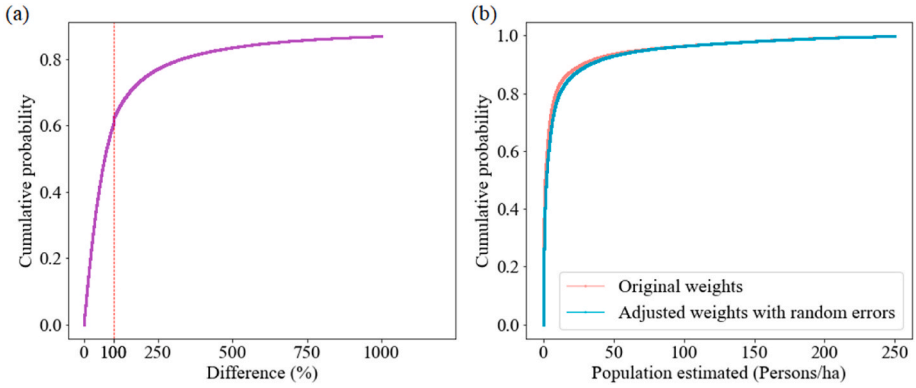


Fig. 11. Cumulative probability curve of (a) per grid difference between CGWNN models trained by population reference with original weights and with adjusted weights (i.e., original weights with random errors), (b) population density estimated by the two CGWNN models. CGWNN, contextualized geographically weighted neural network.

4.2. Comparisons of the methods

Associations between influential factors and population density varied spatially, as the GWR model, allowing local estimates of the associations, outperformed the OLS model that assumes globally unified associations over the study area. After integrating ANN into GWR, the GWNN model further considered nonlinear associations between influential factors and population density, thereby improving accuracy. However, both GWR and GWNN models rely solely on the spatial location of grids, disregarding the influence of contextual disparity on the associations between influential factors and population density. In complex urban environments, nearby grids may exhibit markedly different associations between influential factors and population density due to varying land uses across adjacent parcels. By using the contextual attribute (i.e., land use) as weights for independent variables, the proposed CGWNN model more exhaustively and explicitly estimated these

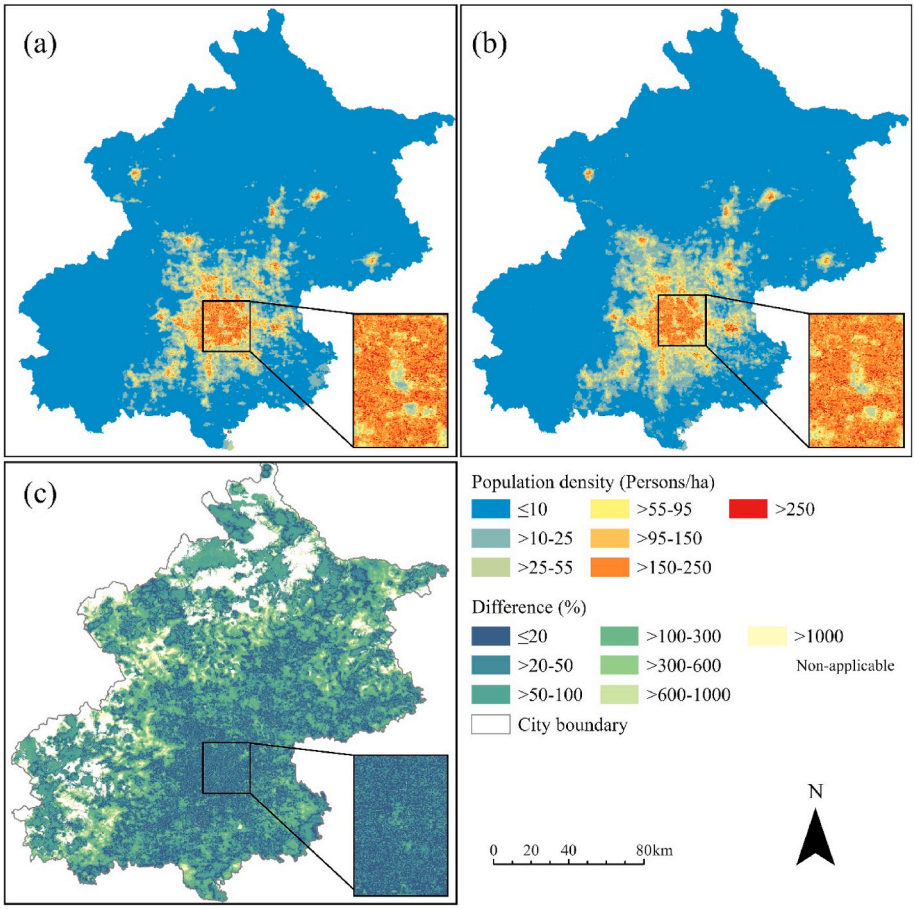


Fig. 12. The 100-m gridded population maps of all areas in Beijing, China in 2020: (a) generated by the CGWNN model, (b) generated by the CGWNN model using adjusted weights with random errors, (c) percentage difference between the two models. CGWNN, contextualized geographically weighted neural network.

contextually varying associations, resulting in a faster convergence rate during training and higher accuracy of population products. The key advantage of CGWNN lies in its ability to capture nonlinear associations between influential factors and population density (i.e., via neural networks), while also accounting for contextual variations in these associations (i.e., via the contextualized geographically weighted structure). Compared to the previously proposed models that also combined ANN with GWR (Du et al., 2020; Zhang et al., 2022), the CGWNN model in this study stands out by not relying on the idealized and pre-defined spatial weight matrix, thereby allowing more flexible spatial relationships to fit complex geographical processes from a knowledge-driven perspective (Feng et al., 2021). By incorporating contextual information rather than relying solely on spatial coordinates, which are merely surrogates for factors affecting associations, the CGWNN model offers a more precise and context-aware estimation of population density.

The applicability and generalizability of the CGWNN model were supported by our uncertainty and generalizability analyses. Results demonstrated that relatively small uncertainties in the dependent variable, population density, have minimal impact on the accuracy of the model, indicating a relatively strong robustness of the CGWNN model. Additionally, the model was successfully applied to generate the population product in Shanghai, a city with relatively different population distribution patterns compared to Beijing, where the model was trained. Results showed that the CGWNN model outperformed well-known population datasets, including WorldPop, GHS-POP, GPW, and LandScan Global, highlighting the relatively strong **generalizability** and **usefulness** of the CGWNN model across broader geographic areas.

4.3. *Interpretations of variable contributions*

Urban land use, representing the most extensive human modification on Earth, reflects the socio-economic functions of urban parcels (Chen et al., 2021), which are directly related to population distribution. This is further supported by the improved accuracy of the GWR-LU model, which included land use as an independent variable, compared to the GWR model without it. Although land use is a direct and valuable factor for population estimation, it typically provides information only at the parcel level. However, residential parcels may contain retail stores and public service facilities, while industrial parcels may include staff dormitories or office buildings. As such, commonly available land use data may not fully capture the detailed composition and functional diversity within a parcel (Chen et al., 2021), thus limiting its ability to directly represent population-related characteristics at the grid level. Nevertheless, land use still reflects the predominant socio-economic function of a parcel to a considerable extent and thus provides contextual information for the grids within it. This socio-economic context may better serve as a contextual variable influencing the associations between influential factors and population density, rather than being directly used as an independent variable, for population estimation. This is evidenced by the higher accuracy of the CGWNN model compared to the GWNN model in this study. Previous studies have also indicated that land use types can impact the relationships between influential factors and population density (Wang et al., 2018), highlighting the importance of incorporating land use as a contextual variable for more accurate population estimation.

The top three most influential factors in the CGWNN model, identified by SHAP values, were the distances to the nearest school,

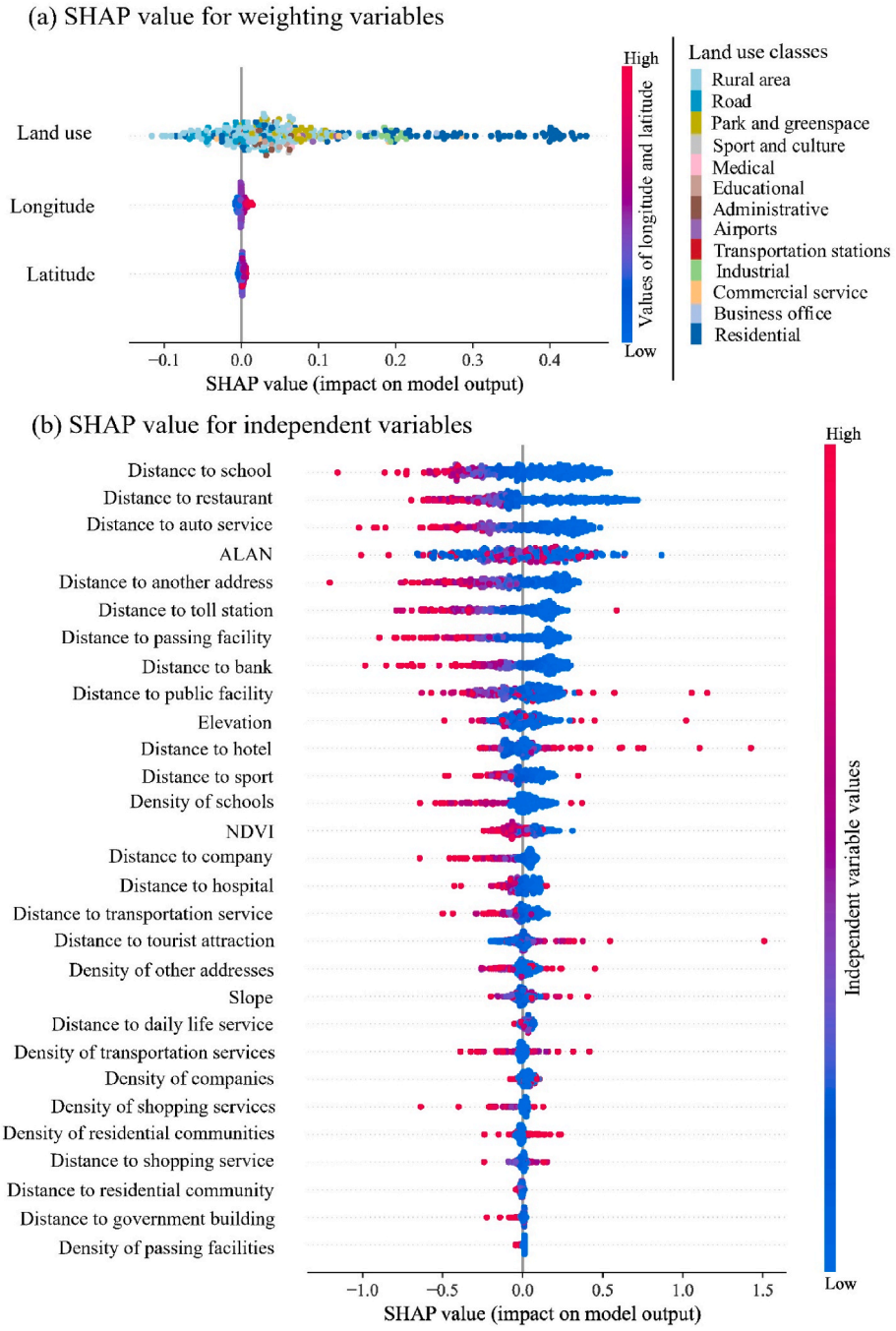


Fig. 13. Associations between population density and variables calculated by SHapley Additive exPlanations (SHAP) based on the contextualized geographically weighted neural network model: (a) weighting variables (i.e., longitude, latitude, and land use), (b) independent variables. The variable importance ranking is based on the sum of SHAP magnitudes. "Density of *x*" denotes the density of the POIs of the category *x*, and "Distance to *x*" denotes the distance to the nearest POI of the category *x*. ALAN, artificial light at night; NDVI, normalized difference vegetation index; POI, point of interest.

restaurant, and auto service. These factors generally showed negative associations with population density, suggesting that areas with better access to these essential services and amenities tend to have higher population densities. This finding is consistent with previous studies that employed similar factors for population mapping (Qiu et al., 2020). However, the densities of schools, restaurants, and auto services themselves were found to generally have negative associations with population density. This may be because this study focused on estimating residential populations, and areas with high densities of these amenities, particularly in 100-m grids, might have limited residential space. Among the POI-derived independent variables, the densities of certain categories of POIs were excluded from population estimation based on Boruta variable selection, as they were found to be unimportant for population estimation in this study. Given the potential positioning uncertainties of POI data, calculating POI density at a fine spatial scale (i.e., 100m) may introduce inaccuracies in data representation, leading to excessive noise or errors that could reduce their relevance for population estimation (Yeow et al., 2021).

The association between ALAN and population density was complex, as both positive and negative associations between ALAN and population density were found. While ALAN is typically positively associated with population density when using large analysis units (e.g., counties

Table 5
Relative accuracy of the population products at the township level in Shanghai.

Datasets	MAE	MAPE	RMSE	R^2
CGWNN	33,015.658	0.693	44,784.591	0.903
WorldPop	36,097.446	0.562	49,690.096	0.893
Global Human Settlement Population Grid	45,086.691	0.397	64,623.873	0.890
Gridded Population of the World	42,139.426	0.766	56,933.810	0.848
LandScan Global	52,771.033	0.913	72,923.586	0.831

CGWNN is the contextualized geographically weighted neural network model trained in Beijing. MAE, mean absolute error (persons); MAPE, mean absolute percentage error; RMSE, root mean square error (persons).

or larger administrative units) (Tan et al., 2018; Wang et al., 2018), this association may not hold at finer scales (e.g., 100m or finer). This is because, at larger spatial units, local variations tend to be averaged out, and ALAN primarily captures the general intensity of human activities, which closely related to population distribution. However, at finer scales, ALAN intensity is influenced by more localized and diverse factors beyond residential population density, such as commercial activities and transportation infrastructure (Wu et al., 2024), implying that areas with strong illumination can also have low population density. Similarly, higher values of elevation and slope were not necessarily associated with lower population density in this study, which differs from previous studies conducted over broad regions (Tremblay & Ainslie, 2021). This is likely because our study focuses on urban population estimation within a relatively small area, where variations in topography could be limited. Moreover, NDVI was negatively associated with population density, consistent with previous studies (Dobbs et al., 2017; Zhang et al., 2021), indicating its reliability in estimating population density.

4.4. Strengths and limitations

This study has three major advantages. First, a novel CGWNN was developed for gridded population estimation based on the nonlinear and spatially varying associations between influential factors and population density. The model also accounted for abrupt variations in these associations across nearby parcels. Such variations typically occur in complex urban environments, where rapid changes in land use can lead to substantial shifts in population density over short distances. Second, this study utilized publicly available land use data, EULUC-China, to account for the socio-economic functions of urban parcels that are relevant to population distribution. As EULUC-China data cover impervious surfaces throughout the country, our method can be extended to larger areas for more accurate population estimation. Third, this study directly used fine-scale samples to produce the gridded population product covering the study area, employing an approach that did not require census data for top-down disaggregation (Wardrop et al., 2018). As the national census is often conducted every 10 years in most countries, and could be even inaccessible in resource-poor settings, using such a method could be more feasible to estimate population distribution for broader use (Leasure et al., 2020).

There were also limitations in this study. First, population density in sampling grids for model training contained uncertainties (Zhuang et al., 2021). While our uncertainty analysis indicated that these inaccuracies had a minimal impact on the results, it would be preferred to use more accurate measures — such as micro-census data collected through household surveys — to obtain fine-scale population density for population estimation. Second, while the CGWNN model demonstrated relatively strong generalizability in data-rich environments like Shanghai, the applicability of the model in data-scarce areas, particularly in Global South cities with prevalent informal settlements and irregular land use patterns, requires further validation. Future studies could leverage more satellite-derived ancillary data to enhance

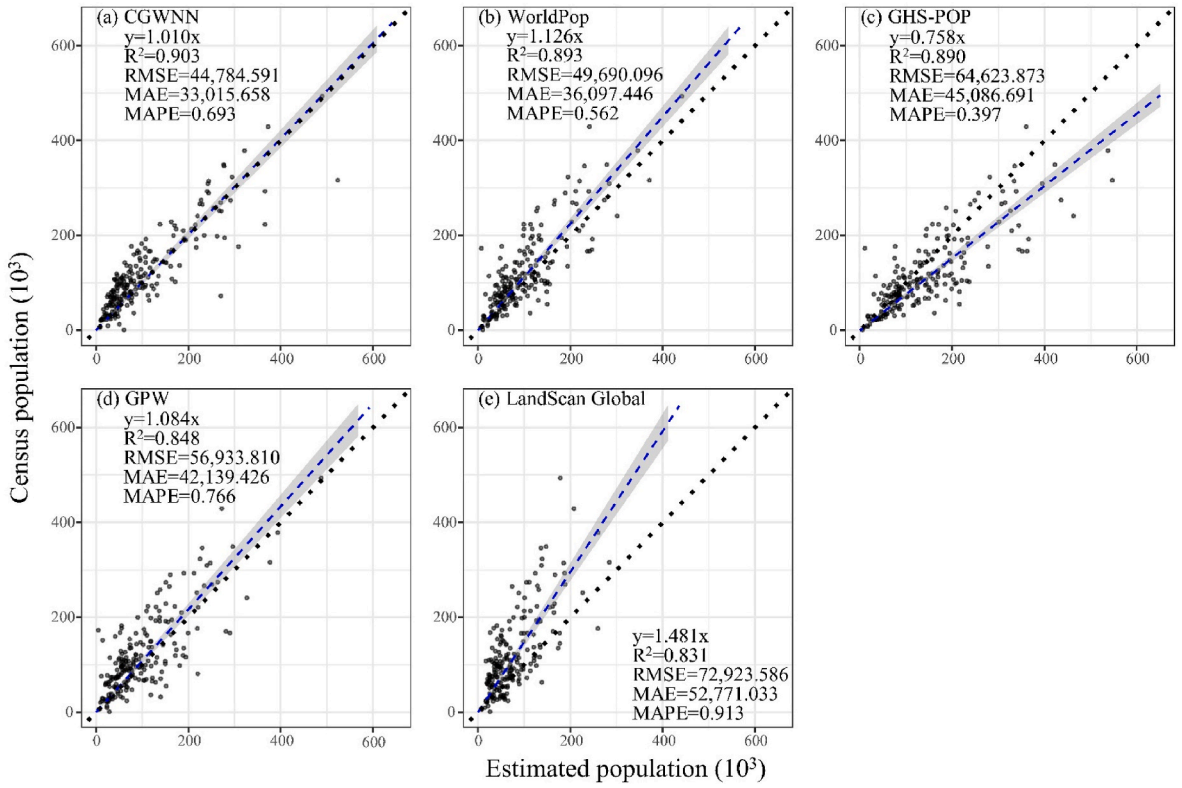


Fig. 14. Scatterplots between the census population and the estimated population in Shanghai generated by (a) the contextualized geographically weighted neural network (CGWNN) model (trained in Beijing), (b) WorldPop, (c) GHS-POP, (d) GPW, (e) LandScan Global. GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World; MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean square error.

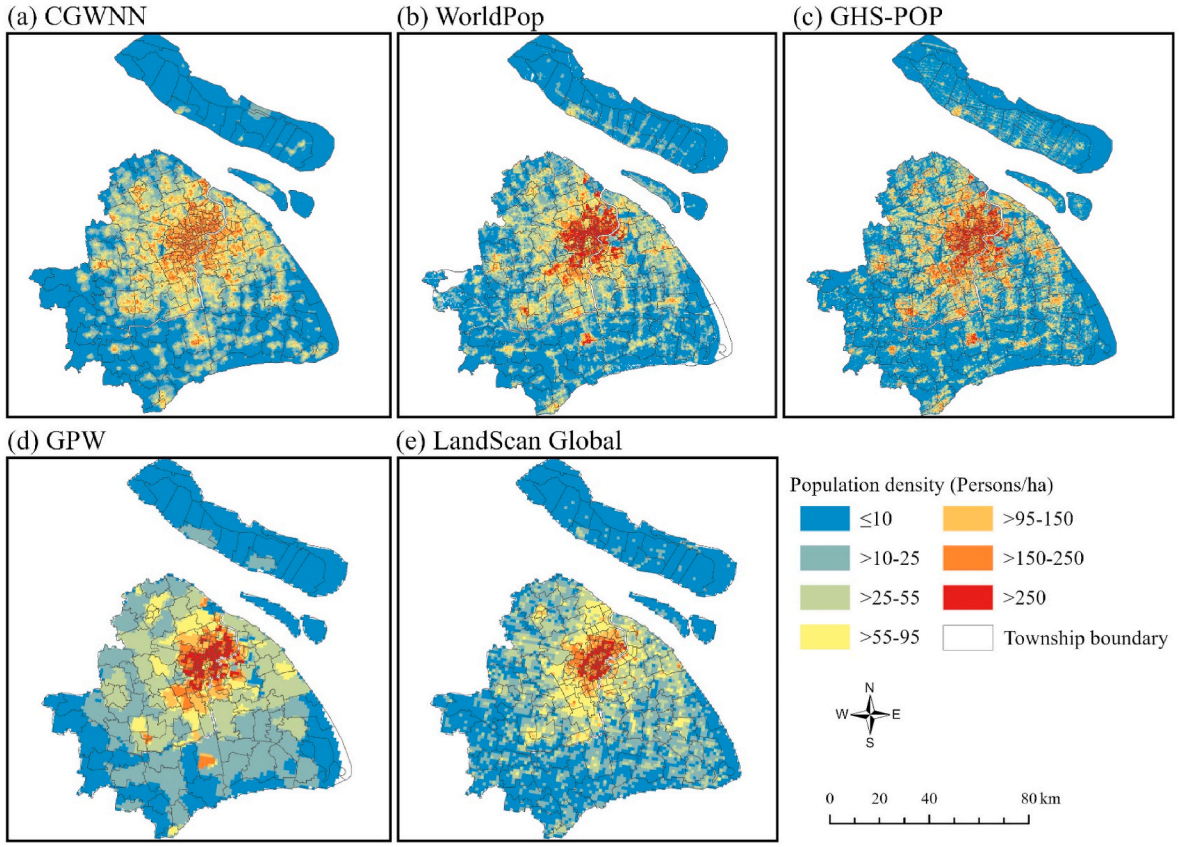


Fig. 15. The 100-m gridded population maps of Shanghai city, China in 2020: (a) generated by the CGWNN model (trained in Beijing), (b) WorldPop, (c) GHS-POP, (d) GPW, (e) LandScan Global datasets. CGWNN, contextualized geographically weighted neural network; GHS-POP, Global Human Settlement Population Grid; GPW, Gridded Population of the World.

modeling applicability in these areas. For instance, satellite-based slum mapping could be used to capture informal settlements (Verma et al., 2019), while high-resolution satellite imagery could facilitate land use classification (Zhou et al., 2020), providing relatively detailed land use categories for CGWNN-based population estimation. Third, this study only considered land use as a contextual variable, but other factors, such as crime rates and pollutant concentrations, could also lead to abrupt variations in the associations between influential factors and population density across nearby parcels. Future studies could incorporate these contextual variables as needed to better estimate population distribution. Fourth, the EULUC-China dataset reflected land use in 2018, which was not aligned with targeting population estimates in 2020, leading to potential uncertainties. However, as land use and its associated socio-economic patterns tend to remain stable over short periods, especially in developed cities like Beijing (Mulligan, 2013), the extent of such uncertainties is likely to be minimal. Fifth, the POI data were acquired from Amap, which is one of the largest web map service providers in China, offering freely accessible and reliable POIs. The dataset has been widely utilized for population estimation in China (Chen et al., 2024). Although Amap data have also been available in some other countries, future research should consider using more extensive datasets, such as Foursquare or Overture POIs, to further enhance model applicability across different regions.

4.5. *Usefulness of gridded population data*

Spatially accurate population data are vital for understanding and responding to various social, economic, and environmental issues (Elvidge et al., 1997; Rahman & Alam, 2021). The high-resolution gridded population datasets generated by our modeling approach provide standardized, georeferenced cells of population density, which are essential for effective planning and resource allocation across multiple fields. For instance, in urban planning, population data enable more informed decisions regarding infrastructure development (Shi et al., 2020), such as greenspace planning, as areas with higher population density may require larger public spaces to meet the needs of residents (Wu et al., 2023). Additionally, analyzing population data can inform emergency response strategies, enabling authorities to allocate resources more effectively during natural disasters or public health emergencies (Martins, 2021; Smith et al., 2019). Overall, the availability of high-resolution population data enhances the capacity for targeted interventions and sustainable development practices.

5. Conclusion

This study developed a novel CGWNN model to estimate population density at the grid level in urban areas, based on remote sensing, social sensing, and land use data. The CGWNN captured the nonlinear and spatially varying associations between influential factors and population density, and allowed them to differ by land-use contexts across grids. It outperformed other models that ignored such contextual heterogeneities, demonstrating the benefits of integrating spatial contexts into population estimation. The most important variables in the CGWNN model were the distances to the nearest school, restaurant, and auto service, all of which were negatively associated with population density. Additionally, ALAN intensity showed both positive and negative associations with population density in different regions, suggesting that increased ALAN did not necessarily indicate higher population density in urban areas. Our modeling method has potential to accurately estimate complex urban population distribution and, more broadly, other

spatial occurrences that exhibit contextually heterogeneous distributions, such as the gross domestic product and carbon emissions, which holds promise for various fields depending on these essential data.

Funding

This study was supported by the National Key R&D Program of China (2023YFC3604703), National Natural Science Foundation of China (42271433), Fundamental Research Funds for the Central Universities (413000159), and Seed Fund Program for WHU-DKU Joint Scientific Research Platform of Wuhan University (WHUDKUZZJJ202308).

CRediT authorship contribution statement

Ge Qiu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Yuchen Li:** Writing – review & editing, Validation. **Kun Qin:** Writing – review & editing, Data curation. **Chen Li:** Writing – review & editing, Data curation. **Shuhan Yang:** Software. **Chun Yin:** Writing – review & editing. **Yaolin Liu:** Supervision, Resources. **Shaoqing Dai:** Writing – review & editing, Visualization, Validation, Supervision, Formal analysis. **Peng Jia:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2025.103708.

References

Bakillah, M., Liang, S., Mobasheri, A., Arsanjani, J. J., & Zipf, A. (2014). Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science, 28*(9), 1940–1963. https://doi.org/10.1080/13658816.2014.909045

Boo, G., Darin, E., Leasure, D. R., Dooley, C. A., Chamberlain, H. R., Lazar, A. N., & Tatem, A. J. (2022). High-resolution population estimation using household survey data and building footprints. *Nature Communications, 13*(1), 10. https://doi.org/10.1038/s41467-022-29094-x

Boriah, S., Chandola, V., & Kumar, V. (2008). Similarity measures for categorical data: A comparative evaluation. *Proceedings of the 2008 SIAM international conference on data mining, carioli, A*.

Carioli, A., Schiavina, M., Freire, S., & MacManus, K. (2023). GHS-POP R2023A - GHS population grid multitemporal (1975-2030). https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE.

Center For International Earth Science Information Network-CIESIN-Columbia University. (2018). *Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision, 11*. https://doi.org/10.7927/H46M34XX. *Version 4.11*.

Chen, Y. H., Xu, C. C., Ge, Y., Zhang, X. X., & Zhou, Y. N. (2024). A 100 m gridded population dataset of China's seventh census using ensemble learning and big geospatial data. *Earth System Science Data, 16*(8), 3705–3718. https://doi.org/10.5194/essd-16-3705-2024

Chen, B., Xu, B., & Gong, P. (2021). Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities. *Big Earth Data, 5*(3), 410–441. https://doi.org/10.1080/20964471.2021.1939243

Cheng, L. X., Wang, L. Z., Feng, R. Y., & Yan, J. N. (2021). Remote sensing and social sensing data fusion for fine-resolution population mapping with a multimodel neural network. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 5973–5987. https://doi.org/10.1109/jstars.2021.3086139

Dai, S. Q., Ren, Y., Zuo, S. D., Lai, C. Y., Li, J. J., Xie, S. Y., & Chen, B. C. (2020). Investigating the uncertainties propagation analysis of CO_2 emissions gridded maps at the urban scale: A case study of jinjiang city, China. *Remote Sensing, 12*(23), 3932. https://doi.org/10.3390/rs12233932

Didan, K. (2021). *MODIS/terra vegetation indices 16-Day L3 global 250m SIN grid V061*. https://doi.org/10.5067/MODIS/MOD13Q1.061

Ding, J.-W., Lu, D.-G., & Dong, Y. (2025). Seismic spatiotemporal assessment of indoor occupant casualties in regional buildings: A Bayesian network approach incorporating population density dynamics. *International Journal of Disaster Risk Reduction, 126*, Article 105637. https://doi.org/10.1016/j.ijdrr.2025.105637

Dobbs, C., Nitschke, C., & Kendal, D. (2017). Assessing the drivers shaping global patterns of urban vegetation landscape structure. *Science of the Total Environment, 592*, 171–177. https://doi.org/10.1016/j.scitotenv.2017.03.058

Du, Z. H., Wang, Z. Y., Wu, S. S., Zhang, F., & Liu, R. Y. (2020). Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. *International Journal of Geographical Information Science, 34*(7), 1353–1377. https://doi.org/10.1080/13658816.2019.1707834

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing, 18*(6), 1373–1379.

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F. C., & Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing, 13*(5), 922. https://doi.org/10.3390/rs13050922

Feng, L. W., Wang, Y. M., Zhang, Z., & Du, Q. Y. (2021). Geographically and temporally weighted neural network for winter wheat yield prediction. *Remote Sensing of Environment, 262*, Article 112514. https://doi.org/10.1016/j.rse.2021.112514

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A, 30*(11), 1905–1927. https://doi.org/10.1068/a301905

Gaughan, A. E., Stevens, F. R., Huang, Z. J., Nieves, J. J., Sorichetta, A., Lai, S. J., & Tatem, A. J. (2016). Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Scientific Data, 3*, Article 160005. https://doi.org/10.1038/sdata.2016.5

Gong, P., Chen, B., Li, X. C., Liu, H., Wang, J., Bai, Y. Q., & Xu, B. (2020). Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Science Bulletin, 65*(3), 182–187. https://doi.org/10.1016/j.scib.2019.12.007

Griffith, D. A. (2006). Hidden negative spatial autocorrelation. *Journal of Geographical Systems, 8*(4), 335–355. https://doi.org/10.1007/s10109-006-0034-9

Harris, R., Dong, G. P., & Zhang, W. Z. (2013). Using contextualized geographically weighted regression to model the spatial heterogeneity of land prices in beijing, China. *Transactions in GIS, 17*(6), 901–919. https://doi.org/10.1111/tgis.12020

Jacobsen, K., & Passini, R. (2010). *Analsysis of ASTER GDEM elevation models. International archives of the photogrammetry*. Remote Sensing and Spatial Information Sciences:[2010 Canadian Geomatics Conference And Symposium Of Commission I, ISPRS Convergence In Geomatics-Shaping Canada's Competitive Landscape] 38 (2010), Nr. Part 1.

Jia, P., & Gaughan, A. E. (2016). Dasymetric modeling: a hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Applied Geography*, 100–108. https://doi.org/10.1016/j.apgeog.2015.11.006

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. *Applied Geography, 50*(2), 99–107.

Jia, P., Shi, X., & Xierali, I. M. (2019). Teaming up census and patient data to delineate fine-scale hospital service areas and identify geographic disparities in hospital accessibility. *Environmental Monitoring and Assessment, 191*, 303. https://doi.org/10.1007/s10661-019-7413-4

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the boruta package. *Journal of Statistical Software, 36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Leasure, D. R., Jochem, W. C., Weber, E. M., Seaman, V., & Tatem, A. J. (2020). National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences of the United States of America, 117*(39), 24173–24179. https://doi.org/10.1073/pnas.1913050117

Lei, Z., Xie, Y. J., Cheng, P. G., & Yang, H. C. (2023). From auxiliary data to research prospects, a review of gridded population mapping. *Transactions in GIS, 27*(1), 3–39. https://doi.org/10.1111/tgis.13020

Mallick, S. K., Das, P., Maity, B., Rudra, S., Pramanik, M., Pradhan, B., & Sahana, M. (2021). Understanding future urban growth, urban resilience and sustainable development of small cities using prediction-adaptation-resilience (PAR) approach. *Sustainable Cities and Society, 74*, 103196. https://doi.org/10.1016/j.scs.2021.103196

Martin, D. (2011). *Directions in population GIS*. https://doi.org/10.1111/j.1749-8198.2011.00440.x, 5(9.

Martins, P. R. (2021). Relationship between population density and COVID-19 incidence and mortality estimates: A county-level analysis. *Journal of Infection and Public Health, 14*(8), 1087–1088. https://doi.org/10.1016/j.jiph.2021.06.018

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer, 55*(1), 31–42.

Mulligan, G. F. (2013). Revisiting the urbanization curve (reprinted from cities, vol 32, pg 113-122, 2013). *Cities, 32*, S58–S67. https://doi.org/10.1016/j.cities.2013.07.009

Piao, S. L., Fang, J. Y., Zhou, L. M., Guo, Q. H., Henderson, M., Ji, W., & Tao, S. (2003). Interannual variations of monthly and seasonal normalized difference vegetation index (NDVI) in China from 1982 to 1999. *Journal of Geophysical Research-Atmospheres, 108*(D14), 4401. https://doi.org/10.1029/2002jd002848

Prechelt, L. (1998). Early stopping - But when? In G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 55–69). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-49430-8_3.

Qiu, G., Bao, Y. H., Yang, X. C., Wang, C., Ye, T. T., Stein, A., & Jia, P. (2020). Local population mapping using a random forest model based on remote and social sensing data: A case study in zhengzhou, China. *Remote Sensing, 12*(10), 1618. https://doi.org/10.3390/rs12101618

Rahman, M. M., & Alam, K. (2021). Clean energy, population density, urbanization and environmental pollution nexus: Evidence from Bangladesh. *Renewable Energy, 172*, 1063–1072. https://doi.org/10.1016/j.renene.2021.03.103

Roni, R., & Jia, P. (2020). An optimal population modeling approach using geographically weighted regression based on high-resolution remote sensing data: A case study in Dhaka City, Bangladesh. *Remote Sensing, 12*(7), 1184. https://doi.org/10.3390/rs12071184

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 of the United States of America, 109*(40), 16083–16088. https://doi.org/10.1073/pnas.1211658109

Shi, Y., Yang, J. Y., & Shen, P. Y. (2020). Revealing the correlation between population density and the spatial distribution of urban public service facilities with Mobile phone data. *ISPRS International Journal of Geo-Information, 9*(1), 38. https://doi.org/10.3390/ijgi9010038

Sims, K., Reith, A., Bright, E., Kaufman, J., Pyle, J., Epting, J., & Rose, A. (2023). *LandScan global 2022*. https://doi.org/10.48690/1529167

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications, 10*, 1814. https://doi.org/10.1038/s41467-019-09282-y

Song, J. C., Tong, X. Y., Wang, L. Z., Zhao, C. L., & Prishchepov, A. V. (2019). Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach. *Landscape and Urban Planning, 190*, Article 103580. https://doi.org/10.1016/j.landurbplan.2019.05.011

Tan, M. H., Li, X. B., Li, S. J., Xin, L. J., Wang, X., Li, Q., & Xiang, W. L. (2018). Modeling population density based on nighttime light images and land use data in China. *Applied Geography, 90*, 239–247. https://doi.org/10.1016/j.apgeog.2017.12.012

Thomson, D. R., Leasure, D. R., Bird, T., Tzavidis, N., & Tatem, A. J. (2022). How accurate are worldpop-global-unconstrained gridded population data at the cell-level?: A simulation analysis in urban Namibia. *PLoS One, 17*(7). https://doi.org/10.1371/journal.pone.0271504

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240. https://doi.org/10.2307/143141

Tremblay, J. C., & Ainslie, P. N. (2021). Global and country-level estimates of human population at high altitude. *Proceedings of the National Academy of Sciences of the United States of America, 118*(18), Article e2102463118. https://doi.org/10.1073/pnas.2102463118

Verma, D., Jana, A., & Ramamritham, K. (2019). Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat International, 88*, Article 101981. https://doi.org/10.1016/j.habitatint.2019.04.008

Wang, L. T., Wang, S. X., Zhou, Y., Liu, W. L., Hou, Y. F., Zhu, J. F., & Wang, F. T. (2018). Mapping population density in China between 1990 and 2010 using remote sensing. *Remote Sensing of Environment, 210*, 269–281. https://doi.org/10.1016/j.rse.2018.03.007

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., & Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences of the United States of America, 115*(14), 3529–3537. https://doi.org/10.1073/pnas.1715305115

Wei, J., Li, Z. Q., Cribb, M., Huang, W., Xue, W. H., Sun, L., & Song, Y. M. (2020). Improved 1 km resolution $PM_{2.5}$ estimates across China using enhanced space-time extremely randomized trees. *Atmospheric Chemistry and Physics, 20*(6), 3273–3289. https://doi.org/10.5194/acp-20-3273-2020

Wu, S. B., Chen, B., Webster, C., Xu, B., & Gong, P. (2023). Improved human greenspace exposure equality during 21st century urbanization. *Nature Communications, 14*(1), 6460. https://doi.org/10.1038/s41467-023-41620-z

Wu, B., Wang, Y., Huang, H. L., Liu, S. Y., & Yu, B. L. (2024). Potential of SDGSAT-1 nighttime light data in extracting urban main roads. *Remote Sensing of Environment, 315*, Article 114448. https://doi.org/10.1016/j.rse.2024.114448

Yeow, L. W., Low, R., Tan, Y. X., & Cheah, L. (2021). Point-of-Interest (POI) data validation methods: An urban case study. *ISPRS International Journal of Geo-Information, 10*(11), 735.

Zhang, L. F., Li, T. W., & Wu, J. A. (2022). Deriving gapless CO_2 concentrations using a geographically weighted neural network: China, 2014-2020. *International Journal of Applied Earth Observation and Geoinformation, 114*, Article 103063. https://doi.org/10.1016/j.jag.2022.103063

Zhang, W. M., Randall, M., Jensen, M. B., Brandt, M., Wang, Q., & Fensholt, R. (2021). Socio-economic and climatic changes lead to contrasting global urban vegetation trends. *Global Environmental Change-Human and Policy Dimensions, 71*, Article 102385. https://doi.org/10.1016/j.gloenvcha.2021.102385

Zhou, W., Ming, D. P., Lv, X. W., Zhou, K. Q., Bao, H. Q., & Hong, Z. L. (2020). SO-CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sensing of Environment, 236*, Article 111458. https://doi.org/10.1016/j.rse.2019.111458

Zhuang, H. M., Liu, X. P., Yan, Y. C., Ou, J. P., He, J. Y., & Wu, C. J. (2021). Mapping multi-temporal population distribution in China from 1985 to 2010 using landsat images via deep learning. *Remote Sensing, 13*(17), 3533. https://doi.org/10.3390/rs13173533