Remote Sensing of Environment 328 (2025) 114894

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Identification and evaluation of nighttime light pollution in residential gathering area of megacities based on SDGSAT-1 glimmer imagery☆

Yixuan Wu [a,b,1], Chenhao Huang [a,b,1,2,*], Yang Ye [c], Linlu Mei [d,e], Yalan Liu [d,e], Dacheng Wang [d,e], Weirong Chen [a,b], Jinsong Deng [a,b,2,**]

[a] *State Key Laboratory of Soil Pollution Control and Safety, Zhejiang University, Hangzhou 310058, Zhejiang, China*
[b] *College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China*
[c] *Institute of Spatial Information for City Brain (ISICA), Hangzhou City University, Hangzhou 310015, Zhejiang, China*
[d] *International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China*
[e] *Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China*

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
SDGSAT-1 GLI
Megacities
Nighttime light pollution
Population downscaling
Mismatch index
Refined urban planning
Sustainable Development Goals

ABSTRACT

Nighttime light pollution has become an increasingly serious issue in rapidly urbanizing megacities. It not only disrupts circadian rhythms and affects mental health, but also leads to energy waste and undermines the stability of urban and surrounding ecosystems, posing a significant threat to sustainable development. This study evaluated nighttime light pollution in the residential gathering areas of two typical megacities in China (Beijing and Shanghai) using 40-m SDGSAT-1 glimmer imagery (reflecting actual supply) and population grids (reflecting human demand) refined by the high-performance Random Forest model (with R^2 values of 0.93 for Beijing and 0.81 for Shanghai). By integrating urban functional zoning data to supplement the demand for nighttime lighting, a Nighttime Light Supply-Demand Mismatch Index (NLSDMI) was developed to quantify the imbalance of nighttime light between supply side and demand side. The results showed that Shanghai's nighttime light pollution area covered 78.25 km^2 (15.10 %), a higher proportion than Beijing's 115.61 km^2 (11.29 %) of the study area. Shanghai also exhibited higher peak NLSDMI values. In both cities, residential zones were among the primary contributors to nighttime light pollution. Additionally, in Beijing, the largest share was distributed in parks and green spaces, while in Shanghai, the second major distribution was found in industrial zones. The spatial patterns of nighttime light pollution reflected the distinct characteristics of the two megacities: Beijing focuses on cultural and administrative functions, while Shanghai tends to play its role as an economic hub. Accordingly, feasible countermeasures, including targeted lighting strategy formulation, urban land-use planning refinement and energy-saving lighting technology innovation, were proposed to mitigate light pollution and promote urban sustainability. This study demonstrated the promising potential of SDGSAT-1 glimmer imagery in advancing light pollution assessment and urban management. It also provides practical pathways toward the achievement of multiple Sustainable Development Goals (SDGs), especially SDG 3 (Good Health and Well-being), SDG 7 (Affordable and Clean Energy), and SDG 11 (Sustainable Cities and Communities). Future research should focus on enhancing data accuracy, improving validation methods, and exploring the applicability of findings to cities with diverse types and scales, thus providing broader theoretical support and practical guidance for global nighttime light pollution management.

☆ This article is part of a Special issue entitled: 'SDGSAT-1 for SDGs' published in Remote Sensing of Environment.
* Corresponding author at: College of Environmental and Resource Sciences, Zhejiang University, China.
** Corresponding author at: State Key Laboratory of Soil Pollution Control and Safety, Zhejiang University, China.

E-mail addresses: huangchenhao@zju.edu.cn (C. Huang), jsong_deng@zju.edu.cn (J. Deng).

[1] Indicates that both authors contributed equally to this research.
[2] Indicates co-corresponding authorship.

https://doi.org/10.1016/j.rse.2025.114894
Received 30 December 2024; Received in revised form 4 June 2025; Accepted 29 June 2025

1. Introduction

With the rapid urbanization, human economic and cultural activities have become increasingly frequent, and the urban populations have continued to grow. The widespread application of nighttime lighting has invigorated socio-economic development and greatly facilitated production and daily life. However, the intensity of nighttime lighting has risen sharply in recent years, with its spread becoming unplanned and surpassing actual needs (Huang et al., 2021; Niu et al., 2021; Stone, 2018). This uncontrolled trend has led to severe light pollution, which poses profound challenges to human health, energy efficiency, and ecological sustainability. The escalating severity of this issue makes it an urgent topic that needs attention and resolution. To tackle the series of global challenges confronting human society in the 21st century, the United Nations (UN) set forth 17 Sustainable Development Goals (SDGs) (UN, 2015). Many of them align with the challenges posed by light pollution, emphasizing the importance of implementing effective measures. Among them, SDG 7 (Affordable and Clean Energy) advocates for improving energy efficiency. However, disorderly nighttime lighting not only results in significant electricity waste but also increases carbon emissions, thereby hindering sustainable energy development (Gaston and Miguel, 2022). Additionally, SDG 3 (Good Health and Well-being) aims to ensure healthy lifestyles and promote the physical and mental health of all individuals. Urban nighttime light pollution has been shown to disrupt the human circadian rhythm, lead to sleep disorders and cardiovascular diseases, and even impair mental health (Levin et al., 2020). Moreover, SDG 11 (Sustainable Cities and Communities) emphasizes the importance of maintaining prosperity and resource balance during urban development, while striving to build sustainable communities. However, excessive nighttime lighting can disturb urban and surrounding ecosystems, diminishing the quality of life and the aesthetic experience of nighttime landscapes (Lis et al., 2024). Therefore, conducting in-depth research on urban nighttime light pollution, depicting its spatial characteristics, and proposing practical solutions can help to minimize unnecessary energy waste and health losses while meeting the human rational needs of production and daily life, and eventually achieve sustainable urban development (Barua et al., 2024; Tavares et al., 2021; Zielinska-Dabkowska and Bobkowska, 2022).

Satellite remote sensing technology, with its advantages of broad coverage, real-time monitoring, and high timeliness, has provided valuable spatiotemporal data for urban-related studies (Zhu et al., 2019). Unlike traditional daytime remote sensing products, nighttime satellite images provide a large-scale indication of Earth's illumination intensity, showing the potential to better assess light pollution (Bagheri et al., 2023; Ye et al., 2024). Especially for accurate urban monitoring, the demand for higher-precision nighttime light (NTL) data is becoming increasingly urgent. Historically, commonly used NTL data include DMSP-OLS and NPP-VIIRS. However, DMSP-OLS suffers from low spatiotemporal resolution, data saturation, and pixel blooming, all of which significantly degrade data quality and limit its applicability (Davies and Smyth, 2018). Although NPP-VIIRS data offers improved spatial resolution compared to DMSP-OLS, its resolution remains insufficient to meet the requirements for detailed urban monitoring. The Sustainable Development Science Satellite 1 (SDGSAT-1), launched in 2021, holds promise to address the limitations of existing NTL data and introduce new scientific advancements. As the first global scientific satellite dedicated to supporting the UN 2030 Agenda for Sustainable Development (2030 Agenda), it is equipped with three advanced sensors: a Thermal Infrared Spectrometer, a Glimmer Imager, and a Multispectral Imager, each designed to meet the scientific research needs of SDGs (Guo et al., 2022). Among them, the glimmer (GLI) imagery features an innovative design employing color bands and a panchromatic band. The spatial resolution of the color bands is 40 m, and that of the panchromatic band is 10 m. Compared to other NTL data, its high resolution and multiple bands provide the most precise range, intensity, and type of illumination ever collected, allowing for detailed characterization of human activity patterns. Currently, scholars have conducted extensive SDGSAT-1 GLI-based studies. Li et al., 2023a developed a dynamic village-scale demarcation method for built-up areas using SDGSAT-1 GLI data, providing a new perspective for fine feature extraction. Wu et al. (2024) proposed a method for extracting urban road networks from SDGSAT-1 GLI and validated it in multiple cities with complex road distributions. Liu et al. (2024b) proposed a novel index by integrating NTL intensity information from SDGSAT-1 GLI data and building volume information from Digital Surface Model (DSM) data to extract built-up areas more accurately. These research cases demonstrate the significant potential of SDGSAT-1 GLI data in urban-related studies.

Light pollution is intrinsically linked to human activities (Walker et al., 2020), making population data an integral part of the study. Given the relatively focused study area and the need to match NTL images, high-resolution population gridded data is essential for detailed spatial analyses. Traditional population data is typically derived from census, while numerous studies have transformed them into gridded formats, resulting in several well-known open datasets. For example, the Gridded Population of the World (GPW) (Tobler et al., 1997) provides global population maps through efficient simple computational methods but cannot capture details. Datasets such as LandScan (Dobson et al., 2000), Global Rural-Urban Mapping Project (GRUMP) (Balk et al., 2006), and WorldPop (Tatem, 2017) utilize data-intensive dasymetric mapping approaches, leveraging fine-scale spatial auxiliary data as inputs for modeling to capture population heterogeneity. However, in high-density urban areas, the allocation accuracy still faces challenges (Xu et al., 2024). In this context, studies have emerged that use auxiliary data to downscale existing datasets. Ye et al. (2019) utilized a Random Forest (RF) model, integrating remote sensing images and points of interest (POI), to downscale county-level census data into 100 × 100 m, revealing a population map with higher accuracy than the WorldPop. Leveraging SDGSAT-1, Liu et al., 2023a refined the WorldPop population distribution to a 10-m resolution. Lei et al. (2024) applied a multi-scale geographically weighted regression model incorporating building footprints, NTL and POI to allocate county-level population to 100-m raster.

Due to the unordered sprawl of nighttime lighting caused by the accelerated urbanization, there has been an increasing number of studies on adopting NTL imagery to monitor and evaluate the nighttime lighting environment. Kuechly et al. (2012) explored the relationship between NTL intensity and land use, revealing that road-related upward light sources accounted for 31.6 % of the sources of light pollution. Tang et al. (Tang et al., 2020) combined night view photos of Changsha with POI data to analyze the spatial patterns of lighting in the urban area, identifying that commercial aggregation was the dominant source of light pollution. These studies focus on the sources of light pollution across different functional areas but have not explored the differences in how functional zones contribute to the generation and distribution of light pollution. Ye et al. (2020) integrated NPP-VIIRS data with population density to evaluate the balance between supply and demand for NTL from a human-centered perspective. Zhao et al. (2021) utilized NTL images and POI data to assess light pollution by examining its impact on urban residential environments. These studies have addressed the impact of nighttime lighting on human well-being. However, these existing studies have not comprehensively and finely revealed the spatial heterogeneity of the distribution and level of light pollution from the perspective of light supply and demand, thereby supporting increasingly sophisticated urban planning and lighting management.

Given the above background, this study focused on typical megacities in China, which is characterized by high economic prosperity and population density (The State Council of China, 2014), as case study areas. By utilizing high-resolution data exemplified by SDGSAT-1 GLI imagery and incorporating multi-source auxiliary data such as urban functional zones, the research aimed to quantifying the mismatch between supply and demand of urban nighttime lighting. Overall, this study developed a Nighttime Light Supply-Demand Mismatch Index (NLSDMI) to identify zones and evaluate levels of nighttime light pollution. Specifically, the phased targets were: (1) to obtain high-quality SDGSAT-1 NTL imagery; (2) to generate reliable 40-m population grids; and (3) to construct NLSDMI to analyze the spatial supply-demand relationship of NTL, and propose recommendations to boost SDGs such as SDG 3, 7, and 13.

In response to the current research gap, the main contributions of this study include:

(1) **Detailed characterization of NTL supply patterns based on new satellite imagery:** Leveraging high resolution SDGSAT-1 GLI data to comprehensively depict actual NTL supply patterns at a 40-m fine scale;
(2) **Refined exploration of NTL demand patterns with AI-driven model and multi-source data:** Utilizing 40-m resolution population grid data via RF-based downscale and urban functional zoning data to analyze NTL demand of residential gathering area in megacities, revealing the spatial distribution of actual regional lighting needs;
(3) **Developed a refined remote-sensing index for assessing urban nighttime light pollution:** Constructing a spatially refined index to evaluate the imbalance between NTL supply and demand, providing scientific supports for light pollution management and optimizing urban lighting environments.

The rest of this paper is structured as follows: Section 2 introduces the study areas and data; Section 3 describes the methodology; Section 4 analyzes the results of nighttime light pollution assessment; Section 5 discusses the findings; and Section 6 summarizes the conclusions.

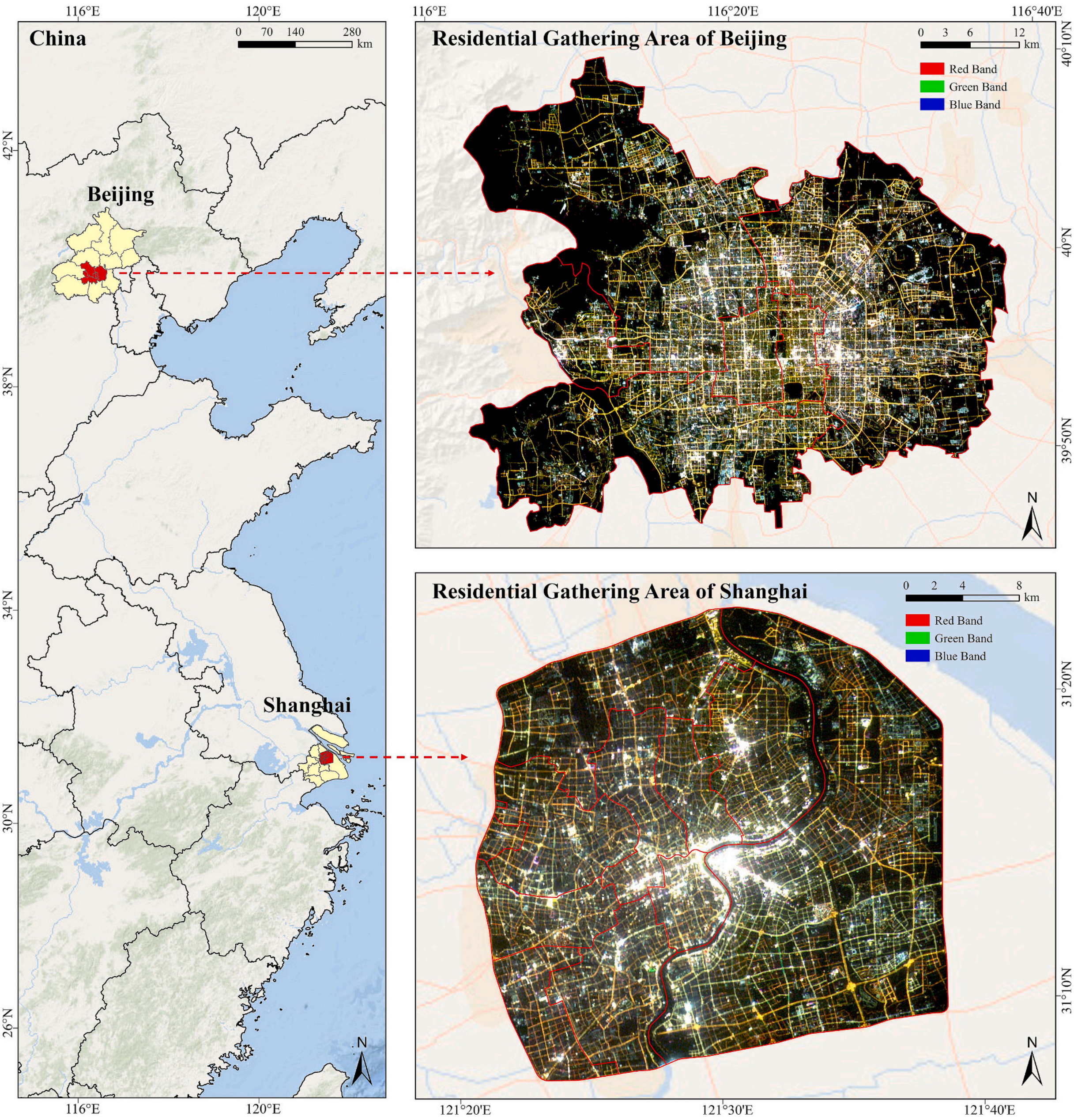

Fig. 1. Study areas and corresponding SDGSAT-1 GLI RGB imagery.

2. Study area and data

2.1. Study area

This study focused on the residential gathering areas of Beijing and Shanghai in China (Fig. 1). These two cities were selected because they are the most representative megacities in mainland China, characterized by high levels of urbanization and vibrant economic activities, ranking among the top global megacities. According to Globalization & World Cities (GaWC), Beijing and Shanghai have been classified as Alpha+ cities since 2008 and have maintained this status (https://gawc.lboro.ac.uk/gawc-worlds/the-world-according-to-gawc/). They are also the only two megacities in mainland China at this level. In addition, the Global Cities Index (GCI) rankings show that Beijing has consistently ranked in the top ten over the past five years, while Shanghai has remained in the top twenty (https://www.kearney.cn/article/-/insights/303836250). As the capital city located in the northern part of the North China Plain, Beijing serves as China's political and cultural hub, with a GDP of 4.98 trillion CNY and a population of 21 million in 2024 (https://tjj.beijing.gov.cn/tjsj_31433/tjgb_31445/ndgb_31446/202503/t20250319_4038820.html). Positioned at the Yangtze River estuary, Shanghai is regarded as the most significant economic, financial and shipping center of China, with a GDP of 5.39 trillion CNY and a population of 25 million in 2024 (https://tjj.sh.gov.cn/tjgb/20250324/a7fe18c6d5c24d66bfca89c5bb4cdcfb.html).

As representative examples of urbanization in China, the two cities have witnessed a significant rise in NTL intensity and a rapid expansion in NTL range over recent years, exacerbating nighttime light pollution. This emerging issue negatively impacts life quality of residents and significantly impedes energy efficiency and urban sustainability. Residential gathering area with the highest population density and urbanization levels, usually serves as a key pivot for commerce, culture and transportation, playing a core role in supporting urban development. Therefore, selecting the residential gathering area of these two megacities as representative cases provides valuable insights into the typical characteristics of urban nighttime light pollution and responding strategies for its mitigation.

In this study, the residential gathering area of Beijing is defined according to the "Beijing City Master Plan (2016-2035)", covering 1387 km^2 and including the districts of Dongcheng, Xicheng, Haidian, Fengtai, Chaoyang, and Shijingshan. And the residential gathering area of Shanghai is defined within the outer ring road based on the outer ring expressway from OpenStreetMap road data (Ta et al., 2020; Yue et al., 2021), which covers 662 km^2, including the districts of Huangpu, Hongkou, Jing'an, and parts of Pudong New District, Minhang District, Baoshan District, and Jiading District.

2.2. Data and preprocessing

2.2.1. SDGSAT-1 GLI imagery

The SDGSAT-1 GLI data used in this study comes from SDGSAT-1 developed and operated by International Research Center of Big Data for Sustainable Development Goals (CBAS). Through the SDGSAT-1 Open Science Program initiated by CBAS, researchers around the world can download SDGSAT-1 data by submitting research proposals online (https://www.sdgsat.ac.cn/). The satellite operates at an orbital altitude of approximately 505 km with a swath width of 300 km. The NTL images obtained by the Glimmer Imager on board SDGSAT-1 contain four bands: a panchromatic band with a spatial resolution of 10 m and the RGB bands with a spatial resolution of 40 m. From the perspective of data quality, the SDGSAT-1 GLI images adopted in this study are from the RGB bands rather than the panchromatic band. Although the panchromatic images have higher spatial resolution, they have more obvious stripes and salt-and-pepper noise, requiring more complex preprocessing (currently there is no mature method system); on the contrary, RGB data has less noise and richer information, and is more suitable for direct applications in large-scale light pollution modeling (Wang et al., 2024).

In order to obtain higher-quality images initially and reduce the complexity of preprocessing, weather forecasts and flight schedules of Beijing Capital International Airport and Shanghai Hongqiao International Airport were referred to, and clear days with no clouds or minor clouds were selected for image acquisition. The specific acquired dates and product IDs of screened SDGSAT-1 GLI data are shown in the Table 1.

To ensure the usability and accuracy of the SDGSAT-1 GLI data, this study conducted necessary preprocessing, which involved three steps:

1. **Potential noise removal:** The SDGSAT-1 RGB GLI imagery consists of three bands, where each band reflects light information in a standard RGB image. Visual inspection and statistical analysis of pixel values showed that noise points exhibit abnormal patterns across the bands. Specifically, noise pixels display the minimum value in at least one band, while other bands show values greater than the minimum. Based on this observation, potential noise points were identified as those with minimum values in at least one band and higher-than-minimum values in other bands. The potential noise was extracted from each band, and noise points were removed accordingly.
2. **Radiometric calibration:** To obtain physically meaningful NTL radiance information, the denoised data from each band were first converted into the NTL radiance values. The calibration parameters required for converting the SDGSAT-1 GLI data into NTL radiance values can be found in the header file of the original SDGSAT-1 GLI data. The file provides detailed calibration parameters for each band of the Glimmer Image for Urbanization, with the calibration coefficients last updated on March 1, 2022. The specific conversion formula is as follows (Liu et al., 2024a):

$$L = DN \times GAIN + BIAS \tag{1}$$

Where L represents the radiance value of the SDGSAT-1 GLI image, with units of W·m^{-2}·sr^{-1}·μm^{-1}, DN denotes the raw digital number of each band, $GAIN$ and $BIAS$ are the calibration parameters officially provided.

Next, calculate the total radiance values for quantifying light pollution based on the light intensity contribution of the RGB three-color bands. The grayscale brightness of SDGSAT-1 GLI imagery is calculated as follows (Grundland and Dodgson, 2007):

$$B = 0.2989 \times L_{Red} + 0.5870 \times L_{Green} + 0.1140 \times L_{Blue} \tag{2}$$

Where B represents the grayscale brightness of the SDGSAT-1 GLI data, with units of W·m^{-2}·sr^{-1}·μm^{-1}, L_{Red}, L_{Green}, L_{Blue} denote the radiance values of the red, green, and blue bands, respectively.

3. **Geometric correction:** After comparing with the high-resolution standard color remote sensing imagery base map, a slight misalignment of roads was observed. To improve data quality, control points were used for image registration, allowing for more accurate application of SDGSAT-1 GLI imagery.

2.2.2. Other geospatial and statistical data

The population data used in this study was sourced from the dataset the University of Southampton shared on the WorldPop platform (https://hub.worldpop.org/geodata/summary?id=49919). The latest 2020 WorldPop population data with a 100-m resolution was downloaded for analysis. The dataset consists of six types, and this study focused on the data that has been adjusted to align with the national population totals reported by the UN. Estimates were made only for areas mapped as containing built-up settlements. The population data for each district was adjusted based on the permanent resident population figures from the "2021 Statistical Yearbook" and then resampled from the original 100-m resolution to a 40-m resolution.

Table 1
SDGSAT-1 GLI data.

Municipality	Acquired Date	Product ID
Beijing, China	2022-01-03	KX10_GIU_20220103_E116.94_N39.61_202200111720_L4A
Shanghai, China	2022-04-10	KX10_GIU_20220410_E119.90_N31.84_202200092994_L4A

POIs are geographical points that represent important locations for human activities. Integrating POI data with multi-source remote sensing data can significantly enhance the accuracy of population distribution mapping (Guo et al., 2023b). The POI data used in this study were sourced from Amap, and six types of POI records closely related to the population in 2020 and 2022 were selected, including food, shopping, transportation facilities, residential areas, life services, and recreation. Since the spatial distribution of each POI category differs, the six POI data types were processed separately. The residential POI category directly reflected population distribution, and thus, the point-to-raster method was employed to convert POI data into raster grids. The remaining five categories, service-oriented POIs, had a diminishing spatial influence with distance, so the euclidean distance was calculated to assess their spatial impact (Li et al., 2023b). This results in six corresponding spatial variables, each with a resolution of 40 m.

The road density data was derived from OpenStreetMap (www.openstreetmap.org). Road density was calculated at a spatial resolution of 40 m using the line density tool in ArcGIS.

The building density data was derived from building footprint and height data for 77 cities nationwide in 2019, obtained from Baidu Map. Since building footprints and heights were unlikely to change significantly over time, this dataset was used as a reference. The calculation method is as follows: a 40 × 40 m fishnet was created for the study area, and the proportion of each grid cell covered by buildings was calculated. The grid data was then converted into GeoTIFF format, yielding building density with a spatial resolution of 40 m.

The Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), and Normalized Difference Water Index (NDWI) for 2020 and 2022 were derived from Landsat 8 satellite imagery available on Google Earth Engine (GEE). The satellite launched in 2013 has a spatial resolution of 30 m. The data used in this study was identified by the ID LANDSAT/LC08/C02/T1_L2. These three auxiliary variables were resampled to a spatial resolution of 40 m.

Digital Elevation Model (DEM) data comes from the Copernicus Digital Elevation Model (COP-DEM), a global 30-m resolution dataset released by the European Space Agency (ESA). This data was then resampled to 40 m to align with the data scale.

The urban functional zoning data was derived from the 2018 dataset of land use types in Chinese cities (Gong et al., 2020), which includes 5 main categories and 12 subcategories. This dataset was created using 10-m satellite imagery from Sentinel-2 A/B in 2018, OpenStreetMap, NTL data (Luojia-1), POI data from Amap (including categories and quantities), and Tencent social location data as input features. In this study, the data was validated and corrected using the updated POI data, resulting in the revised urban functional zoning data.

A summary introduction of all data used in this study are listed in Table 2.

Table 2
Data list.

Data Name	Description	Source	Purpose
SDGSAT-1 GLI imagery	SDGSAT-1 GLI RGB band images (40-m resolution)	International Research Center of Big Data for Sustainable Development Goals	Measuring the actual supply of NTL
WorldPop population	Constrained Individual countries 2020 UN adjusted (100-m resolution)	Online database	Measuring the actual demand for NTL
Road density	Road density raster calculated using Line Density tool based on OSM road network	OpenStreetMap	WorldPop population raster downscaling
Building density	Building density raster calculated using the Point Density tool based on building footprint data	Baidu Map	
NDBI NDVI NDWI	The remote-sensing index calculated on the GEE platform	Landsat 8 satellite from United States Geological Survey (available on GEE)	
DEM	The global 30-m Copernicus Digital Elevation Model (COP-DEM)	European Space Agency	
POI_food POI_shop POI_transport POI_recreation POI_service	POI raster calculated using the Euclidean Distance tool	Amap	
POI_ residential	POI raster produced using the Point-to-Raster tool		
Demographic data	Statistical data of permanent population in each district in 2022 from the Statistical Yearbook of the corresponding municipalities published in 2023	Government open data	Adjusting population estimation
Urban functional zoning data	Verified and corrected urban functional zoning data based on updated POI data	Existing studies	Measuring the actual demand for NTL

3. Methods

The overall methodological framework applied in this study is shown in Fig. 2. The first component corresponds to Section 2.2 (Data and Preprocessing), which describes the preprocessing of SDGSAT-1 GLI data and other auxiliary datasets. The subsequent two components constitute the core of this framework. The second component corresponds to Section 3.1 (Population downscaling estimation), where the procedures for deriving estimated high-resolution population raster datasets based on the RF model are elaborated. The third component corresponds to Section 3.2 (Nighttime light pollution assessment), detailing the construction of the NLSDMI. After completing the above analysis process, this study proposed policy recommendations based on the research results obtained, which were discussed in Section 5.4.

3.1. *Population downscaling estimation*

The study area consisted of the residential gathering areas of two typical megacities, where there are dense artificial light sources, showing the characteristics of high dynamic range and complex color temperature (Yang et al., 2021). To accurately capture internal diversity of nighttime lighting and match SDGSAT-1 NTL data, higher-resolution population data is necessary to support refined analysis. However, the

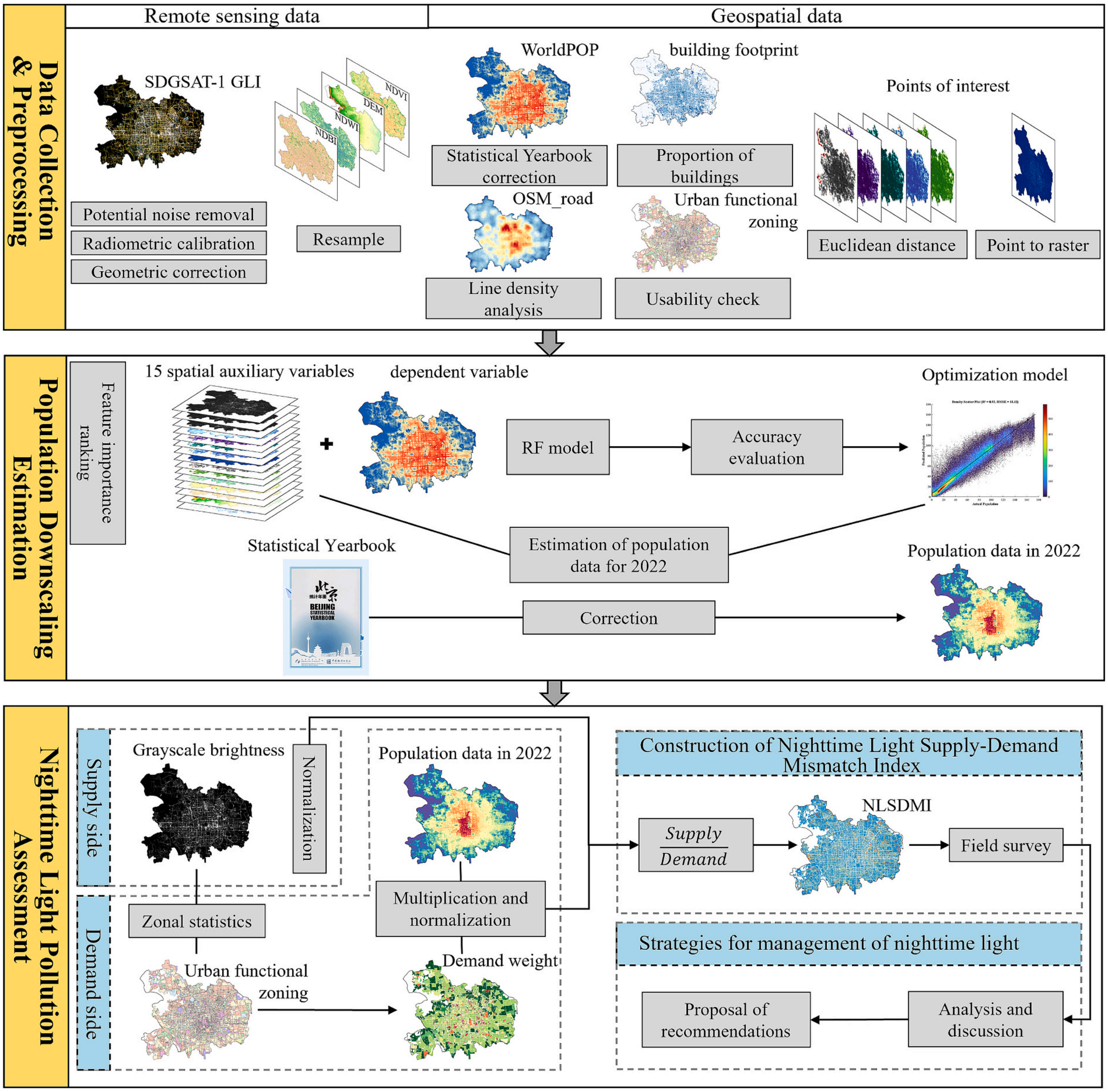

Fig. 2. Overall methodological framework.

population grid data currently in common use mainly come from WorldPop and LandScan. The former has a spatial resolution of 100 m but is currently only updated to 2020, while the latter meets the time validity of year 2022 but has a spatial resolution of only 1 km. Therefore, it is necessary to adjust the existing population grid to achieve time-effective and spatial-detailed research.

To achieve this, this study employed RF regression model for population downscaling estimation, following these three steps:

(1) **Model Construction:** A nonlinear regression relationship was established based on the corrected 2020 WorldPop population grid at 40 m resolution, combined with other 15 spatial auxiliary variables with the same resolution, to develop a parameter-optimal RF model;
(2) **Estimation of Population Data:** The trained RF model was applied to 2022 auxiliary variables at 40 m resolution to estimate the population distribution for 2022;
(3) **Regional Adjustment:** The estimated population data was further refined using official demographic data to improve accuracy.

3.1.1. *Principle of RF regression model*

This study employed the RF model for regression tasks. RF was an ensemble machine learning algorithm based on decision trees, first proposed by Breiman (2001). It enhanced the robustness and accuracy of predictions by constructing multiple decision trees and aggregating their outputs through methods such as averaging or majority voting. The main processing steps were as follows: (1) Randomly sampling multiple training set from the sample data using the bootstrap method; (2) Constructing independent decision trees for each training set; (3) Randomly selecting a subset of features at each split node to determine the best split; (4) Aggregating the predictions of all decision trees to produce the final output. This structure provided RF with several notable advantages, such as high prediction accuracy, robustness to outliers, and effective prevention of overfitting. RF could be applied to classification task or regression task and effectively model nonlinear relationships between the dependent variable and multiple independent variables. It had been successfully utilized in downscaling studies and exhibited overall robust performance (Yang et al., 2020; Zhang et al., 2022).

3.1.2. *Construction of auxiliary variable database*

This study employed the 2020 population gridded data with a 40-m resolution after resampled from the 100-m WorldPop raster and calibrated it using demographic data from statistical yearbooks, as one of the key dependent variables for population downscaling estimation. The purpose of this correction is to reduce the resampling bias of the original WorldPop data in characterizing the population distribution at smaller spatial scales, thereby providing more reliable support for subsequent modeling.

In selecting other auxiliary variables, this study focused on acquiring all possible spatial information closely related to population distribution. Furthermore, the feature importance evaluation in the RF model was utilized to identify key auxiliary variables and improve the efficiency and interpretability of the model. Specifically, this process involved testing different combinations of variables and conducts repeated training to evaluate the contribution of each variable to the model performance, then assess its explanatory power for the spatial distribution of the population. After eliminating redundant or irrelevant variables with small contributions through the above feature engineering process, this study screened out 15 key variables as auxiliary factors for population downscaling estimation. The auxiliary variables include: (1) Satellite optical bands: SDGSAT-1 GLI Red Band, Green Band, and Blue Band; (2) Urban built environment factors: Road density, Building density, and DEM; (3) Remote sensing indices: NDBI, NDVI, and NDWI; (4) POI-drived factors: POI_food, POI_shop, POI_transport, POI_residential, POI_service, and POI_recreation. The visualized groupings of the variables were presented in the supplementary material (Fig. S1 and Fig. S2).

3.1.3. *Execution of RF regression model*

This study employed MATLAB (Version: R2024a) to build parameter-optimal RF regression models. The specific steps are outlined as follows:

(1) **Data standardization:** All 40-m raster data were standardized to the World Geodetic System 1984 geographic coordinate system, ensuring consistent dimensions and grid alignment. And the dependent and independent variables in 2020 were imported into MATLAB.

(2) **Model training and prediction:** Then the regression relationship between the 40-m population data and other auxiliary variables in 2020 was modeled using the RF. The optimal parameters of the model were determined through iterative optimization using grid search and 5-fold cross-validation. Expressly, the number of decision trees was set to 50–1000, the maximum depth of a single tree is set to 0–20, the minimum number of samples required for node splitting was set to 2–20, and the minimum number of samples in each leaf node was set to 1–10 to enhance the stability and predictive accuracy of the model. In addition, the split ratios of the training set and validation set were 70 % and 30 % respectively. Out-of-bag (OOB) prediction was enabled to evaluate model performance, eliminating the need for a separate test dataset.

(3) **Feature importance analysis:** After model training, feature importance was then calculated based on the OOB data to assess the contribution of each auxiliary variable to the target dependent variable. Based on the variable ranking, significantly redundant variables were eliminated, and steps (1)–(3) were repeated until the model performance reached the optimal level.

(4) **Model accuracy evaluation:** The optimal model was applied to obtain the fitted values of the 40-m resolution population data in 2020 and then compared with the actual values. The performance of the model was reflected by the Coefficient of Determination (R^2) (Hutengs and Vohland, 2016), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) (Hodson, 2022), and Mean Squared Logarithmic Error (MSLE) calculated (Abdelrahim and Yücel, 2025). The formulas for these evaluation metrics are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{5}$$

$$MSLE = \frac{1}{n}\sum_{i=1}^{n} [log(y_i + 1) - log(\widehat{y}_i + 1)]^2 \tag{6}$$

Where y_i represents the actual value, $\widehat{y}_i$ represents the predicted value, and $\overline{y}$ represents the mean of the actual values. i denotes the i-th spatial pixel, n is the total number of spatial pixels.

3.1.4. *40-m population girds estimation*

This study assumed that the regression relationship for the same regional data can be transferred and remain generally consistent between similar years, which has been supported by several studies (Knibbe et al., 2014; Masselot et al., 2018; Wu et al., 2025). Specifically, the relationship between the population raster and auxiliary variables in 2020 captured by RF was applied to estimate population raster in 2022. As such, the model parameters trained on 2020 data were also applied to the 2022 auxiliary variables.

To ensure that the estimated population data is more accurate and aligns with reality, this study conducted a district-level adjustment of the RF-estimated population based on the 2022 resident population figures for each district from the 2023 Statistical Yearbook of each megacity. The specific correction formula is as follows:

$$A_{corrected,i,j} = A_{ij} \times \frac{P_i}{\sum A_{ij}} \tag{7}$$

Where A_{ij} represents the population number on the j-th grid cell of the i-th district after downscaling; P_i represents the resident population of the i-th district (from the statistical yearbook); $\sum_j A_{ij}$ represents the total downscaled population for all grid cells in the i-th district, used to calculate the correction ratio; $A_{corrected,i,j}$ represents the corrected population for the j-th grid cell in the i-th district.

3.2. *Nighttime light pollution assessment*

From the dual perspectives of supply and demand, the mismatch between nighttime lighting and human activities was quantified by establishing a gridded index NLSDMI. Then the areas affected by light pollution was indenfied where NTL exceeded actual human demand. Specifically, areas where NLSDMI greater than a certain threshold were classified as light-polluted, while areas where NLSDMI less than or equal to the threshold were classified as non-polluted (the threshold was dynamically determined based on the distribution of NLSDMI in different regions). In the identified light pollution areas, the higher the NLSDMI, the higher the mismatch degree and the more prominent the pollution problem.

3.2.1. *Supply-side NTL intensity quantification*

The supply-side NTL intensity was simply represented by actual lighting intensity derived from SDGSAT-1 GLI images. The radiance values of the three bands (RGB) were synthesized and converted into grayscale brightness values according to Section 2.2.1. These brightness values were then normalized using min-max normalization, as shown in

Eq. (8).

$$S_{normal} = \frac{B - B_{min}}{B_{max} - B_{min}} \tag{8}$$

Where S_{normal} represents the normalized actual NTL intensity, B represents the grayscale brightness derived from the RGB data as calculated before, measured in W·m^{-2}·sr^{-1}·μm^{-1}.

3.2.2. *Demand-side NTL intensity quantification*

The demand-side NTL intensity primarily reflects human rational demand for nighttime illumination. Since the residential gathering areas of megacities often consist of multiple functional zones, the demand for NTL varies across different zones and cannot be directly represented by population density alone. Therefore, this study introduced a functional zone type-specific demand weight system to quantify the differences in NTL demand.

Hence, the demand weights were determined by both urban functional zoning and actual lighting intensity. Since the original urban functional zoning data collected excluded road areas, roads were not considered in demand weight determination. The specific steps for calculating the weight were as follows: (1) The Zonal Statistics as Table tool in ArcGIS was used to calculate the number of grid cells and the total grayscale brightness value for each of the 11 secondary functional categories (Residential, Business office, Commercial service, Industrial, Transportation stations, Airport facilities, Administrative, Educational, Medical, Sports and cultural, Park and green space); (2) The raster data used in this study had a resolution of 40 m, meaning that each grid cell covered an area of 1600 m^2. Based on this, the total area of each functional zone was calculated; (3) The total grayscale brightness value of each functional zone was then divided by its total area to obtain the unit-area NTL intensity, which was taken as the demand weight for each functional zone as shown in Eq. (9).

$$W_i = \frac{\sum_{j=1}^{n_i} B_{ij}}{\sum_{j=1}^{n_i} A_{ij}} \tag{9}$$

Where W_i represents the weight of the i-th urban functional zone, B_{ij} represents the grayscale brightness of the j-th grid cell in the i-th urban functional zone, A_{ij} represents the area of the j-th grid cell in the i-th urban functional zone. n_i represents the total number of grid cells in the i-th urban functional zone. The demand weights for each functional zone in the two megacities calculated are shown in Table 3.

The demand-side NTL intensity was calculated as the product of population and demand weight, as shown in Eq. (10). The demand-side NTL intensity was normalized using min-max normalization, as shown in Eq. (11).

$$D_j = P_j \times W_i \tag{10}$$

$$D_{normal} = \frac{D_j - D_{min}}{D_{max} - D_{min}} \tag{11}$$

Where D_j represents the demand intensity of the j-th grid cell, P_j represents the number of individuals in the j-th grid cell, W_i represents the weight of the i-th urban functional zone, D_{normal} represents the normalized demand light intensity.

Table 3
NTL demand weights of Beijing and Shanghai.

Zonal Function	Beijing	Shanghai
Industrial	0.07	0.09
Park and green space	0.06	0.10
Airport facilities	0.06	0.07
Transportation stations	0.18	0.19
Educational	0.11	0.12
Residential	0.11	0.11
Business office	0.29	0.29
Commercial service	0.12	0.23
Sport and cultural	0.16	0.23
Administrative	0.14	0.11
Medical	0.15	0.14

3.2.3. *NLSDMI establishment and light pollution area identification*

Finally, the normalized supply-side NTL intensity was divided by the normalized demand-side NTL intensity to obtain NLSDMI, as shown in Eq. (12).

$$NLSDMI = \frac{S_{normal}}{D_{normal}} \tag{12}$$

Where S_{normal} represents the normalized actual NTL intensity, D_{normal} represents the normalized demand NTL intensity, and *NLSDMI* represents the nighttime light supply-demand mismatch index. Visual representations of the data used in constructing the index were included in the supplementary material (Fig. S3 and Fig. S4). Furthermore, the derived NLSDWI was spatially analyzed using Global and Local Moran's I to intuitively reveal its distribution pattern.

The NLSDWI developed in this study is a continuous spatial metric designed to quantify the degree of mismatch between NTL and actual human demand, which are positively associated with the risk of light pollution. According to a widely accepted definition, "Light pollution is the presence of any unwanted, inappropriate, or excessive artificial lighting" (Smith et al., 2023). Therefore, it is necessary to quantitatively determine the extent of affected areas to provide scientific references for light pollution management. In this study, thresholds were set as the mean plus one standard deviation of the log-transformed NLSDMI values. Areas with NLSDMI exceeding this threshold were classified as light-polluted, while areas with values less than or equal to the threshold were classified as non-polluted.

Specifically, the threshold determination involved three steps: (1) The original NLSDMI distribution was right-skewed (long right tail), so a base-10 logarithmic transformation ($\log_{10}$) was applied, resulting in an approximately normal distribution of the transformed values. (2) According to the empirical rule of normal distribution, about 68 % of values lie within the range of mean ± one standard deviation. Values exceeding the mean plus one standard deviation were considered significantly higher than average or anomalously elevated, which served as a robust and interpretable criterion to identify light-polluted areas. (3) Since the threshold was determined in the log-transformed domain, it was finally converted back to the original scale using the exponential function.

3.2.4. *Field survey*

To validate the accuracy of identifying light pollution areas derived from multi-source data in this study, field surveys were conducted in both Beijing and Shanghai. Given the substantial investment of manpower, material, and resources for field surveys, a representative sampling strategy was adopted. Specifically, within various functional zones of each city, the central points of pixel with relatively high (top 5 %) or low (bottom 5 %) NLSDMI value were randomly selected as survey points, and their geographic coordinates (latitude and longitude) were recorded. The subsequent point screening process comprehensively compared long-term remote-sensing images (such as NPP-VIIRS and high-resolution visible light images), land use maps, and urban planning documents to ensure that the selected locations were stable in both spatial and functional attributes.

The field surveys were carried out on clear nights with minimal cloud cover and relatively stable air quality conditions, specifically during the peak period of artificial lighting (20:00–22:00, UTC + 8). At each selected location, the following investigation procedures were undertaken: (1) the nighttime environment was documented through on-site photography; (2) illuminance levels (in LUX) were measured using three TA636A digital light meters, and the average value was calculated to ensure measurement reliability; and (3) the number of passing vehicles and pedestrians was recorded over a standardized 5-min observation period. These measurements were subsequently used to estimate the local nighttime light pollution level (expressed as a relative value, i.e., the average illuminance measured by the three instruments divided by the sum of the number of vehicles and pedestrians).

4. Results

4.1. Quality evaluation of SDGSAT-1 GLI imagery

4.1.1. Denoising result

In this study, potential noise was identified through visual inspection combined with pixel-based statistical analysis. Taking the local Red band of SDGSAT-1 GLI in Beijing as an example, the denoised result obtained were shown in Fig. 3. It can be observed that the potential noise identified was mainly distributed near the road network, especially in the intersection of roads where human activities are complex and lighting facilities are dense. This is because the dynamic light sources represented by vehicle lights in these areas are easily captured by the sensor as abnormal values, and the spectral superposition of multiple static light sources may also cause interference in the sensor's recognition. Moreover, the denoising method proposed effectively removed salt-and-pepper noise and saturation overflow artifacts in most urban blocks. Image distortions were significantly reduced and clarity was improved as rough and blurry areas were eliminated.

4.1.2. Spatial distribution analysis of SDGSAT-1 NTL intensity

The SDGSAT-1 GLI images of the residential gathering areas of the two megacities obtained by preprocessing according to the method in Section 2.2.1 are shown in Figs. 4 and 5. It can be observed that the NTL was primarily concentrated in the central regions (such as the commercial districts within Third Ring Road of Beijing and the Bund and Lujiazui along both banks of the Huangpu River in Shanghai), with relatively sparser distribution in the surrounding areas, though scattered small-scale clusters were also present. It's worth noting that the road network was visible due to high spatial resolution of SDGSAT-1, reflecting the spatial characteristics of urban transportation of these megacities: Beijing's transportation network presents a mixed structure of ring, radial and chessboard shapes, while Shanghai's road network presents the characteristics of ring and grid. Additionally, when considering the separate radiation values for each RGB band, it is

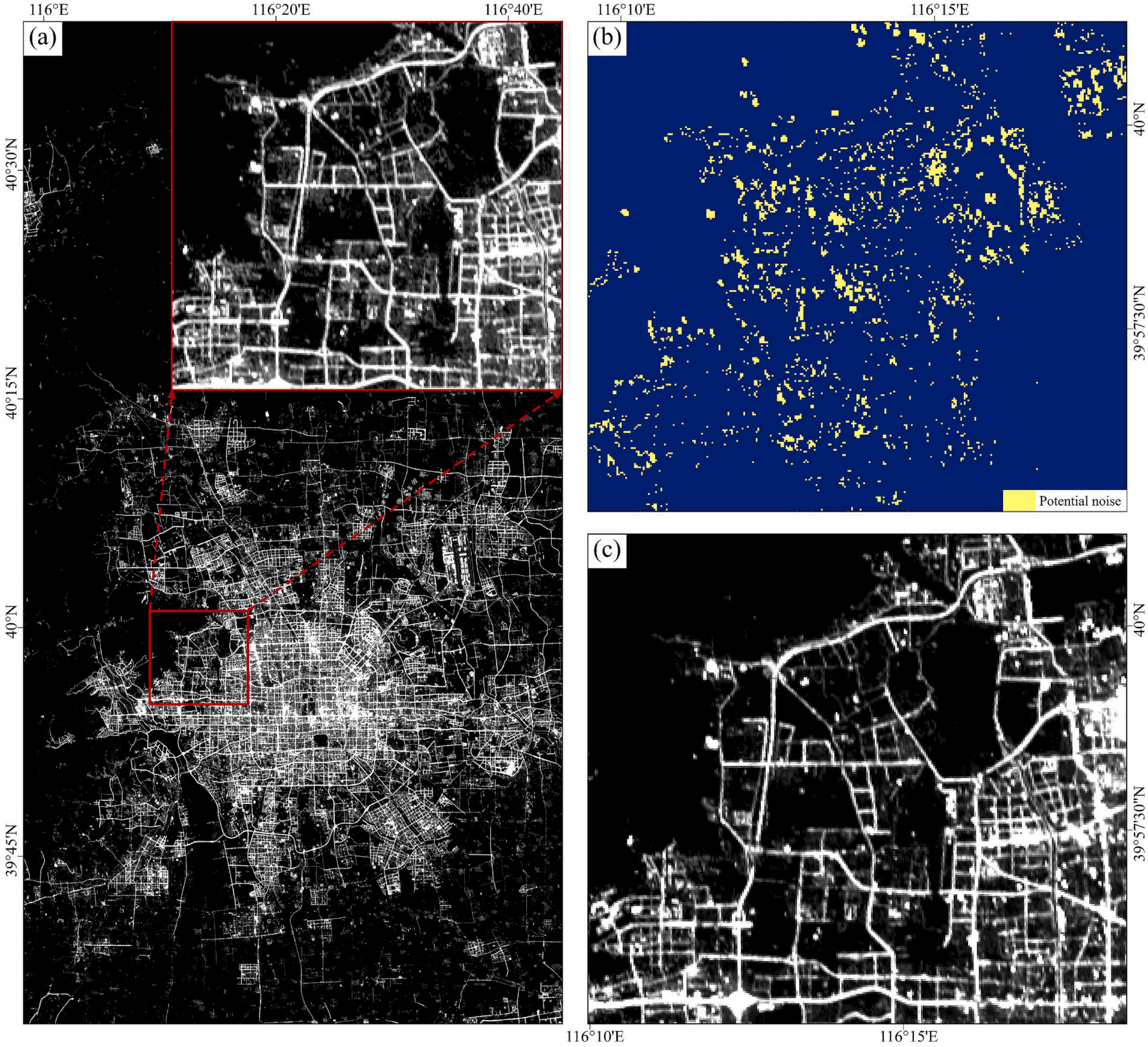

Fig. 3. (a) Red band GLI image of Beijing with noise; (b) Results of potential noise identification; (c) Partial image after potential noise removal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

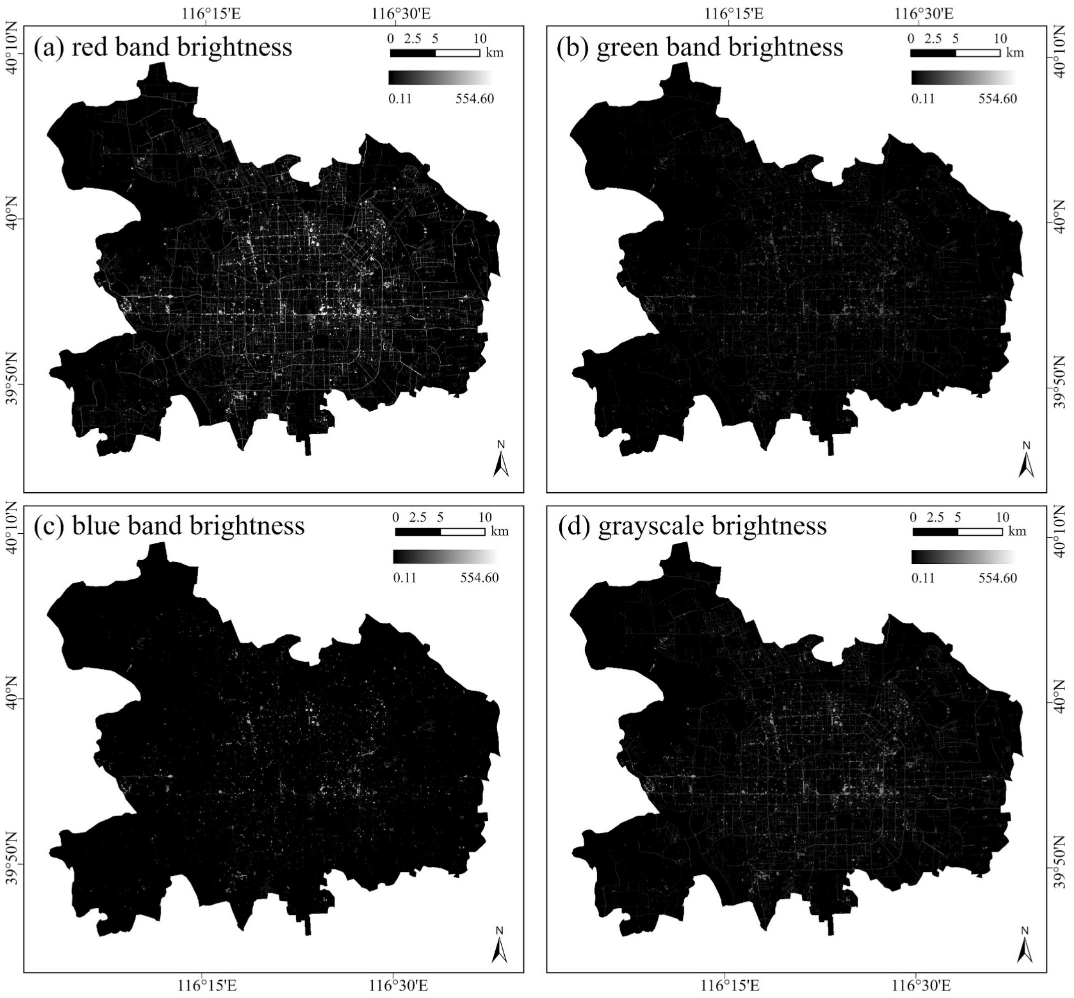

Fig. 4. NTL radiance values (W·m^{-2}·sr^{-1}·cm^{-1}) of SDGSAT-1 GLI images for the residential gathering area of Beijing: (a)-(d) represent the red, green, blue bands and grayscale brightness, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

evident that the radiation in the Blue Band was lower compared to the Red Band and Green Band, particularly near the road areas. This is because the current nighttime artificial lighting sources Beijing and Shanghai are mainly Light Emitting Diode (LED) that emit warm yellow light and white light, while cool blue light has been restricted in layout due to the need to avoid its damage to human eyes (B. Guo et al., 2023a).

4.2. *Results of population downscaling estimation*

4.2.1. *Performance evaluation of RF models*

This study employed the RF model to establish the regression relationship between population data and auxiliary variables. Model performance was assessed using R^2、RMSE, MAE and MSLE. Table 4. presents the accuracy evaluation results of parameter-optimal RF models: The RF trained in the residential gathering area of Beijing achieved an R^2 of 0.93, demonstrating an excellent fitting relationship. Additionally, the RMSE and MAE were 11.11 and 7.60, respectively, indicating minimal error between predicted and actual values. Furthermore, the MSLE was 0.30, suggesting that the model maintained high predictive accuracy across different population density regions with relatively low errors. While the model trained in the residential gathering area of Shanghai yielded an R^2 of 0.81, which, although slightly lower than Beijing, still reflected a strong fitting relationship, confirming the effectiveness of prediction in this region. The RMSE and MAE were 36.70 and 26.06, respectively, indicating some level of error, but the overall predicted trend remained consistent with the actual population distribution. Meanwhile, the MSLE was 0.46, suggesting slightly higher errors in some low-population-density areas, but without significantly affecting overall prediction accuracy. Overall, the RF models developed in this study demonstrated strong predictive capability in both study areas. The RF performed with higher accuracy in Beijing, while the Shanghai model maintained satisfactory interpretability.

In order to intuitively demonstrate the relationship between the predicted value and the actual value of the RF model, scatter density diagrams were further drawn, as shown in Fig. 6. The colors in the image represented point density, with red indicating high-density areas and blue indicating low-density areas. Overall, the RF regression model demonstrated favorable fitting performance in both study areas, with data points in both figures distributed along a straight line. The model performed better in Beijing, achieving a higher degree of agreement between predicted and actual values. Although the Shanghai model exhibited some deviation, it still demonstrated good applicability. Among them, the scatter distribution in Beijing was closer to the "$y = x$" line, with a regression equation slope of 0.8876 and an intercept of 5.95. This suggested that the model's predictions were relatively close to the actual values, with a smaller overall error and better fitting performance. It was also observed that data points were more densely clustered within the low to medium population range (0–100), implying higher prediction accuracy in this range. Similarly, the scatter plot for Shanghai also exhibited an overall trend, with a regression equation slope of 0.7186 and an intercept of 33.88, indicating that the predicted

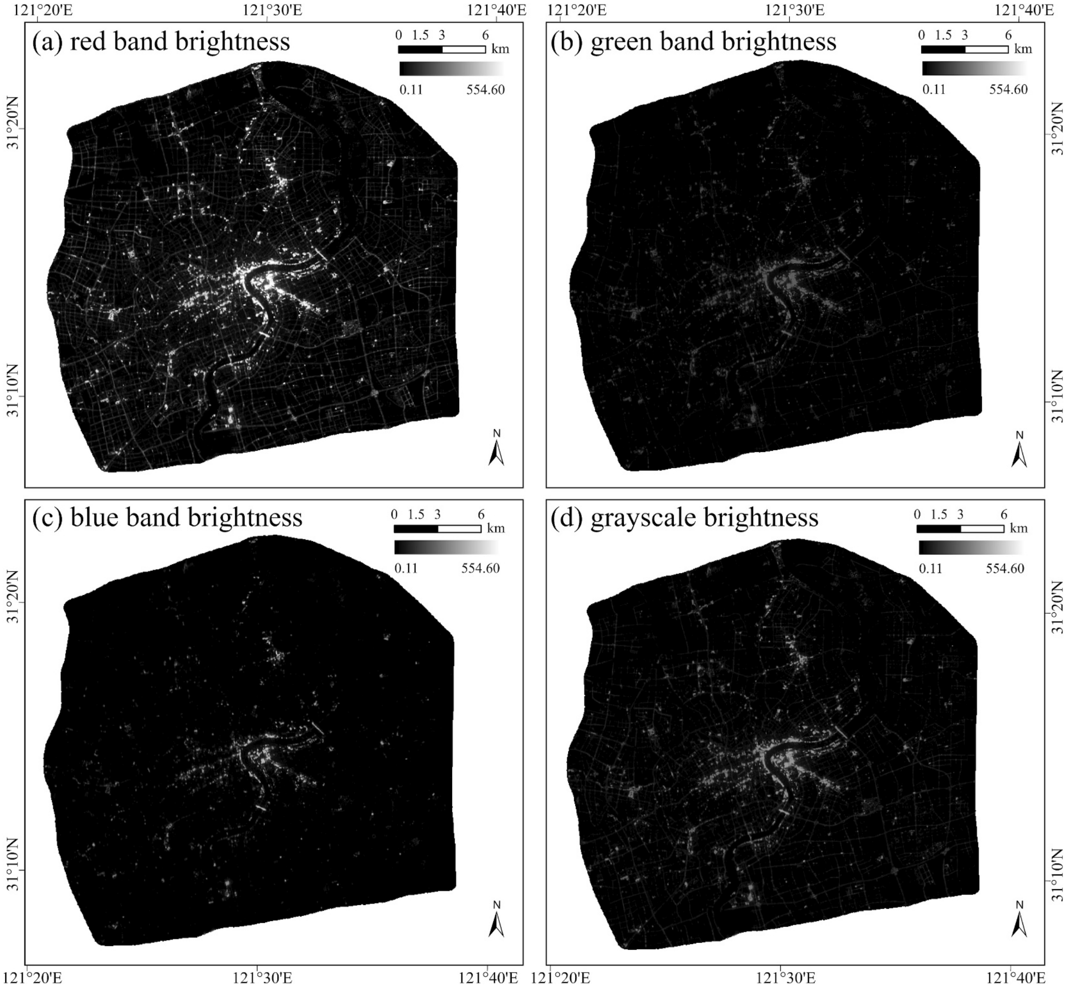

Fig. 5. NTL radiance values (W·m^{-2}·sr^{-1}·cm^{-1}) of SDGSAT-1 GLI images for the residential gathering area of Shanghai: (a)-(d) represent the red, green, blue bands and grayscale brightness, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Table of RF accuracy evaluation.

Evaluation index	Beijing	Shanghai
R^2	0.93	0.81
RMSE	11.11	36.70
MAE	7.60	26.06
MSLE	0.30	0.46

values in Shanghai were generally lower than the actual values. The error increased, and the scatter points became more dispersed, particularly in high-population-density areas (100−300). Meanwhile, high-density points were mainly concentrated in low-population areas (0–100), suggesting that the model maintained relatively stable prediction accuracy in these regions.

In general, the results above indicate that the proposed method is effective and can be applied to subsequently estimate the 40-m spatial population distribution for 2022.

4.2.2. *Feature importance screening of RF models*

Analyzing the feature importance of auxiliary variables is crucial to understanding the spatial distribution pattern of population and ensuring the robustness of regression relationships. By identifying key variables that significantly influence population distribution, the modeling process can be further optimized, ultimately enhancing the robustness and accuracy of estimated population. This study screened out key auxiliary variables by conducting feature importance analysis in combination with OOB data, and quantified the contribution of each variable to the model (as shown in Fig. 7).

Through experiments and a comprehensive consideration of the training models for the two cities, the final 15 key variables for RF training and prediction were determined based on the ranking of their contribution, as shown in Fig. 7. The results indicate that all auxiliary variables contribute to the model to varying degrees, validating the rationality and effectiveness of the selected variables. However, their contributions differ. Among these, the contribution of road density is the highest for both Beijing and Shanghai, followed by the remote sensing index NDBI, which measures the extent of surface building coverage. These two auxiliary variables play a critical role in the model. Notably, the POI_residential variable has the lowest contribution and even exhibits a negative value in the Beijing model, suggesting that this variable fails to effectively distinguish population distribution across different areas. Additionally, the importance of POI variables remains relatively balanced, indicating that different POI (e.g., commercial, transportation, and recreational facilities) have little difference in their impact on the spatial distribution of the population.

4.2.3. *Spatial pattern of the estimated 40-m population grid*

Based on the trained optimal RF models and 2022 auxiliary variables, 40-m resolution population raster data for the residential gathering areas of Beijing and Shanghai in 2022 were estimated. To ensure that the high-resolution population raster data remained consistent with

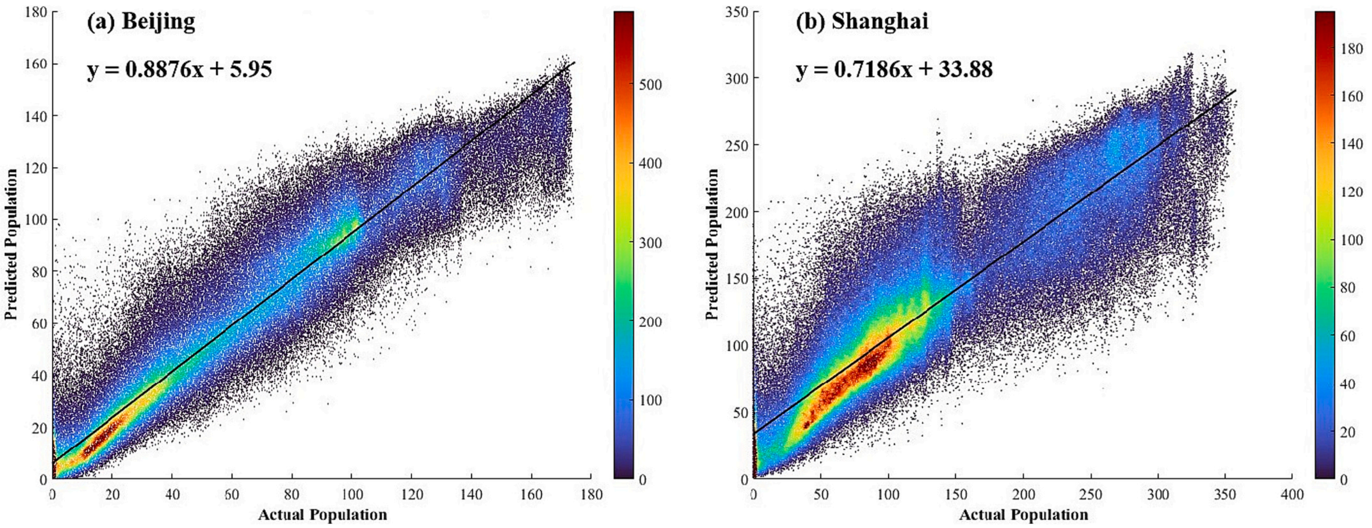

Fig. 6. Scatter density plots of trained RF regression models: (a) Beijing; (b) Shanghai.

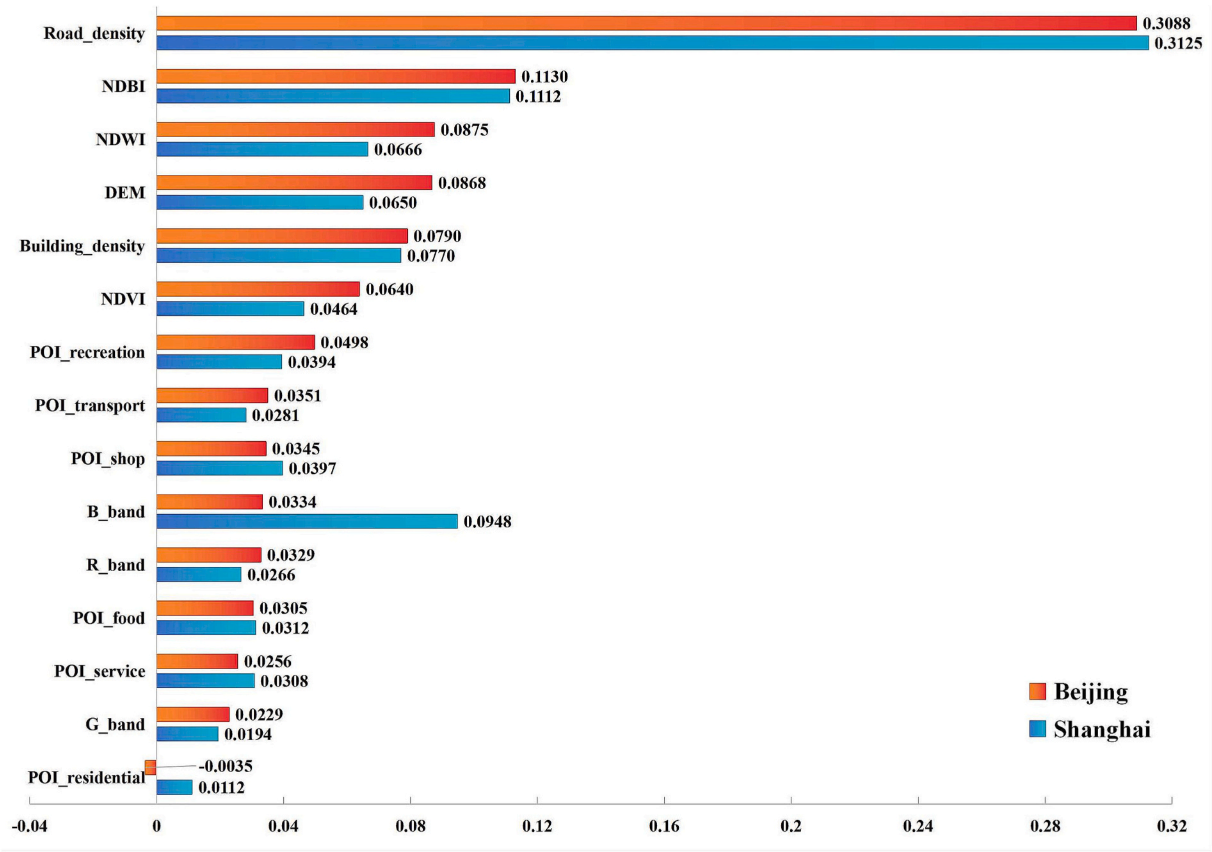

Fig. 7. Feature importance ranking.

macro-level statistics while preserving fine-scale details, these preliminarily estimated rasters were calibrated using district-level official demographic data. The results are presented in Fig. 8, compared with the original 100-m resolution WorldPop population data from 2020.

The spatial aggregation of the population is the result of long-term economic development and urbanization processes, exhibiting a high degree of stability (Sato and Yamamoto, 2005). Therefore, a reasonable population estimation should align with real-world spatial patterns. As shown in Fig. 8, the 40-m population data estimated for 2022 preserves the internal urban population structure and remains consistent with the overall trend of the original data. For instance, in Shanghai, the population was primarily concentrated in districts west of the Huangpu River, such as Hongkou, Jing'an, Huangpu, and the sides of Putuo, Xuhui, and Yangpu districts that are closer to the city center. In contrast, the population in Pudong New District was relatively lower, mainly clustered along the Huangpu River and around the Oriental Pearl Tower, with an overall decreasing trend toward the east. Similarly, in Beijing, the population was predominantly concentrated in Xicheng and Dongcheng districts, as well as in the sides of Haidian and Chaoyang districts that were closer to the city center. Whereas the population density decreased

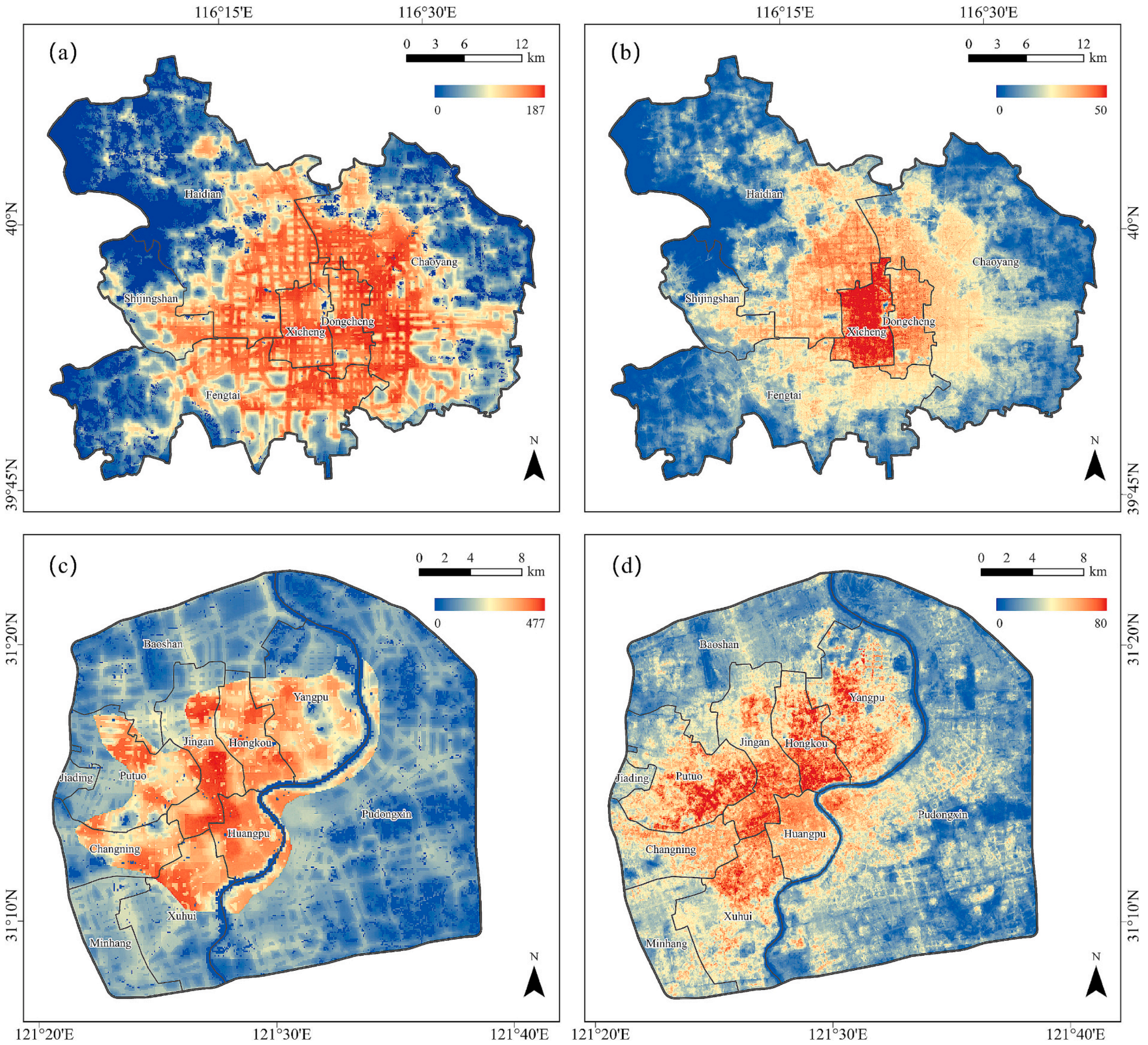

Fig. 8. Results of population downscaling estimation:
(a) 100-m WorldPop data of Beijing in 2020; (b) 40-m estimiated population data of Beijing in 2022; (c) 100-m WorldPop data of Shanghai in 2020; (d) 40-m estimiated population data of Shanghai in 2022.

in areas further from the core region. These results align with the urban development patterns of Shanghai and Beijing, as well as the spatial distribution trends observed in the original WorldPop population data, indicating that the RF model effectively comprehended the spatial characteristics of population distribution.

Furthermore, from the perspective of visual observation, the RF model effectively optimized the spatial distribution of the original population data, making the adjusted population distribution more reasonable and natural. For example, as for 100-m resolution WorldPop population data, the population boundaries in Shanghai exhibited a jagged distribution (Fig. 8c), resulting in relatively abrupt boundaries. In contrast, the estimated 40-m data (Fig. 8d) achieved a smooth transition, demonstrating a more continuous population gradient. Similarly, there was an overestimation of the population along the roads within Beijing's Fifth Ring, evident from the numerous red linear patches (Fig. 8a). Nevertheless, the downscaled results at 40-m resolution (Fig. 8b) eliminated this bias to a great extent. The results above indicate that the 40-m population rasters, processed through refined downscaling models, shows improved capability in detailing local population aggregation, thereby enhancing data quality.

4.3. Identification and evaluation of nighttime light pollution

4.3.1. Spatial analysis of NLSDMI

The calculation values, statistical results, and spatial patterns of NLSDMI in the residential gathering areas of the two megacities are shown in Fig. 9.

Based on the formulas established in Section 3.2.3, the NLSDMI was calculated in both megacities, as shown in Fig. 9(a) and Fig. 9(b). According to statistics, in the residential gathering area of Beijing, the NLSDMI reached a maximum value of 64.69, with an average of 0.47. In the residential gathering area of Shanghai, the NLSDMI had a maximum value of 185.56 and an average of 0.49, indicating the presence of prominent light pollution. These results highlight the spatial extent and level of NTL supply-demand mismatch in the two megacities, with Shanghai exhibiting a broader range of mismatch severity compared to Beijing.

In addition, the spatial statistical analyses on the derived NLSDWI are shown in Fig. 9(c)-(f). From the results of global spatial autocorrelation, the Global Moran's I value for Beijing was 0.58 and for Shanghai was 0.49, both significantly greater than zero ($Z > 2.58$, $P < 0.01$), indicating a strong positive spatial correlation and clear spatial

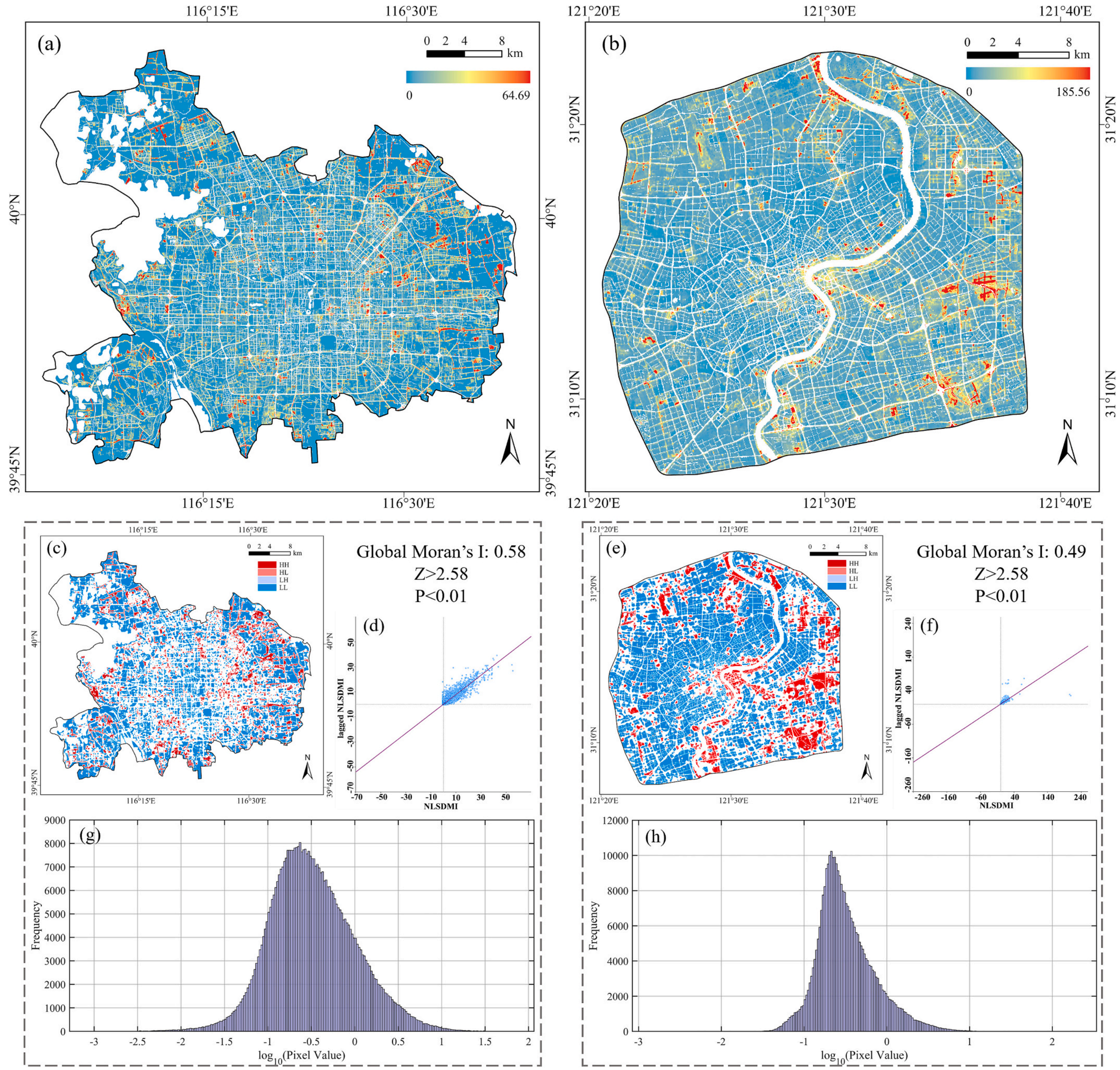

Fig. 9. Spatial patterns of NTL mismatch: NLSDMI of residential gathering areas in (a) Beijing and (b) Shanghai; Local Moran's I cluster maps of NLSDMI in (c) Beijing and (e) Shanghai (HH: high-high cluster; HL: high-low outlier; LH: low-high outlier; LL: low-low cluster); Moran scatter plots of NLSDMI in (d) Beijing and (f) Shanghai; Logarithmic histograms of NLSDWI for (g) Beijing and (h) Shanghai.

clustering of NTL mismatch in both regions. Compared with Shanghai, the higher Global Moran's I value for Beijing suggested a more concentrated spatial distribution of light mismatch phenomena. The local spatial autocorrelation maps (Local Moran's I cluster maps) further revealed the spatial distribution patterns of the NLSDWI. In both Beijing and Shanghai, the dominant cluster types were LL (low-low) and HH (high-high), indicating that the NLSDWI tends to exhibit homogeneous clustering in most areas.

As shown in Fig. 9(g) and Fig. 9(h), the NLSDMI values for the Beijing and Shanghai study areas exhibit an approximately normal distribution after logarithmic transformation. This indicates a significant reduction in data skewness and effective compression of extreme values. Such a distribution pattern is more suitable for subsequent statistical analyses and enhances the stability and reliability of the results.

4.3.2. *Spatial analysis of light pollution*

Based on the threshold determination approach proposed in Section 3.2.3, areas with NLSDMI values greater than 1.00 in Beijing and greater than 0.72 in Shanghai were identified as light-polluted areas, while those with values less than or equal to these thresholds were classified as non-polluted regions, as illustrated in Fig. 10(a) and Fig. 10(b). Spatially, light pollution in Beijing was more evenly distributed, with a noticeable tendency to align with roads, potentially affecting residents living alongside these areas. In contrast, light pollution in Shanghai displayed a localized clustering pattern, with high-pollution areas mainly clustered in regions where industrial and residential areas are mixed (such as Pudong and Baoshan). In some industrial zones, the NLSDMI even exceeds 180, reflecting the combined effects of industrial production and residential lighting.

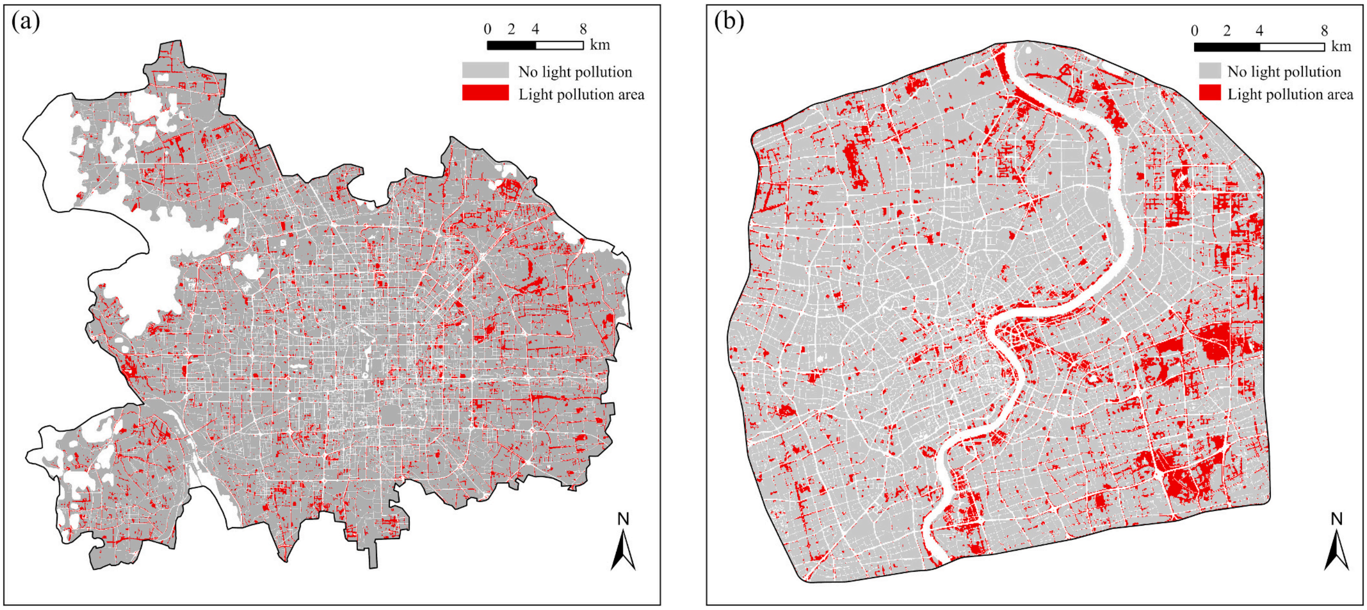

Fig. 10. Identified nighttime light-polluted regions in residential gathering areas of (a) Beijing and (b) Shanghai.

The statistical results of light pollution levels in different functional areas of the residential gathering areas of the two megacities are shown in Table 5. It can be seen that Beijing research area covers an area of 1023.77 km^2, of which 115.61 km^2 (11.29 %) were classified as light-polluted. Parks and green spaces (40.29 %) and residential areas (38.92 %) accounted for the largest shares of the polluted regions. This suggested that large ecological spaces (such as the Olympic Park) might have experienced light pollution diffusion due to facilities like light shows and pathway lighting, and that residential areas might have suffered from intensified pollution caused by inefficient streetlight design. Industrial areas made up only 5.88 % of the light-polluted land, which was significantly lower than in Shanghai. This was likely related to Beijing's industrial relocation policies, which moved industries to peripheral districts such as Tongzhou and Daxing. In Shanghai, the study area encompassed 518.27 km^2, with 78.25 km^2 (15.10 %) identified as light-polluted. Among these, residential areas (33.59 %) and industrial land (30.40 %) represented the largest proportions. The higher proportion of industrial land might be attributed to Shanghai's role as a manufacturing and port economy hub, where industrial parks operating around the clock—such as chemical and automobile manufacturing zones in Lingang and Baoshan—require continuous high-intensity lighting. Although the proportion of light pollution in residential areas was lower than in Beijing, the actual pollution density was higher when considering the smaller study area in Shanghai. In the residential gathering areas of Beijing and Shanghai, residential zones exhibited a notably high share of nighttime light pollution. This highlighted a significant mismatch between light supply and demand, where residents' low demand for nighttime light contrasted with actual excessive illumination. Such over lighting not only resulted in resource waste but also posed potential health risks to residents. Moreover, in both Beijing and Shanghai, land designated for transportation hubs accounted for the most minor proportion of light-polluted areas, at 0.17 % and 0.19 %, respectively.

Table 5
Areas and proportions of light pollution across different functional zones.

	Beijing		Shanghai	
	Area (km^2)	Proportion	Area (km^2)	Proportion
Residential	45.00	38.92 %	26.29	33.59 %
Business office	0.68	0.59 %	3.07	3.93 %
Commercial service	2.60	2.25 %	0.94	1.20 %
Industrial	6.80	5.88 %	23.79	30.40 %
Transportation stations	0.19	0.17 %	0.15	0.19 %
Airport facilities	3.24	2.80 %	3.59	4.58 %
Administrative	2.74	2.37 %	4.22	5.40 %
Educational	5.19	4.49 %	4.41	5.64 %
Medical	0.35	0.30 %	0.43	0.55 %
Sport and cultural	2.24	1.94 %	3.18	4.07 %
Park and green space	46.57	40.29 %	8.18	10.45 %
Total of polluted area	115.61	100.00 %	78.25	100.00 %
Total of study area	1023.77		518.27	

4.3.3. *Results of field survey*

The field survey results are summarized in Tables 6 and 7, and the field photos are provided in the supplementary material (Table S1 and Table S2).

Overall, the results of the field surveys are essentially consistent with the findings of this study. Specifically: (1) In Shanghai, sites 1 and 2, which are classified as Administrative and some Industrial areas that do not operate at night, have very low human traffic at night. However, the NTL intensity is not low, leading to relatively strong light pollution. (2) Sites 6, 8, and 9 in Shanghai, and site 2 in Beijing, are categorized as Business office and Commercial service. During our survey period, most of the Business office areas were in an overtime working state. Therefore, they have a high demand for lighting, and the human traffic is relatively high. Similarly, the light intensity is also relatively high, resulting in moderate light pollution. (3) Sites 10, 11, and 12 in Shanghai have relatively high human traffic, but the light intensity is not very high. Therefore, the light pollution is not severe. (4) Sites 4, 5, and 7 in Shanghai, and sites 3 and 5 in Beijing, are Residential and Park and greenspace. There is a certain degree of light pollution in these areas, indicating that these regions require more moderate nighttime lighting.

It is worth noting that site 3 in Shanghai, which is classified as Park and greenspace, has relatively high human traffic at night. However, its NTL intensity is very high. The field investigation found that the lights used in this area emit dazzling light that is vertically projected downward, resulting in a very high illuminance value and, consequently, strong light pollution. This poses a danger to both the natural ecology of the park and human health.

Table 6
Field survey data of nighttime light pollution in Shanghai.

ID	Functional zone	Geographic coordinate (x, y)	Cars	Individuals	Brightness value (LUX)				Light pollution level
					NO. 1	NO. 2	NO. 3	Mean	
1	Administrative	121.5410, 31.2240	1	1	26.5	31.2	24.5	27.40	13.70
2	Industrial	121.6129, 31.2420	3	1	43.5	41.3	39.8	41.53	10.38
3	Park and greenspace	121.5537, 31.2239	26	35	465.0	472.0	462.4	466.5	7.65
4	Residential	121.6131, 31.2407	5	7	59.4	110.6	57.8	75.9	6.33
5	Park and greenspace	121.6286, 31.2336	8	8	93.2	110.6	85.0	96.3	6.02
6	Business office	121.5318, 31.2484	11	12	128.3	124.7	129.8	127.6	5.55
7	Residential	121.6115, 31.2298	14	10	60.4	50.7	56.0	55.7	2.32
8	Business office	121.6154, 31.1864	27	27	120.2	105.6	113.4	113.1	2.09
9	Commercial service	121.5479, 31.1562	24	9	82.9	82.3	40.5	68.6	2.08
10	Residential	121.6064, 31.1722	12	15	43.7	43.2	38.8	41.9	1.55
11	Commercial service	121.3854, 31.1644	13	35	42.7	62.2	39.9	48.3	1.01
12	Industrial	121.4614, 31.1289	39	25	58.2	54.6	65.8	59.5	0.93

Table 7
Field survey data of nighttime light pollution in Beijing.

ID	Functional zone	Geographic coordinate (x, y)	Cars	Individuals	Brightness value (LUX)				Light pollution level
					NO. 1	NO. 2	NO. 3	Mean	
1	Medical	116.3267, 39.7302	7	12	119.6	114.8	110.3	114.9	6.05
2	Commercial service	116.3649, 39.8507	16	11	137.5	130.2	128.7	132.1	4.89
3	Residential	116.3268, 39.7262	7	4	42.3	35.1	37.3	38.2	3.48
4	Educational	116.3269, 39.7235	7	3	21.9	27.4	26.1	25.1	2.51
5	Residential	116.3267, 39.7254	5	6	17.4	21.7	19.2	19.4	1.77

5. Discussion

5.1. Refined estimation of population distribution grid

In this study, a high-performance RF model was employed to downscale 100-m resolution population data to a finer 40-m resolution. This model was successfully applied to estimate the population distribution for the year 2022, resulting in a very high-resolution spatial population dataset. Compared with previous studies, this study achieved competitive performance in key evaluation metrics and data resolutions ($R^2 = 0.93$ in Beijing and $R^2 = 0.81$ in Shanghai at 40 m). For instance, Zhou et al. (2024) used RF regression with multi-source geospatial data to refine population data ($R^2 = 0.59$ at 100 m), while Liu et al., 2023b applied RF-based downscaling of census data, achieving a relatively higher accuracy ($R^2 = 0.83$ at 150 m). This study effectively leveraged comprehensive auxiliary variables to capture nonlinear relationships, particularly by incorporating 40-m resolution SDGSAT-1 GLI data as one of the key predictors. As a high-precision source of NTL information, this dataset demonstrated superior performance in capturing and representing urban nighttime activity patterns. In addition, considering the inherent differences among various POI data, they were categorized and refined accordingly in this study. This approach effectively revealed the diverse relationships between different types of POIs and population distribution.

As mentioned above, the RF model performed well in both two megacities, but there were certain differences in model performance between them, particularly in terms of the R^2 value. Additionally, the RMSE and MAE metrics indicated that the prediction errors were smaller in Beijing and larger in Shanghai. The MSLE and scatter density plots revealed that, in low population density areas, prediction errors were slightly higher in Shanghai. The differences in model performance may be related to the adaptability of feature variables. In the Shanghai's model, the importance of the SDGSAT-1 GLI Blue Band and building density was higher than in Beijing, suggesting that the population distribution in Shanghai may be influenced by a more complex interplay of factors, leading to less stable relationships between variables and population distribution, which increased the difficulty of model fitting. In contrast, the relationships between feature variables and population distribution in Beijing appeared more direct. Furthermore, differences in urban spatial structure may also contribute to the observed discrepancies in model performance. As a monocentric metropolitan area, Beijing's employment opportunities are primarily concentrated in the urban center, resulting in a more direct relationship between physical features and population distribution, which facilitated more stable model fitting (Huang et al., 2015). In contrast, Shanghai's polycentric and mixed land-use urban structure, particularly in low-density areas, is characterized by urban-rural transitions and new district developments (Li et al., 2024; Ta et al., 2021; Zhang et al., 2019). These factors may lead to a decoupling of physical characteristics from actual residential populations, thereby increasing prediction errors in the model.

The WorldPop dataset used in this study, although undergoing multiple calibration processes and being widely applied, still exhibits spatial distribution biases. These biases are reflected in the model training and feature importance ranking. For example, road density had the largest contribution, which may be because dense road networks improve the accessibility of transportation, logistics, and public services, supporting higher population densities. However, it is also possible that the methodology behind the WorldPop data generation heavily relies on road network information, leading to an over-allocation of population along roads, thus increasing the contribution of this variable. Additionally, the contribution of the POI_residential variable is relatively low, even showing a negative value in the Beijing model. Theoretically, residential POIs should be positively correlated with population distribution, as residential areas are the primary locations of human settlements. However, WorldPop tends to allocate population to roads rather than residential areas, which reduces the contribution of POI_residential to the model.

This study explored feasible methods to acquire refined population grids. High-precision population data with fine resolutions, by revealing spatial distribution patterns of populations, provides crucial support for various research fields that rely on spatially explicit population data, such as urban planning, resource allocation, disaster assessment, and social governance (Duan et al., 2024). The population grids can support the monitoring, evaluation, and decision-making of several SDGs. Specifically, they can be applied to assess SDG 3 (Good Health and Well-being) (Juran et al., 2018), SDG 4 (Quality Education) (Qiu et al., 2019), and SDG 7 (Affordable and Clean Energy) (Gaughan et al., 2019), and are particularly valuable in evaluating SDG 11 (Sustainable Cities and Communities). They can measure disaster exposure to assess SDG 11.5 (Tuholske et al., 2021) and refine the percentage of urban populations living in slums, informal settlements, or lacking adequate housing to contribute to SDG 11.1.1 (Thomson et al., 2022). "Leave no one behind" is one of the core principles of the SDGs, and approximately half of the SDG indicators rely on population data for tracking progress (Qiu et al., 2022). This highlights the critical importance of high-precision grid-based population distribution data in monitoring the progress of the SDGs. Therefore, further advancement of high-precision population data research, with a focus on enhancing its usability and accuracy, remain essential.

5.2. NLSDMI: the most sophisticated nighttime light pollution index to date

Previous studies primarily focused on the phenomenon of light pollution and its impacts (Mu et al., 2021; Xue et al., 2020). Drawing inspiration from the concept of supply-demand relationships (Ye et al., 2024), this study expanded light pollution research by exploring mismatches in nighttime light supply and demand, with a particular emphasis on the imbalance between human needs and the spatial distribution of NTL sources. Additionally, this research was the first to incorporate urban functional zone data. Functional zones, formed by clustering similar socioeconomic activities, are characterized by a primary function (e.g., residential, educational, or commercial). These zones represent critical components of urban areas, facilitating detailed spatial division and offering insights into human interactions and demands within different parts of the urban environment. This study examined the issue from both supply and demand perspectives, integrating urban functional zone data with nighttime light intensity information to compute demand weights for different zones. A formula was developed to quantify the extent of mismatches between NTL supply and demand. Maintaining a reasonable balance between supply and demand is essential, as mismatches could result in resource waste, environmental degradation, and potential risks to human health and ecosystems. The analytical framework proposed in this study provides a fresh perspective on nighttime light distribution and management. It also serves as a valuable reference for future urban lighting planning and the pursuit of sustainable development.

This study introduced the concept of demand weights to address the limitations of defining NTL demand solely based on population size. Instead, demand weights for different urban functional zones were calculated using the average NTL intensity per unit area. The results revealed notable variations in NTL demand weights across different types of urban functional zones. These disparities may be attributed to several underlying factors, including: (1) Differences in the types of socioeconomic activities: The dominant function of each zone largely determines the intensity and characteristics of its lighting demand. Commercial zones, characterized by vibrant nighttime economic activities, typically require high-intensity and prolonged lighting to extend consumption hours and attract visitors. In contrast, residential zones prioritize the well-being of inhabitants, necessitating a balance between safety lighting and minimizing light intrusion. Industrial zones emphasize task-oriented lighting, with illumination concentrated in specific operational areas. (2) Temporal variations in human activity and staggered demand: Human activity intensity and spatial distribution vary considerably across functional zones throughout the day. Mobile phone signaling data suggest that commercial zones exhibit sustained human presence and activity into midnight, while activity in residential areas declines sharply after 22:00. Moreover, certain special-purpose zones, such as hospitals, require continuous high-intensity NTL to ensure the operation of emergency access routes. (3) Variation in lighting equipment types: Functional zones differ significantly in their lighting infrastructure. For instance, airports employ high-mast floodlights and in-ground LED luminaires to meet the safety requirements of aviation environments, illuminating runways, aprons, and adjacent parking areas with high-intensity, large-coverage lighting. (4) Policy-driven regulatory constraints: Certain ecological and natural areas—such as urban parks, wetlands, and green spaces—are subject to strict lighting controls under "dark-sky" protection policies. These regulations aim to reduce light pollution, preserve ecological integrity, and maintain conditions suitable for astronomical observation, thereby significantly limiting artificial illumination in such zones.

As a crucial component of modern urban infrastructure, NTL enhances nighttime safety and functionality, yet it has also led to increasingly severe light pollution. Although light pollution is not explicitly mentioned in the SDGs, it is closely linked to several goals, particularly SDG 3 and SDGs 11–15 (Lyytimäki, 2025). The NLSDMI proposed in this study provides a scientific and quantitative foundation for light pollution management. It identifies redundant and misallocated lighting resources, offering a more accurate reflection of urban lighting suitability and environmental pressure than conventional light intensity metrics. Guided by NLSDMI, street lighting can be dimmed in targeted areas to improve sleep quality (Kyba et al., 2021), supporting SDG 3; Restricting private lighting sources and quantifying light spillover can improve energy efficiency (Bouroussis and Topalis, 2020), contributing to SDG 7; additionally, establishing lighting zone classifications and integrating ecological data into tailored lighting design can promote more sustainable urban environments (Haddock et al., 2019; Tavares et al., 2021), aligning with SDG 11. In summary, light pollution management based on NLSDMI represents a vital pathway toward achieving multiple SDGs in a coordinated and evidence-based manner.

5.3. Differences in nighttime light pollution patterns among megacities with specific functions

This study selected Beijing and Shanghai, two representative megacities in China, as the research areas to explore the extent and spatial patterns of nighttime light pollution based on the NLSDMI developed. The results revealed two major differences: (1) Differences in land-use types with high NTL. In Beijing, a relatively large proportion of light-polluted areas are associated with parks and green spaces, followed by residential areas. Parks and green spaces typically require minimal artificial lighting, and excessive illumination may disrupt local ecosystems and wildlife habitats (Huang et al., 2023), while also reducing urban livability. In contrast, besides residential areas, industrial areas in Shanghai exhibit the second highest proportion of light pollution. This may be due to the low population density in industrial areas, while certain 24-h industrial operations generate high-intensity lighting. (2) Differences in spatial distribution patterns of light pollution. Beijing exhibits a more dispersed and evenly distributed pattern of light pollution. This may be related to its exemplary role as China's political and cultural center, leading to a more balanced urban development. In contrast, Shanghai presents a more concentrated and pronounced pattern of light pollution, with significantly higher NLSDMI values. This is likely a result of its function as an economic center, where intensive commercial and industrial activities drive greater demand for nighttime illumination.

Cities differ significantly in their light demand and pollution characteristics due to variations in functional roles and development levels, which reflect their distinct economic, social, and environmental responsibilities. (1) In terms of urban functional positioning, economically driven cities like Shanghai prioritize economic activities, with frequent nighttime commercial and industrial operations generating high light demand. Politically oriented cities like Beijing focus on administrative and cultural functions, leading to more dispersed light demand. Light pollution tends to be more evenly distributed in these cities, with notable impacts along roads, within administrative zones, and in green spaces. (2) In terms of development, advanced cities with well-established infrastructure experience more frequent and intense nighttime light use, exacerbating light pollution issues. These cities should focus on implementing energy-saving technologies and strengthening light pollution control. Conversely, less developed cities, while currently facing lighter light pollution due to lower economic activity and light demand, may encounter growing challenges as urbanization accelerates. Effective research and management of light pollution require careful consideration of city-type differences. Strategies must be tailored to align with each city's functional roles and development characteristics, ensuring targeted and effective mitigation measures.

5.4. *Strategies for refined management of NTL in residential gathering areas*

As cities continue to develop, achieving the SDGs and advancing the vision of human well-being and sustainable communities necessitate more precise management of light pollution. Building on the findings of this study, this paper offers targeted recommendations from various perspectives to support efforts in mitigating light pollution, enhancing urban lighting environments, and fostering sustainable urban development.

(1) **Adopt targeted lighting strategies.** In residential areas, implement human-centered smart dimming systems that dynamically adjust brightness at night based on actual demand, thereby minimizing disruption to residents' sleep and well-being. In parks and green spaces, reduce unnecessary decorative lighting and enforce a night-time lighting curfew (e.g., 23:00–05:00) to preserve natural darkness and protect ecosystems and wildlife habitats (Xue et al., 2020). In industrial zones, promote the adoption of energy-efficient, environmentally friendly lighting systems and strictly regulate external illumination during non-operational hours to minimize negative effects on surrounding areas.

(2) **Refine urban land-use planning.** Incorporate light-pollution buffer zones into functional zoning (Gaston et al., 2015; Wei et al., 2025): establish green belts or low-illumination corridors between high-intensity commercial districts and residential neighborhoods, using vegetation and terrain to block light spillover and improve nocturnal visual comfort for residents. Impose strict lighting thresholds on ecological corridors—such as parks, riverbanks, and wetlands to balance ecological protection with nighttime mobility and safety.

(3) **Innovate lighting facilities and technologies.** Encourage the use of low-color-temperature, low-intensity luminaires to mitigate the proportion of blue light and its impacts on human health and ecosystems (Lin et al., 2023; Liu et al., 2024). Deploy smart streetlights to achieve on-demand lighting and maximize energy savings (Pardo-Bosch et al., 2022). Promote modular LED fixtures and plug-and-play control units to streamline maintenance and upgrades, extend equipment lifespan, and reduce life-cycle energy consumption.

5.5. *Limitations and prospects*

This study explored and applied SDGSAT-1 GLI data to derive 40-m resolution population estimates for 2022 and conducted a preliminary investigation into light pollution. However, several aspects require further exploration. The limitations of this study are outlined below.

(1) **The availability and accuracy of data require further improvement.** This study employed SDGSAT-1 GLI data with basic preprocessing. However, visual inspection and comparisons with other studies suggest that the data may retain residual noise and cloud artifacts that are challenging to remove. Although this study utilized high-quality NTL images, future research using lower-quality data will need more robust and comprehensive preprocessing methods. Additionally, the urban functional zone data used in this study were relatively outdated. While corrections were made using POI data, discrepancies with the actual urban structure remain. Furthermore, the dataset lacked road information, a critical urban component that should be integrated into future studies to enhance accuracy and completeness.

(2) **More accurate scale transfer models need to be developed.** This study utilized a RF model for population estimation, successfully capturing nonlinear relationships between population data and auxiliary variables. However, the model's performance varied across different cities, with inconsistent results. Future research should prioritize optimizing model parameters and architecture and exploring models with enhanced migratability. Additionally, incorporating auxiliary variables more strongly correlated with the target variable and refining variable selection methods are critical for improving accuracy and applicability of the model.

(3) **The light pollution evaluation method remains to be further optimized.** This study defined urban light pollution through a supply-demand framework, calculated mismatch indices, and developed zonal maps of nighttime light pollution. It further analyzed factors such as population distribution, urban functional zones, and intercity differences. However, the findings were not validated through comprehensive and systematic investigations. Although the field survey based on statistical random sampling conducted by this study can provide some field information, the limitations of the survey coverage, insufficient spatial representation of the sample points, insufficient synchronization of measurement time, irregular instrument operation, and deviations in the counting of pedestrians and vehicles will ultimately affect the accuracy and generalizability of the results. Moreover, the process of determining demand weights could be further improved.

6. Conclusion

This study provides a comprehensive assessment of nighttime light pollution in typical megacities, taking the residential gathering regions of Beijing and Shanghai as case studies. By leveraging SDGSAT-1 GLI imagery and other multi-source high-resolution spatial data, a novel index (NLSDMI) with 40-m resolution was developed. With the help of NLSDMI and its spatial distribution, the imbalance between nighttime lighting supply and human activity demand was finely quantified, and key areas affected by light pollution was further identified. In the residential gathering area of Beijing, the light pollution area reached 115.61 km^2 (11.29 %), mainly distributed in green spaces and residential zones. In Shanghai, the situation in residential gathering are was more severe, with an affected area of 78.25 km^2 (15.10 %), mainly concentrated in residential and industrial zones. These findings highlighted the pressing need for tailored management of urban spaces to enhance urban sustainability, such as targeted lighting strategy, energy-efficient technology, and refined urban planning. Future research should focus on enhancing data accuracy, improving validation methods, and exploring the applicability of findings to cities with diverse types and scales, thus providing broader theoretical support and practical guidance for global SDGs-oriented urban management.

The significance of this study lies in its innovative approach to quantifying and analyzing nighttime light pollution from both the supply and demand perspectives. It is the first to employ high-resolution SDGSAT-1 GLI imagery to provide a fine-scale representation of NTL supply, and to estimate the spatial distribution of lighting demand based on RF-downscaled population data and weighted urban functional zoning. A continuous index of nighttime light pollution (NLSDMI) was constructed to quantify the imbalance between lighting supply and demand, thereby enabling the identification of key light-polluted areas. By revealing the mismatch, this study offers valuable insights for optimizing urban lighting and advancing the achievement of the SDGs.

CRediT authorship contribution statement

Yixuan Wu: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Chenhao Huang:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yang Ye:** Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Linlu Mei:** Supervision, Project administration, Data curation. **Yalan Liu:** Supervision, Project administration, Data curation. **Dacheng Wang:** Supervision, Project administration, Data curation. **Weirong Chen:** Writing – review & editing, Visualization, Validation, Investigation. **Jinsong Deng:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization.

Funding

This research was funded by the Intergovernmental International Science and Technology Innovation Cooperation Program Under National Key Research and Development Plan (2024YFE0198600), the Basic Investigation Program of the Ministry of Science and Technology of the People's Republic of China (2023FY100102), the National Natural Science Foundation of China (22376179), and the Natural Science Foundation of Zhejiang Province (LTGS24D010002).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research findings are a component of the SDGSAT-1 Open Science Program, which is conducted by the Inter-national Research Center of Big Data for Sustainable Development Goals (CBAS). The data utilized in this study is sourced from SDGSAT-1 and provided by CBAS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2025.114894.

Data availability

Data will be made available on request.

References

Abdelrahim, A.I.M., Yücel, Ö., 2025. A machine learning based regression methods to predicting syngas composition for plasma gasification system. Fuel 381, 133575. https://doi.org/10.1016/j.fuel.2024.133575.

Bagheri, S., Karimzadeh, S., Feizizadeh, B., 2023. Investigation and modeling of physical development of urban areas and its effects on light pollution using night light data. IJEG 8, 98–110. https://doi.org/10.26833/ijeg.976495.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: Methods, applications and data. In: Hay, Simon I., Graham, A., Rogers, D.J. (Eds.), Advances in Parasitology, Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications. Academic Press, pp. 119–156. https://doi.org/10.1016/S0065-308X(05)62004-0.

Barua, E., Kabir, Md.N., Islam, Md.A., 2024. Monitoring the trend of nighttime light pollution (NLP) in the Chattogram District, Bangladesh, using DMSP/OLS and VIIRS data. J. Geovis. Spat. Anal. 8, 31. https://doi.org/10.1007/s41651-024-00192-2.

Bouroussis, C.A., Topalis, F.V., 2020. Assessment of outdoor lighting installations and their impact on light pollution using unmanned aircraft systems - the concept of the drone-gonio-photometer. J. Quant. Spectrosc. Radiat. Transf. 253, 107155. https://doi.org/10.1016/j.jqsrt.2020.107155.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Davies, T.W., Smyth, T., 2018. Why artificial light at night should be a focus for global change research in the 21st century. Glob. Chang. Biol. 24, 872–882. https://doi.org/10.1111/gcb.13927.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote. Sens. 66, 849–857.

Duan, H., Shi, Zhongqi, Ge, Ji, Wu, Fan, Liu, Yuzhou, Hong, Zhang, Wang, C., 2024. High-resolution population mapping based on SDGSAT-1 glimmer imagery and deep learning: a case study of the Guangdong-Hong Kong-Marco Greater Bay Area. Intern. J. Digital Earth 17, 2407519. https://doi.org/10.1080/17538947.2024.2407519.

Gaston, K.J., Miguel, A.S., 2022. Environmental impacts of artificial light at night. Annu. Rev. Environ. Resour. 47, 373–398. https://doi.org/10.1146/annurev-environ-112420-014438.

Gaston, K.J., Duffy, J.P., Bennie, J., 2015. Quantifying the erosion of natural darkness in the global protected area system. Conserv. Biol. 29, 1132–1141. https://doi.org/10.1111/cobi.12462.

Gaughan, A.E., Oda, T., Sorichetta, A., Stevens, F.R., Bondarenko, M., Bun, R., Krauser, L., Yetman, G., Nghiem, S.V., 2019. Evaluating nighttime lights and population distribution as proxies for mapping anthropogenic CO2 emission in Vietnam, Cambodia and Laos. Environ. Res. Commun. 1, 091006. https://doi.org/10.1088/2515-7620/ab3d91.

Gong, P., Chen, B., Li, Xuecao, Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., Feng, Y., Gong, Y., Gu, H., Huang, H., Huang, X., Jiao, H., Kang, Y., Lei, G., Li, A., Li, Xiaoting, Li, Xun, Li, Y., Li, Zhilin, Li, Zhongde, Liu, Chong, Liu, Chunxia, Liu, M., Liu, S., Mao, W., Miao, C., Ni, H., Pan, Q., Qi, S., Ren, Z., Shan, Z., Shen, S., Shi, M., Song, Y., Su, M., Ping Suen, H., Sun, B., Sun, F., Sun, J., Sun, L., Sun, W., Tian, T., Tong, X., Tseng, Y., Tu, Y., Wang, H., Wang, L., Wang, X., Wang, Z., Wu, T., Xie, Y., Yang, Jian, Yang, Jun, Yuan, M., Yue, W., Zeng, H., Zhang, K., Zhang, N., Zhang, T., Zhang, Y., Zhao, F., Zheng, Y., Zhou, Q., Clinton, N., Zhu, Z., Xu, B., 2020. Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Sci. Bull. 65, 182–187. https://doi.org/10.1016/j.scib.2019.12.007.

Grundland, M., Dodgson, N.A., 2007. Decolorize: fast, contrast enhancing, color to grayscale conversion. Pattern Recogn. 40, 2891–2896. https://doi.org/10.1016/j.patcog.2006.11.003.

Guo, B., Hu, D., Zheng, Q., 2023a. Potentiality of SDGSAT-1 glimmer imagery to investigate the spatial variability in nighttime lights. Int. J. Appl. Earth Obs. Geoinf. 119, 103313. https://doi.org/10.1016/j.jag.2023.103313.

Guo, H., Dou, C., Chen, H., Liu, J., Fu, B., Li, X., Zou, Z., Liang, D., 2022. SDGSAT-1: the world's first scientific satellite for sustainable development goals. Sci. Bull. 68. https://doi.org/10.1016/j.scib.2022.12.014.

Guo, W., Zhang, J., Zhao, X., Li, Y., Liu, J., Sun, W., Fan, D., 2023b. Combining Luojia1-01 nighttime light and points-of-interest data for fine mapping of population

Spatialization based on the zonal classification method. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 16, 1589–1600. https://doi.org/10.1109/JSTARS.2023.3238188.

Haddock, J.K., Threlfall, C.G., Law, B., Hochuli, D.F., 2019. Light pollution at the urban forest edge negatively impacts insectivorous bats. Biol. Conserv. 236, 17–28. https://doi.org/10.1016/j.biocon.2019.05.016.

Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci. Model Dev. 15, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Huang, C., Zhuang, Q., Meng, X., Guo, H., Han, J., 2021. An improved nightlight threshold method for revealing the spatiotemporal dynamics and driving forces of urban expansion in China. J. Environ. Manag. 289, 112574. https://doi.org/10.1016/j.jenvman.2021.112574.

Huang, C., Ye, Y., Jin, Y., Liang, B., 2023. Research Progress, hotspots, and evolution of nighttime light pollution: analysis based on WOS database and remote sensing data. Remote Sens. 15, 2305. https://doi.org/10.3390/rs15092305.

Huang, D., Liu, Z., Zhao, X., 2015. Monocentric or polycentric? The urban spatial structure of employment in Beijing. Sustainability 7, 11632–11656. https://doi.org/10.3390/su70911632.

Hutengs, C., Vohland, M., 2016. Downscaling land surface temperatures at regional scales with random forest regression. Remote Sens. Environ. 178, 127–141. https://doi.org/10.1016/j.rse.2016.03.006.

Juran, S., Broer, P.N., Klug, S.J., Snow, R.C., Okiro, E.A., Ouma, P.O., Snow, R.W., Tatem, A.J., Meara, J.G., Alegana, V.A., 2018. Geospatial mapping of access to timely essential surgery in sub-Saharan Africa. BMJ Glob. Health 3. https://doi.org/10.1136/bmjgh-2018-000875.

Knibbe, J.S., van der, A.R.J., de Laat, A.T.J., 2014. Spatial regression analysis on 32 years of total column ozone data. Atmos. Chem. Phys. 14, 8461–8482. https://doi.org/10.5194/acp-14-8461-2014.

Kuechly, H.U., Kyba, C.C.M., Ruhtz, T., Lindemann, C., Wolter, C., Fischer, J., Hölker, F., 2012. Aerial survey and spatial analysis of sources of light pollution in Berlin, Germany. Remote Sens. Environ. 126, 39–50. https://doi.org/10.1016/j.rse.2012.08.008.

Kyba, C., Ruby, A., Kuechly, H., Kinzey, B., Miller, N., Sanders, J., Barentine, J., Kleinodt, R., Espey, B., 2021. Direct measurement of the contribution of street lighting to satellite observations of nighttime light emissions from urban areas. Light. Res. Technol. 53, 189–211. https://doi.org/10.1177/1477153520958463.

Lei, Z., Zhou, S., Cheng, P., Xie, Y., 2024. Improved population mapping for China using the 3D building, nighttime light, points-of-interest, and land use/cover data within a multiscale geographically weighted regression model. ISPRS Int. J. Geo Inf. 13, 335. https://doi.org/10.3390/ijgi13090335.

Levin, N., Kyba, C.C.M., Zhang, Q., Sánchez de Miguel, A., Román, M.O., Li, X., Portnov, B.A., Molthan, A.L., Jechow, A., Miller, S.D., Wang, Z., Shrestha, R.M., Elvidge, C.D., 2020. Remote sensing of night lights: a review and an outlook for the future. Remote Sens. Environ. 237, 111443. https://doi.org/10.1016/j.rse.2019.111443.

Li, C., Chen, F., Wang, N., Yu, B., Wang, L., 2023a. SDGSAT-1 nighttime light data improve village-scale built-up delineation. Remote Sens. Environ. 297, 113764. https://doi.org/10.1016/j.rse.2023.113764.

Li, L., Gao, T., Wang, Y., Jin, Y., 2023b. Evaluation of public transportation station area accessibility based on walking perception. Intern. J. Transport. Sci. Technol. 12, 640–651. https://doi.org/10.1016/j.ijtst.2023.01.001.

Li, W., Gu, H., Han, S., Sun, B., 2024. On the changing of urban spatial structure: a glimpse from Shanghai's relocated residents. Cities 147, 104795. https://doi.org/10.1016/j.cities.2024.104795.

Lin, Z., Jiao, W., Liu, H., Long, T., Liu, Y., Wei, S., He, G., Portnov, B.A., Trop, T., Liu, M., Li, X., Wen, C., 2023. Modelling the public perception of urban public space lighting based on SDGSAT-1 glimmer imagery: a case study in Beijing, China. Sustain. Cities Soc. 88, 104272. https://doi.org/10.1016/j.scs.2022.104272.

Lis, A., Zienowicz, M., Kącki, Z., Iwankowski, P., Kukowska, D., Shestak, V., 2024. Park lighting after dark – is it a route or a place? How people feel in park nightscapes (experiment). Landsc. Urban Plan. 248, 105098. https://doi.org/10.1016/j.landurbplan.2024.105098.

Liu, C., Chen, Y., Wei, Y., Chen, F., 2023a. Spatial population distribution data disaggregation based on SDGSAT-1 nighttime light and land use data using Guilin, China, as an example. Remote Sens. 15, 2926. https://doi.org/10.3390/rs15112926.

Liu, L., Cheng, G., Yang, J., Cheng, Y., 2023b. Population spatialization in Zhengzhou city based on multi-source data and random forest model. Front. Earth Sci. 11. https://doi.org/10.3389/feart.2023.1092664.

Liu, Saimiao, Zhou, Y., Wang, F., Wang, S., Wang, Z., Wang, Y., Qin, G., Wang, P., Liu, M., Huang, L., 2024. Lighting characteristics of public space in urban functional areas based on SDGSAT-1 glimmer imagery:A case study in Beijing, China. Remote Sens. Environ. 306, 114137. https://doi.org/10.1016/j.rse.2024.114137.

Liu, Shaoyang, Wang, C., Chen, Z., Li, W., Zhang, L., Wu, B., Huang, Y., Li, Y., Ni, J., Wu, J., Yu, B., 2024a. Efficacy of the SDGSAT-1 glimmer imagery in measuring sustainable development goal indicators 7.1.1, 11.5.2, and target 7.3. Remote Sens. Environ. 305, 114079. https://doi.org/10.1016/j.rse.2024.114079.

Liu, Shaoyang, Wang, C., Wu, B., Chen, Z., Zhang, J., Huang, Y., Wu, J., Yu, B., 2024b. Integrating NTL intensity and building volume to improve the built-up areas' extraction from SDGSAT-1 GLI data. Remote Sens. 16, 2278. https://doi.org/10.3390/rs16132278.

Lyytimäki, J., 2025. Sustainable development goals relighted: light pollution management as a novel lens to SDG achievement. Discov. Sustain. 6, 197. https://doi.org/10.1007/s43621-025-00991-7.

Masselot, P., Chebana, F., Bélanger, D., St-Hilaire, A., Abdous, B., Gosselin, P., Ouarda, T.B.M.J., 2018. Aggregating the response in time series regression models, applied to weather-related cardiovascular mortality. Sci. Total Environ. 628–629, 217–225. https://doi.org/10.1016/j.scitotenv.2018.02.014.

Mu, H., Li, X., Du, X., Huang, J., Su, W., Hu, T., Wen, Y., Yin, P., Han, Y., Xue, F., 2021. Evaluation of light pollution in global protected areas from 1992 to 2018. Remote Sens. 13, 1849. https://doi.org/10.3390/rs13091849.

Niu, W., Xia, H., Wang, R., Pan, L., Meng, Q., Qin, Y., Li, R., Zhao, X., Bian, X., Zhao, W., 2021. Research on large-scale urban shrinkage and expansion in the Yellow River affected area using night light data. ISPRS Int. J. Geo Inf. 10, 5. https://doi.org/10.3390/ijgi10010005.

Pardo-Bosch, F., Blanco, A., Sesé, E., Ezcurra, F., Pujadas, P., 2022. Sustainable strategy for the implementation of energy efficient smart public lighting in urban areas: case study in san Sebastian. Sustain. Cities Soc. 76, 103454. https://doi.org/10.1016/j.scs.2021.103454.

Qiu, Y., Zhao, X., Fan, D., Li, S., 2019. Geospatial disaggregation of population data in supporting SDG assessments: a case study from Deqing County, China. ISPRS Int. J. Geo Inf. 8, 356. https://doi.org/10.3390/ijgi8080356.

Qiu, Y., Zhao, X., Fan, D., Li, S., Zhao, Y., 2022. Disaggregating population data for assessing progress of SDGs: methods and applications. Intern. J. Digi. Earth. 15 (1), 2–29.

Sato, Y., Yamamoto, K., 2005. Population concentration, urbanization, and demographic transition. J. Urban Econ. 58, 45–61. https://doi.org/10.1016/j.jue.2005.01.004.

Smith, K.T., Lopez, B., Vignieri, S., Wible, B., 2023. Losing the darkness. Science 380, 1116–1117. https://doi.org/10.1126/science.adi4552.

Stone, T., 2018. The value of darkness: a moral framework for urban nighttime lighting. Sci. Eng. Ethics 24, 607–628. https://doi.org/10.1007/s11948-017-9924-0.

Ta, N., Zeng, Y., Zhu, Q., Wu, J., 2020. Relationship between built environment and urban vitality in Shanghai downtown area based on big data. Sci. Geogr. Sin. 40, 60–68. https://doi.org/10.13249/j.cnki.sgs.2020.01.008.

Ta, N., Kwan, M.-P., Lin, S., Zhu, Q., 2021. The activity space-based segregation of migrants in suburban Shanghai. Appl. Geogr. 133, 102499. https://doi.org/10.1016/j.apgeog.2021.102499.

Tang, Q., Wang, C., Wang, Y., Su, R., Jiang, J., Cui, H., 2020. Preliminary study on light pollution monitoring based on volunteered passenger aircraft remote sensing. Remote Sens. Technol. Appl. 35, 1360–1367.

Tatem, A.J., 2017. WorldPop, open data for spatial demography. Sci. Data 4, 170004. https://doi.org/10.1038/sdata.2017.4.

Tavares, P., Ingi, D., Araújo, L., Pinho, P., Bhusal, P., 2021. Reviewing the role of outdoor lighting in achieving sustainable development goals. Sustainability 13, 12657. https://doi.org/10.3390/su132212657.

Thomson, D.R., Stevens, F.R., Chen, R., Yetman, G., Sorichetta, A., Gaughan, A.E., 2022. Improving the accuracy of gridded population estimates in cities and slums to monitor SDG 11: evidence from a simulation study in Namibia. Land Use Policy 123, 106392. https://doi.org/10.1016/j.landusepol.2022.106392.

Tobler, W., Deichmann, U., Gottsegen, J., Maloy, K., 1997. World population in a grid of spherical quadrilaterals. Int. J. Popul. Geogr. 3, 203–225. https://doi.org/10.1002/(SICI)1099-1220(199709)3:3<203::AID-IJPG68>3.0.CO;2-C.

Tuholske, C., Gaughan, A.E., Sorichetta, A., de Sherbinin, A., Bucherie, A., Hultquist, C., Stevens, F., Kruczkiewicz, A., Huyck, C., Yetman, G., 2021. Implications for tracking SDG Indicator metrics with gridded population data. Sustainability 13, 7329. https://doi.org/10.3390/su13137329.

UN, 2015. Transforming our World: The 2030 Agenda for Sustainable Development. United Nations, New York.

Walker, W.H., Bumgarner, J.R., Walton, J.C., Liu, J.A., Meléndez-Fernández, O.H., Nelson, R.J., DeVries, A.C., 2020. Light pollution and Cancer. Int. J. Mol. Sci. 21, 9360. https://doi.org/10.3390/ijms21249360.

Wang, N., Hu, Y., Li, X.-M., Lan, Y., Kang, C., Yan, L., Dou, C., Miao, C., 2024. Enhancing SDGSAT-1 night light images using a panchromatic guidance denoising algorithm. Int. J. Appl. Earth Obs. Geoinf. 128, 103748. https://doi.org/10.1016/j.jag.2024.103748.

Wei, J., Zhong, Y., Chen, A., Tang, H., Li, D., 2025. Space-time cube reveals escalating light pollution in China's national parks: impact of boundary geometry and human activities (1992–2021). Environ. Pollut. 377, 126462. https://doi.org/10.1016/j.envpol.2025.126462.

Wu, B., Wang, Y., Huang, H., Liu, S., Yu, B., 2024. Potential of SDGSAT-1 nighttime light data in extracting urban main roads. Remote Sens. Environ. 315, 114448. https://doi.org/10.1016/j.rse.2024.114448.

Wu, S., Ding, Jiale, Wang, Ruoxu, Wang, Yige, Yin, Ziyu, a, and Du, Z., 2025. Using an attention-based architecture to incorporate context similarity into spatial non-stationarity estimation. Int. J. Geogr. Inf. Sci. 0, 1–24. doi:https://doi.org/10.1080/13658816.2025.2456556.

Xu, W., Zhou, Y., Taubenböck, H., Stokes, E.C., Zhu, Z., Lai, F., Li, X., Zhao, X., 2024. Spatially explicit downscaling and projection of population in mainland China. Sci. Total Environ. 941, 173623. https://doi.org/10.1016/j.scitotenv.2024.173623.

Xue, X., Lin, Y., Zheng, Q., Wang, K., Zhang, J., Deng, J., Abubakar, G.A., Gan, M., 2020. Mapping the fine-scale spatial pattern of artificial light pollution at night in urban environments from the perspective of bird habitats. Sci. Total Environ. 702, 134725. https://doi.org/10.1016/j.scitotenv.2019.134725.

Yang, Q., Yuan, Q., Li, T., Yue, L., 2020. Mapping PM2.5 concentration at high resolution using a cascade random forest based downscaling model: evaluation and application. J. Clean. Prod. 277, 123887. https://doi.org/10.1016/j.jclepro.2020.123887.

Yang, Z., Chen, Y., Guo, G., Zheng, Z., Wu, Z., 2021. Using nighttime light data to identify the structure of polycentric cities and evaluate urban centers. Sci. Total Environ. 780, 146586. https://doi.org/10.1016/j.scitotenv.2021.146586.

Ye, T., Zhao, N., Yang, X., Ouyang, Z., Liu, X., Chen, Q., Hu, K., Yue, W., Qi, J., Li, Z., Jia, P., 2019. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. Sci. Total Environ. 658, 936–946. https://doi.org/10.1016/j.scitotenv.2018.12.276.

Ye, Y., Xue, X., Huang, L., Gan, M., Tong, C., Wang, K., Deng, J., 2020. A new perspective to map the supply and demand of artificial night light based on Loujia1-01 and urban big data. J. Clean. Prod. 276, 123244. https://doi.org/10.1016/j.jclepro.2020.123244.

Ye, Y., Tong, C., Dong, B., Huang, C., Bao, H., Deng, J., 2024. Alleviate light pollution by recognizing urban night-time light control area based on computer vision techniques and remote sensing imagery. Ecol. Indic. 158, 111591. https://doi.org/10.1016/j.ecolind.2024.111591.

Yue, W., Chen, Y., Thy, P.T.M., Fan, P., Liu, Y., Zhang, W., 2021. Identifying urban vitality in metropolitan areas of developing countries from a comparative perspective: Ho Chi Minh City versus Shanghai. Sustain. Cities Soc. 65, 102609. https://doi.org/10.1016/j.scs.2020.102609.

Zhang, H., Wang, S., Liu, K., Li, X., Li, Z., Zhang, X., Liu, B., 2022. Downscaling of AMSR-E soil moisture over North China using random Forest regression. ISPRS Int. J. Geo Inf. 11, 101. https://doi.org/10.3390/ijgi11020101.

Zhang, T., Sun, B., Li, W., Dan, B., Wang, C., 2019. Polycentricity or dispersal? The spatial transformation of metropolitan Shanghai. Cities 95, 102352. https://doi.org/10.1016/j.cities.2019.05.021.

Zhao, F., Chu, C., Liu, R., Peng, Z., Du, Q., Xie, Z., Sun, Z., Zeng, H., Xia, J., 2021. Assessing light pollution using POI and Luojia1-01 night-time imagery from a quantitative perspective at City scale. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 14, 7544–7556. https://doi.org/10.1109/JSTARS.2021.3097320.

Zhou, Y., Gu, S., Ge, Z., Li, Y., 2024. Optimizing urban population mapping grid size selection based on random Forest regression. IEEE Geosci. Remote Sens. Lett. 21, 1–5. https://doi.org/10.1109/LGRS.2024.3354262.

Zhu, Z., Zhou, Y., Seto, K.C., Stokes, E.C., Deng, C., Pickett, S.T.A., Taubenböck, H., 2019. Understanding an urbanizing planet: strategic directions for remote sensing. Remote Sens. Environ. 228, 164–182. https://doi.org/10.1016/j.rse.2019.04.020.

Zielinska-Dabkowska, K.M., Bobkowska, K., 2022. Rethinking sustainable cities at night: paradigm shifts in Urban Design and City lighting. Sustainability 14, 6062. https://doi.org/10.3390/su14106062.