Article:
Lu, B., Luo, S., Yue, P. et al. (2026) Beyond global metrics: A geographically weighted framework for exploring multi-source spatial data validation. Computers Environment and Urban Systems, 128. 102450. ISSN: 0198-9715

https://doi.org/10.1016/j.compenvurbsys.2026.102450

Lu B, Luo S, Yue P, Dong G, **Comber L** (in press). Beyond global metrics: A geographically weighted framework for exploring multi-source spatial data validation. Paper accepted for publication in *Computers, Environment and Urban Systems* (April 2026).

Beyond Global Metrics: A Geographically Weighted Framework for Multi-Source Spatial Data Validation

Abstract

With the increasing availability of multi-source spatial datasets, ensuring data consistency and accuracy has become a critical challenge in spatial analysis. Traditional global indicators such as Mean Error (ME), Mean Absolute Error (MAE), Mean Relative Error (MRE), Root Mean Square Error (RMSE), Correlation Coefficient (CC), and Univariate Linear Regression (ULR) have been widely used to evaluate overall discrepancies between pair-wise datasets. However, these global metrics fail to reveal spatial details of discrepancies between these datasets. To address this limitation, we established a spatially explicit and scale aware validation framework by embedding conventional global indicators within a geographically weighted (GW) modelling context by introducing GW ME, GW MAE, GW MRE, GW RMSE, GW CC, and univariate GWR to provide spatially detailed assessments of discrepancies. The study also emphasizes the crucial role of bandwidth selection in determining the scale of analysis, with large bandwidths smoothing local variations and small bandwidths preserving fine-scale details. Using GPWv4 data, the Gridded Population of China Dataset (GPCD) and the Seventh National Population Census (NPC7) of China uniformly collected in 2020 as a case study, we validate the effectiveness of this approach at the county level. Results show that GW indicators successfully reveal spatial heterogeneities in data discrepancies, which are overlooked by global measures. It is noteworthy that the local validation framework is always straightforward to apply to other types of spatial data sets. Overall, this study provides a systematic and scalable method for multi-source spatial data validation to enhance quality control, improve integration strategies, and facilitate more informed decision-making in spatial analysis and geographic studies.

Keywords: Spatial Heterogeneity, Geographically Weighted Models, Geographically Weighted Regression, Data Consistency, Population Data

1. Introduction

With the rapid advancements in the fields of geographic information science, spatio-temporal big data mining, and remote sensing, the volume and variety of spatial and temporal datasets being generated have increased exponentially (Atluri, Karpatne, & Kumar, 2018). These datasets are often characterized by diverse formats, varying spatial resolutions, and distinct scales, reflecting their disparate sources and mechanisms of generation (Lü, et al., 2019; S. Wang & Li, 2021). Notably, while these datasets may semantically represent similar or even the same geographic phenomena, significant differences exist in terms of their accuracy, spatial distribution, and statistical properties (Luo, et al., 2024). Critical issues and considerable challenges lie in the discrepancies in data characteristics across multiple sources for their utilization. These discrepancies further exacerbate uncertainties in the outcomes of multi-source data analysis, presenting indispensable barriers to achieving reliable and robust conclusions in relative studies (Noardo, 2022; Zhang, et al., 2024).

As one of the most critical data sources in spatial data science, population data has been applied in various fields (Dong, Yang, Cai, & Xu, 2017; Ehrlich, et al., 2018; Lu, Shi, et al., 2024). In particular, the accuracy and availability of population data are crucial for addressing global challenges and advancing smart city initiatives (Karimi, Haghi Kashani, Akbari, & Mahdipour, 2021). In these studies, the inherent discrepancies among population data from multiple sources usually demonstrated noteworthy impacts on the outcomes (Hamdi, et al., 2022; Lu, Shi, et al., 2024), even irreconcilable differences at certain scales (Hanberry, 2022). With the increasing diversity in the methods of population data generation and dissemination, these multi-sourced datasets have become more complex and varied. For instance, census data published by governmental statistical agencies, dynamic (Deville, et al., 2014; Doyle, Hung, Farrell, & McLoone, 2014) and social media data (Alexander, Polimis, & Zagheni, 2022; Z. Wang, et al., 2019), and estimation models based on remote sensing (RS) and geographic information systems (GIS) techniques (Stevens, Gaughan, Linard, & Tatem, 2015; Xie, Weng, & Weng, 2015). These datasets usually exhibit substantial disparities in spatial resolution, temporal coverage, spatio-temporal distributions, and statistical information, which poses great challenges for attempts to integrate and analyze these datasets, particularly in interdisciplinary cases where inconsistencies among data sources might lead to divergent results and heightened uncertainties (Chin, et al., 2025).

To address the inconsistencies and uncertainties associated with multi-source population data, a number of studies have been conducted to highlight the strengths and limitations of widely used gridded population data products. Xu et al. (2021) focused on the 2015 population distribution across the Dian-Gui-Qian region of Southwest China, comparing five popular grid population datasets with census data supplemented with high-resolution Google Earth imagery. Hierink et al. (2022) conducted an evaluation of six popular gridded population datasets within a geographic accessibility modeling framework in sub-Saharan Africa. Hall, Stroh, and Paya (2012) assessed the quality of four global gridded population datasets in terms of global statistical accuracy. Yin et al. (2021) employed statistical methods to cross-compare the consistency and discrepancies among four widely used grid population datasets. These studies collectively underscore the importance and necessity of validating multi-source population datasets to fit them to specific analytical needs, geographic contexts, and case objectives. It is worth noting that these studies predominantly relied on global analysis methods that summarize discrepancies using a single value, including traditional statistical metrics such as mean error (ME), mean relative error (MRE), mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (CC), as well as univariate linear regression (ULR). While these methods are effective in reflecting the overall consistency and discrepancy between different spatial datasets, the global techniques fall short of capturing localized variations in their comparative characteristics, limiting their ability to express the spatial details of differences and analyze spatial heterogeneity in depth (Goodchild, 2004).

To reflect the principles of spatial dependence (Tobler, 1970) and spatial heterogeneity (Goodchild, 2004), geographically weighted regression (GWR) was proposed to highlight spatial heterogeneities in data relationships by conducting place-based weighted regression to produce spatially varying coefficient estimates (Chris Brunsdon, Fotheringham, & Charlton, 1996). Following the development of GWR, a variety of geographically weighted (GW) techniques have been introduced, including GW summary statistics (C. Brunsdon, Fotheringham, & Charlton, 2002), GW principal components analysis (Harris, Brunsdon, & Charlton, 2011) and GW discriminant analysis (Chris Brunsdon, Fotheringham, & Charlton, 2007). In addition, a series of studies have extended the GW framework to spatial error diagnostics, including local correlation estimation (Harris & Juggins, 2011), the comap for non-stationary kriging models (Harris, Chris, & and Charlton, 2013), geographically weighted correspondence error and transition matrices for spatially explicit land cover changes and classification errors (A. Comber & and Tsutsumida, 2023; A. Comber, Chris, Martin, & and Harris, 2017; A. J. Comber, 2013; Münch, 2020). Specifically, Tsutsumida et al. (2019) further developed a set of geographically weighted diagnostics, including local MAE, RMSE, and CC for continuous raster data, providing a reference framework for spatial error structure exploration. Comber et al. (2012) adopted the GWR technique to analyze the local differences between land cover observations of field data and fuzzy classifications from remote sensing images. These GW models form a novel and continually evolving locally technical framework for identifying spatially non-stationary features or patterns (Lu, et al., 2023; Lu, Hu, et al., 2024). In this study, we will apply the localized validation technical framework for multi-source spatial data sets within the GW modeling framework. In this way, the conventional limitation in expressing spatial details of differences between multi-source spatial datasets is necessarily overcome by incorporating GW techniques. This transforms data validation from a global assessment into a spatially explicit diagnostic process, enabling the identification of where discrepancies are concentrated and how their magnitudes change spatially. Importantly, the spatial heterogeneity emphasized in this study is not treated as a causal explanation of data discrepancies, but rather as an inherent property of how differences or consistencies between multi-source spatial datasets are distributed across space. The primary aim is to examine whether these discrepancies are spatially stationary or non-stationary, which cannot be addressed using conventional global indicators.

Specifically, we will take tree data sets widely adopted, i.e. the Gridded Population of the World, Version 4 (GPWv4) dataset (Center For International Earth Science Information Network-CIESIN-Columbia University, 2017) with an output resolution of 30 arc-seconds, gridded population of China dataset (GPCD) available through the Resource and Environment Science and Data Center (RESDC, http://www.resdc.cn/) with an output resolution of 1 kilometer grids, and the 7th National Population Census data of China, released in 2021 which accurately describes population over 2,800 administrative units, for a case study to calculate and analyze the localized consistencies and discrepancies between them. Despite the increasing use of population datasets from different sources, there has been insufficient attention to their inherent differences and how these discrepancies might affect analysis results, biased or even contradictory conclusions (Hanberry, 2022). Traditionally, global statistical indicators can provide an overall sense of agreement but fail to reveal where large local deviations occur, which could critically affect results and interpretations, particularly for local studies (Lu, Dong, Yue, & Qin, 2024). To fill this gap, we propose integrating a local framework that uses GW techniques to validate the spatially varying correspondences or discrepancies among multi-source spatial datasets. Although population data are the primary focus of this study, the framework can also be applied to other types of spatial data, especially those representing the same geographic entities but created through different methods or sources (Uhl & Leyk, 2022).

The remainder of this paper is structured as follows: Section 2 outlines the methodological design and population data sources for the case study; Section 3 presents the validation results based on global and local indicators; Section 4 discusses the implications and limitations of this technical framework; and Section 5 concludes with key findings and anticipates future directions.

2. Methodology and data

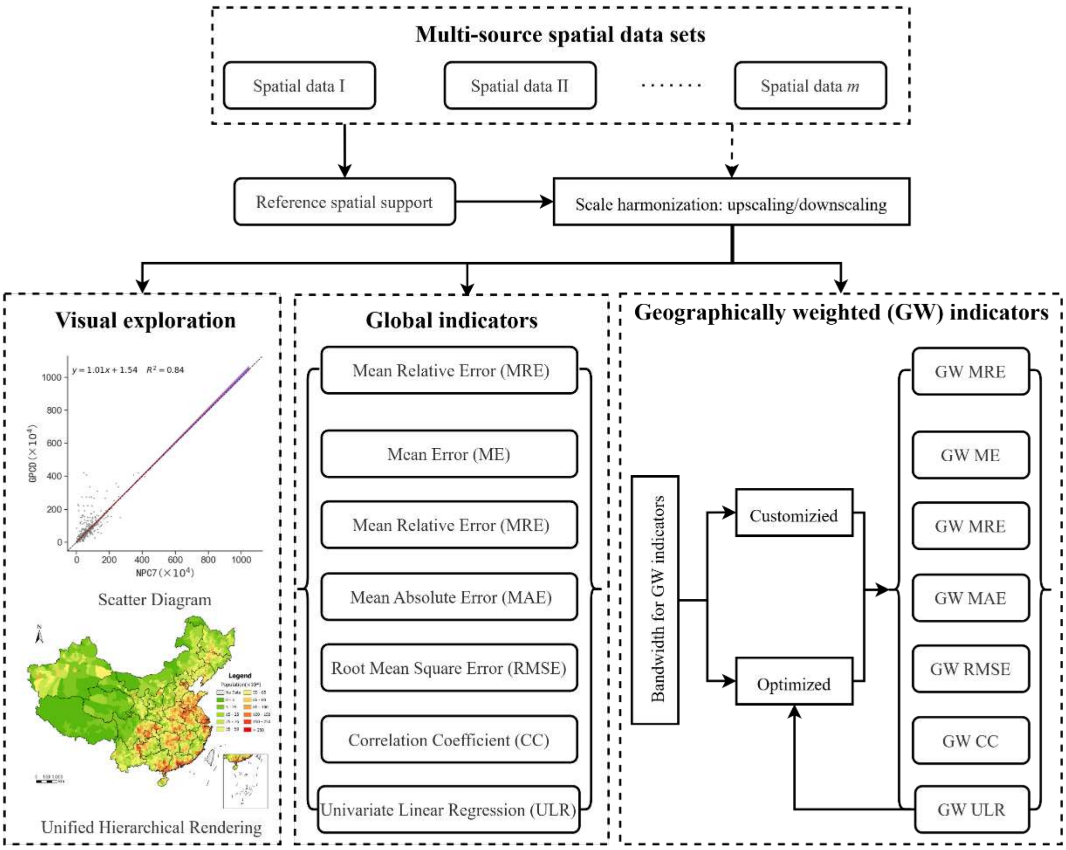


Figure 1 Technical framework of validating multi-source spatial data from both qualitative and quantitative (global and local) perspectives

As illustrated in Figure 1, the validation process begins by standardizing multi-source spatial data in terms of coordinate reference system, spatial and temporal resolutions. Following this, a qualitative description is conducted via two key means:

- ✓ Map comparison – All datasets are displayed using a unified hierarchical rendering to analyze their spatial distribution and patterns.
- ✓ Visual exploration – Scatter diagrams are typically used to visually evaluate the bias among all datasets, often incorporating univariate linear regression lines.

From a scientific research perspective, evaluating the consistency of multi-source spatial data mainly depends on quantitative metrics, which is the primary focus of this study. The following sections will introduce traditional global indicators and local indicators.

2.1 Global and local indicators

In previous studies, global methods, including traditional statistical metrics such as ME, MRE, MAE, RMSE, and CC, as well as ULR, are usually adopted to evaluate the consistency between multi-source spatial data sets (e.g. Hall, et al., 2012; Y. Xu, et al., 2021; Yin, et al., 2021). They can be expressed as follows (Tsutsumida, et al., 2019):

$$\mathrm{ME} = \frac{1}{N}\sum(x_i - y_i) \tag{1}$$

$$\mathrm{MAE} = \frac{1}{N}\sum|x_i - y_i| \tag{2}$$

$$\mathrm{MRE} = \frac{1}{N}\sum\frac{|x_i - y_i|}{y_i} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum(x_i - y_i)^2} \tag{4}$$

$$\mathrm{CC} = \frac{\mathrm{cov}(\mathbf{x},\mathbf{y})}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}} \tag{5}$$

$$y_i = \beta x_i + \varepsilon_i \tag{6}$$

where x_i, y_i are observations of pair-wise multi-source datasets at spatial unit *i*, *N* is the number of observations in each data set, $\mathrm{cov}(\mathbf{x},\mathbf{y})$ is the covariance between the pair-wise multi-source datasets, σ_x, σ_y are the standard deviations of each data set, β and ε_i is the coefficient and random error of the ULR. Note that the ULR slope is equivalent to the Pearson CC when both the dependent and independent variables are standardized to zero mean and unit variance. While these global indicators are effective in reflecting the overall consistency and discrepancy between pair-wise multi-source datasets, they naturally fall short of capturing localized variations to express the spatial details of differences.

GW summary statistics (GWSS, C. Brunsdon, et al., 2002) and GWR primarily provide foundational approaches for locally evaluating multi-source datasets. GWSS applies the GW modelling framework to compute summary statistics, including mean, standard deviation, covariance, and CC. Effectively the GW framework allows local versions of the above global indicators of accuracy to be calculated, from weighted data subsets to generate spatially distributed accuracy or correspondence measures. Specifically, unlike traditional global correlation analysis, GW CC can provide spatially varying correlations between spatial data sets from different sources by incorporating spatial heterogeneity and a location-wise weighting scheme. Its expression can be defined as follows,

$$CC(u_i, v_i) = \frac{Cov_{(\mathbf{x},\mathbf{y})}(u_i,v_i)}{SD_{\mathbf{x}}(u_i,v_i) \times SD_{\mathbf{y}}(u_i,v_i)} \tag{7}$$

where (u_i, v_i) is the coordinate of spatial unit *i*, $Cov_{(\mathbf{x},\mathbf{y})}(u_i, v_i)$ is the GW covariance between the pair-wise multi-source datasets, $SD_{\mathbf{x}}(u_i, v_i)$ and $SD_{\mathbf{y}}(u_i, v_i)$ are the GW standard deviations of each data set at spatial unit i. They can be formulized as eq. (8) and eq. (9)

$$Cov_{(\mathbf{x},\mathbf{y})}(u_i, v_i) = \frac{\sum_j w_{ij}(x_j - \bar{x}_i)(y_j - \bar{y}_i)}{\sum_j w_{ij}} \tag{8}$$

$$SD_{\mathbf{x}}(u_i, v_i) = \frac{\sqrt{\sum_j \left(x_j - \bar{x}(u_i, v_i)\right)^2 w_{ij}}}{\sum_j w_{ij}} \tag{9}$$

where w_{ij} is the spatial weight of each observation j for the calibration location i, which can be calculated by a kernel function based on the distance between location i and j (Gollini, Lu, Charlton, Brunsdon, & Harris, 2015; Lu, Harris, Charlton, & Brunsdon, 2014). $\bar{x}(u_i, v_i)$ is the GW mean of data set **x**, and it can be calculated via eq. (10).

$$\bar{x}(u_i, v_i) = \frac{\sum_j x_j w_{ij}}{\sum_j w_{ij}} \tag{10}$$

Meanwhile, GWR quantifies the local relationship between one or more independent variables and the dependent variable. Therefore, univariate GWR can be used to replace the ULR for locally evaluating the observations of pair-wise spatial data sets from different sources. Accordingly, the univariate GWR can be expressed as eq. (11):

$$y_i = \beta(u_i, v_i) x_i + \varepsilon_i \tag{11}$$

where $\beta(u_i, v_i)$ is the localized coefficient at location i. Unlike global ULR model, algebraic equivalence between univariate GWR slopes and GWCC does not strictly hold under different GW estimators, i.e. weighted least squares and weighted covariance-variance terms. As a result, although the two measures are closely related, they are not guaranteed to be numerically identical at each location, particularly when spatial weights vary across space. They provide complementary GW perspectives on local association patterns between the pair-wise spatial data sets.

Building upon the aforementioned concepts of GW models, we further propose the local versions of key error indicators, including GW ME, GW MRE, GW MAE, and GW RMSE. Their respective mathematical formulations are expressed as follows (Tsutsumida, et al., 2019):

$$\mathrm{ME}_i = \frac{\sum_{j=1}^{n} w_{ij}(x_i - y_i)}{\sum_{j=1}^{n} w_{ij}} \tag{12}$$

$$\mathrm{MAE}_i = \frac{\sum_{j=1}^{n} w_{ij}|x_i - y_i|}{\sum_{j=1}^{n} w_{ij}} \tag{13}$$

$$\mathrm{MRE}_i = \frac{\sum_{j=1}^{n} w_{ij}|x_i - y_i|/y_i}{\sum_{j=1}^{n} w_{ij}} \tag{14}$$

$$\mathrm{RMSE}_i = \sqrt{\frac{\sum_{j=1}^{n} w_{ij}(x_i - y_i)^2}{\sum_{j=1}^{n} w_{ij}}} \tag{15}$$

When applying these local indicators, bandwidth plays a crucial role by defining the spatial scale over which local information is aggregated, as different bandwidths correspond to different levels of spatial generalization. It defines the spatial scale over which GW indicators are computed. In this study, bandwidth is explicitly conceptualized as a scale parameter rather than as a tuning parameter aimed at optimizing individual validation metrics.. To ensure scale consistency across indicators, it is recommended to use a consistent bandwidth for all calculations to ensure meaningful comparisons across different indicators. Among all the GW indicators, the GWR technique offers well-established bandwidth optimization methods, such as the cross-validation (CV) approach and the Akaike Information Criterion (AIC) based selection method (Lu, Hu, et al., 2024). Therefore, the optimized bandwidth obtained through the univariate GWR model can be used as a reference bandwidth for computing all the GW indicators. Note, however, this optimized bandwidth is not assumed to be theoretically optimal for all the GW indicators; instead, it is adopted as a data-adaptive and internally consistent reference scale at which all GW metrics are initially evaluated. On the other hand, users may also customize the bandwidth to facilitate localized comparisons at different spatial scales, as the bandwidth will directly affect the varying degrees of all the GW indicators, i.e. the larger the bandwidth, the smoother the GW indicators change, and *vice versa*.

2.2 Data for Case Study

In this study, we conduct a case study within the Chinese Mainland to locally validate the 2020 GPCD and GPWv4 dataset with the 7^{th} National Population Census data (NPC7). The analysis is performed at the county-level, covering over 2,800 administrative units. Both global and local indicators are calculated to evaluate the correlation and discrepancies between these datasets.

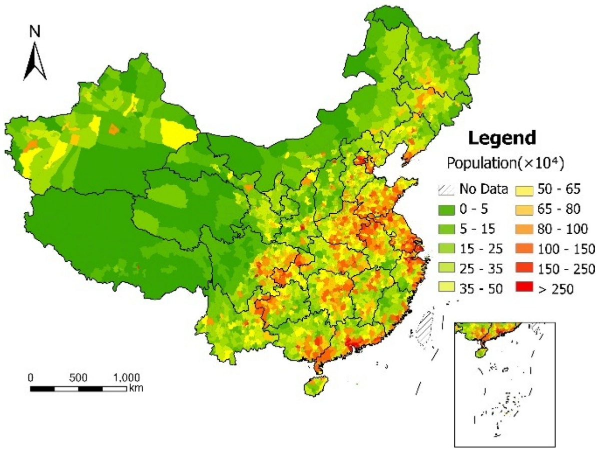
The GPCD was released by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (IGSNRR, CAS), and is available through the Resource and Environment Science and Data Center (RESDC, http://www.resdc.cn/). In this study, the population collected in 2020 was disaggregated onto a spatial grid using a multi-factor weighting method, which considered factors closely related to population, such as land use types, nighttime light intensity, and residential settlement density (X. Xu, 2020).

The GPWv4 dataset was developed by the Center for International Earth Science Information Network (CIESIN) at Columbia University and the Socioeconomic Data and Applications Center (SEDAC) at NASA (Center For International Earth Science Information Network-CIESIN-Columbia University, 2017). It is publicly available in raster format on the official website of SEDAC (https://sedac.ciesin.columbia.edu/) and the data set of 2020 was accessed. In this dataset, population estimates were allocated to a 30-arc-second grid using an areal-weighting method, excluding lakes, rivers, and ice-covered areas.

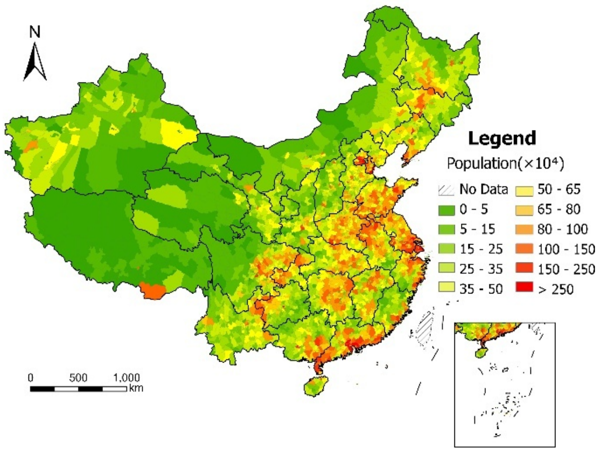
The National Population Census is conducted by the Chinese government and is publicly available on the official website of the National Bureau of Statistics of China (https://www.stats.gov.cn/). This study utilizes data from the 7th National Population Census, conducted in 2020.

To facilitate validation, we initially harmonized these three datasets at the county-level. Figure 2(a) displays a map of China's provincial administrative divisions, and Figures 2(b-d) present the maps of these three population datasets with a unified hierarchical rendering. Visually, the population is predominantly concentrated in the southeastern counties, while the northwestern regions, particularly Tibet, Qinghai, and Inner Mongolia, apparently exhibit low population densities. The spatial distributions of the three datasets are generally consistent with each other. However, it is noteworthy that in low- or medium-density population provinces, such as Heilongjiang and Jilin, the population estimates from both GPWv4 and GPCD often exhibit a certain degree of overestimation compared to NPC7. Furthermore, discrepancies also exist between GPWv4 and GPCD themselves, particularly in provinces like Xinjiang and Tibet.

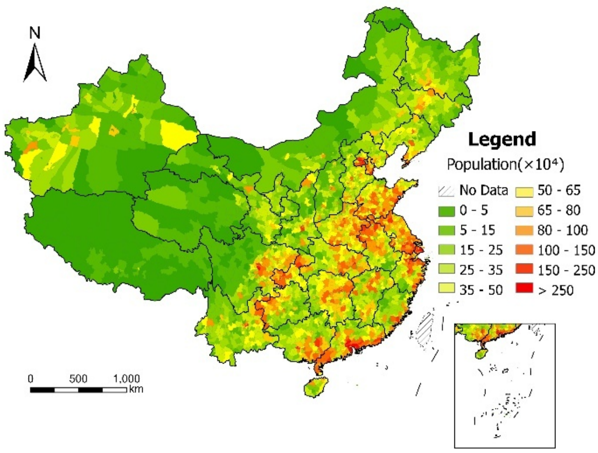
(a) China's provincial administrative divisions



(b) GPCD



(c) GPWv4

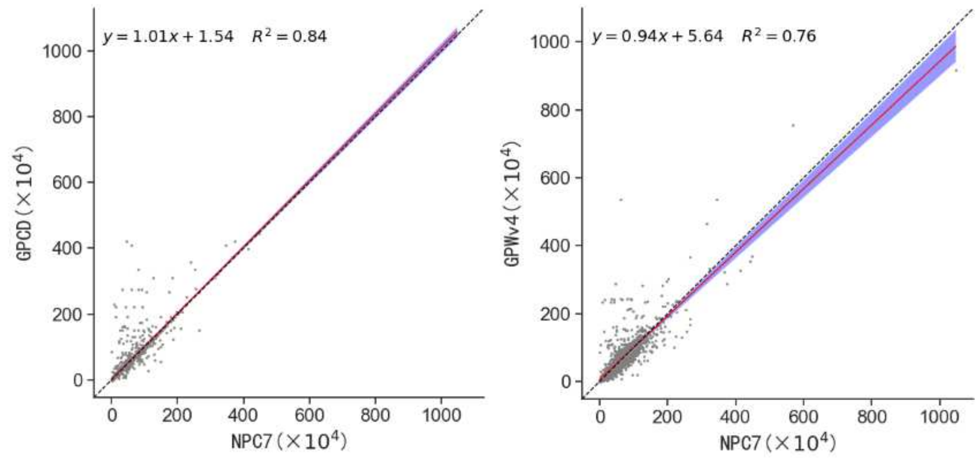


(d) NPC7

Figure 2 Maps of multi-source population data sets

Considering the authoritative status and high accuracy of the 7th National Population Census (NPC7), we upscale the other two datasets, GPCD and GPWv4, to match the spatial resolution and administrative boundaries of NPC7 at the county level. This upscaling process serves as a spatial harmonization step to ensure that all datasets share a common spatial support, which is essential for meaningful validation. Although spatial aggregation may reduce fine-scale variability inherent in gridded datasets, it allows discrepancies to be evaluated consistently at the administrative-unit level. This alignment ensures that subsequent comparisons are conducted on a consistent spatial basis. NPC7 is then used as the reference dataset for both global and local evaluations, providing a benchmark against which the correspondence and consistency of GPCD and GPWv4 can be assessed. Moreover, indirect comparisons between GPCD and GPWv4, facilitated through their respective differences with NPC7, also offer valuable insights into their mutual discrepancies and underlying structural variations.

Figure 3(a-c) displays pairwise scatterplots for GPCD, GPWv4, and NPC7. Each plot features the fitted regression line in red and the identity line as a grey dashed line. These results further support the maps presented in Figure 2. The slopes of the fitted lines, measuring 1.01, and 0.94, respectively, indicate that the datasets are generally consistent with one another overall. However, the R^2 values of 0.84 and 0.76, respectively, suggest notable deviations in the finer details, highlighting the need for further exploration using quantitative indicators to understand their correspondences better.



(a) GPCD *vs.* NPC7 (b)GPWv4 *vs.* NPC7

Figure 3 Pairwise scatterplots between GPCD, GPWv4, and NPC7

It should be noted that the purpose of this study is not to preserve or assess grid-level population patterns, but to evaluate spatially varying consistencies among multi-source datasets under a unified spatial framework. Therefore, any aggregation-related effects introduced by upscaling or downscaling are treated as a controlled and necessary component of the validation design rather than a source of analytical bias.

3. Results of the case study

3.1 Global indicators

The results of six global indicators, i.e. ME, MAE, MRE, RMSE, CC, and the coefficient estimate (β) from the ULR are reported in Table 1, to evaluate the consistency and discrepancies among GPCD, GPWv4 and NPC7 at the county-level. From them, we can see that:

- ✓ ME quantifies the average discrepancy between two datasets, accounting for both the magnitude and direction of the error. The ME of NPC7-GPCD and NPC7-GPWv4 are less than 0, which indicates both GPCD and GPWv4 tend to underestimate NPC7, with GPWv4 showing a larger average underestimation.

- ✓ MAE, MRE, and RMSE provide different perspectives on absolute and relative differences between two datasets. RMSE, in particular, squares the error terms, amplifying the impact of larger discrepancies, and suggesting substantial errors between two datasets. GPWv4 has a considerably larger MAE and RMSE when compared to NPC7 than GPCD does, indicating larger absolute errors overall.
- ✓ All pairs show strong positive linear correlations (CC > 0.85), meaning they generally trend in the same direction, even if their absolute values differ. Besides, all the coefficient estimates β being close to 1, especially for NPC7-GPCD, suggest significant linear relationships but also underestimations or overestimations between different datasets.

Table 1 Global evaluation indicators of the differences between three data sets

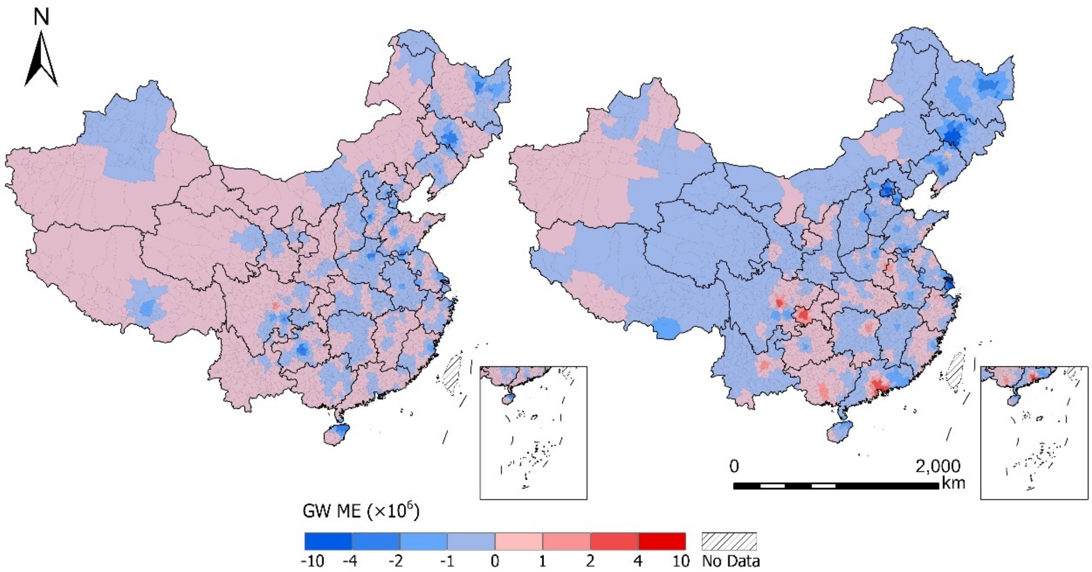
Datasets	ME($\times10^4$)	MAE($\times10^4$)	MRE($\times10^4$)	RMSE($\times10^4$)	CC	β
NPC7-GPCD	-1.95	4.11	0.97	20.23	0.92	1.01
NPC7-GPWv4	-2.59	10.44	0.23	24.38	0.87	0.94
GPWv4-GPCD	0.65	9.31	2.60	17.98	0.94	0.96

Although these global indicators offer insights into the overall differences and consistency between the three datasets, they do not identify specific locations or units where discrepancies arise, nor do they account for spatial variations in these discrepancies across different regions. This limitation underscores the need for local indicators to better comprehend the spatially varying correspondences and inconsistencies between different datasets.

3.2 Local indicators

3.2.1 Local indicators calculated with optimized bandwidth from univariate GWR

In the first place, we employed the AIC approach and the adaptive Bi-square kernel function to optimize the bandwidth for the univariate GWR model (see the details in Lu et al. 2014). Once the optimal bandwidth is determined, the same bandwidth and kernel function are consistently applied to compute other localized indicators, including GW ME, GW MAE, GW MRE, GWRMSE, and GW CC. In this case, the optimal bandwidths were found to be 20 (the number of nearest neighbors) for GPCD and 26 (the number of nearest neighbors) for GPWv4, which are then uniformly applied to their respective spatial weight calculations.

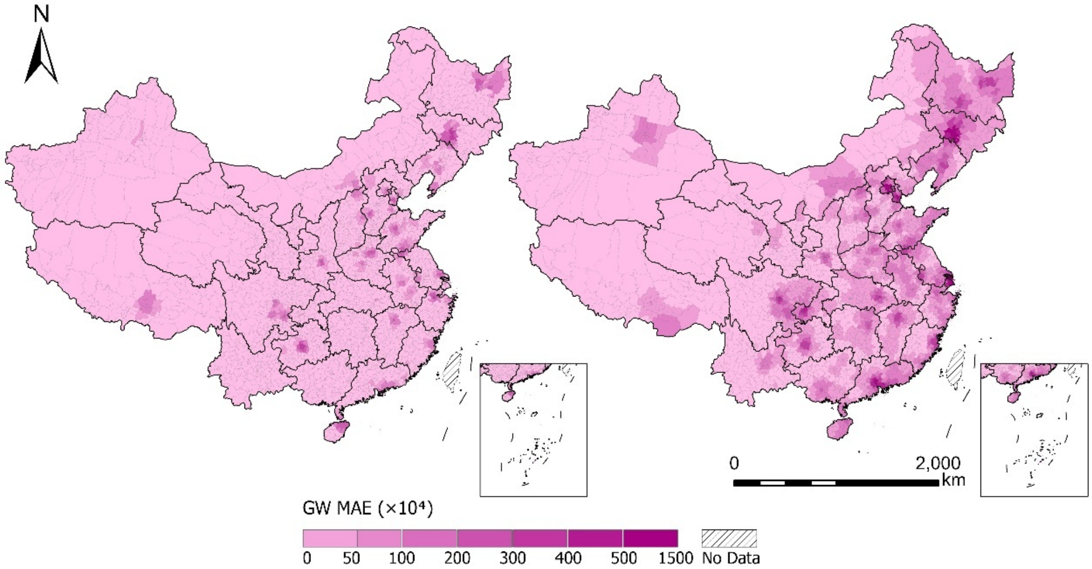


(a) NPC7-GPCD (b) NPC7-GPWv4

Figure 4 GW ME results with optimized bandwidth from univariate GWR

Figure 4(a-b) presents the spatial distribution of GWME for NPC7-GPCD and NPC7-GPWv4. A simultaneous comparison of Figure 4(a) and Figure 4(b) indicates that, when compared to NPC7, GPCD exhibits widespread local overestimation while GPWv4 shows underestimation. Besides, an individual analysis of each figure indicates that large errors are concentrated in provincial capital cities and economically developed regions, underscoring the spatial heterogeneity of the sign and magnitude of discrepancies between the two datasets. For instance, taking Figure 4(b) as an example, in Guiyang (Guizhou Province) and Changchun (Jilin Province), GPWv4 underestimates population counts compared to NPC7. In contrast, positive GW MEs in Chengdu (Sichuan Province) and Changsha (Hunan Province) indicate the overestimation of GPWv4.

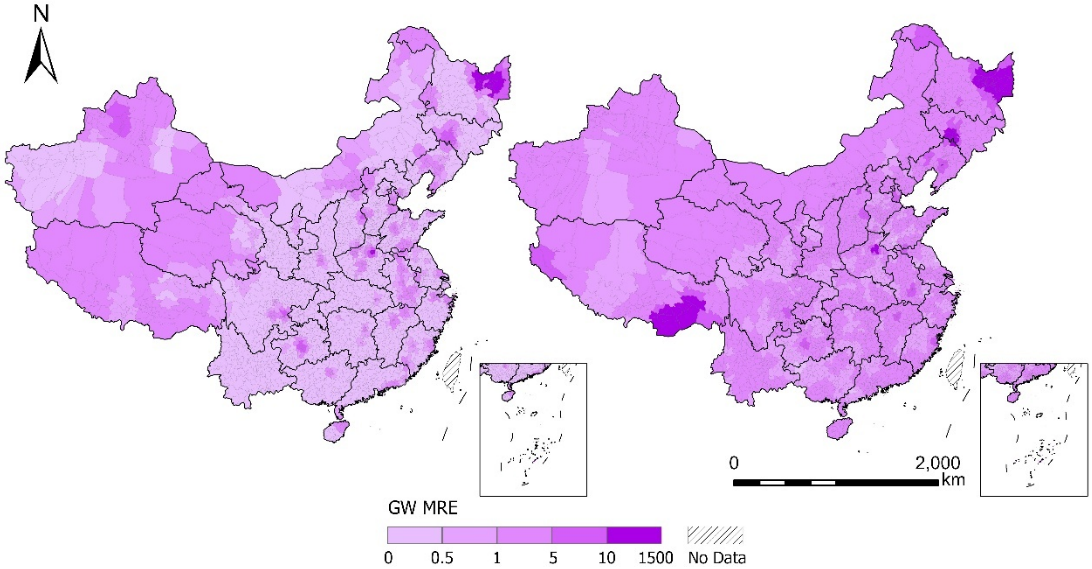
Figure 5(a-b) presents the spatial distribution of GWMAE for NPC7-GPCD and NPC7-GPWv4. Unlike GWME, GWMAE eliminates the influence of positive and negative signs, offering a clearer and more straightforward representation of the magnitude of discrepancies between different datasets. The GWMAE results, when compared to NPC7, show for both GPCD and GPWv4 that major errors are concentrated in provincial capital cities, a pattern that is highly consistent with the findings from GWME.



(a) NPC7-GPCD (b) NPC7-GPWv4

Figure 5 GW MAE results with optimized bandwidth from univariate GWR

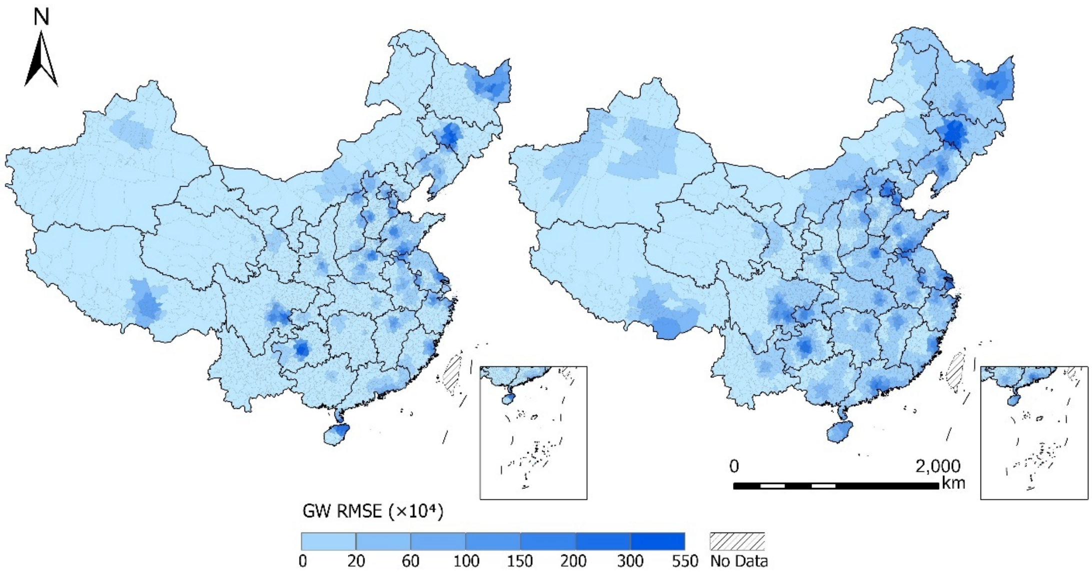
Figure 6(a-b) presents the spatial distribution of GWMRE for NPC7-GPCD and NPC7-GPWv4. Unlike GWME and GWMAE, GWMRE calculates the relative proportion of local errors, making it a more effective metric for comparing discrepancies across both densely and sparsely populated areas, as it mitigates biases related to variations in population size. Comparing Figure 6(a) and Figure 6(b) reveals that GPCD exhibits a larger spatial extent of relatively low GWMRE with their respective optimal bandwidths. Comparing Figures 4 and 5, and specifically considering the NPC7-GPWv4 results in these figures, it can be observed that, some areas with high GW MEs and GW MAEs, such as Changsha (Hunan Province) and Wuhan (Hubei Province), show low GW MREs. In contrast, certain regions with relatively low GW MEs and GW MAEs, like northeastern counties in Heilongjiang Province and Zhengzhou (Henan Province), exhibit high GW MREs. In this regard, GW MRE offers a more nuanced and fair assessment of errors by considering variations in population density.



(a) NPC7-GPCD (b) NPC7-GPWv4

Figure 6 GW MRE results with optimized bandwidth from univariate GWR

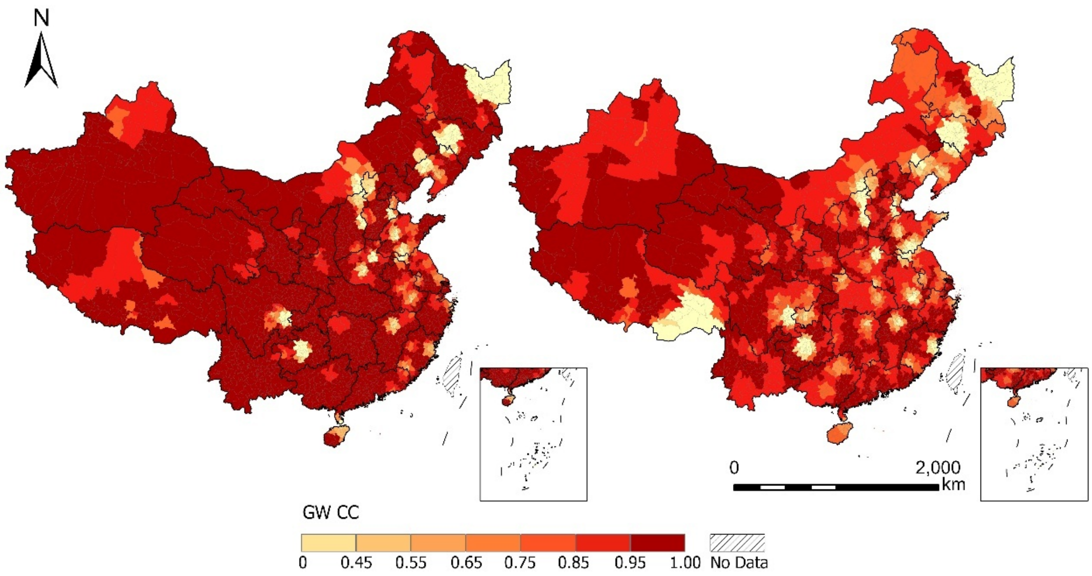
Figure 7(a-b) presents the spatial distribution of GW RMSE for NPC7-GPCD and NPC7-GPWv4. GW RMSE identifies local areas where substantial errors occur, providing a more detailed understanding of regional discrepancies. Compared to the GW MAE presented in Figure 5, GW RMSE highlights those extreme discrepancies in Beijing and Guiyang (Guizhou Province), both for NPC7-GPCD and NPC7-GPWv4.



(a) NPC7-GPCD (b) NPC7-GPWv4

Figure 7 GW RMSE results with optimized bandwidth from univariate GWR

Figure 8(a-b) presents the spatial distribution of GWCC for NPC7-GPCD and NPC7-GPWv4. Overall, the results show apparent variations in GWCC across different provinces and regions. Visually stronger correlations could be observed in the western regions than in the eastern areas, although they are economically developed and population-concentrated. Specifically, low GWCCs, indicating weak correlations can be observed in Heilongjiang, Jilin, Sichuan, Gansu and Tibet. The observed inconsistency can be attributed to the spatialization methods of GPWv4, which used the area-weighting method by assuming an even distribution within each grid. This assumption works effectively in most of regions, but might lead to substantial disparities in areas with uneven distributions, like urban-rural junctions.

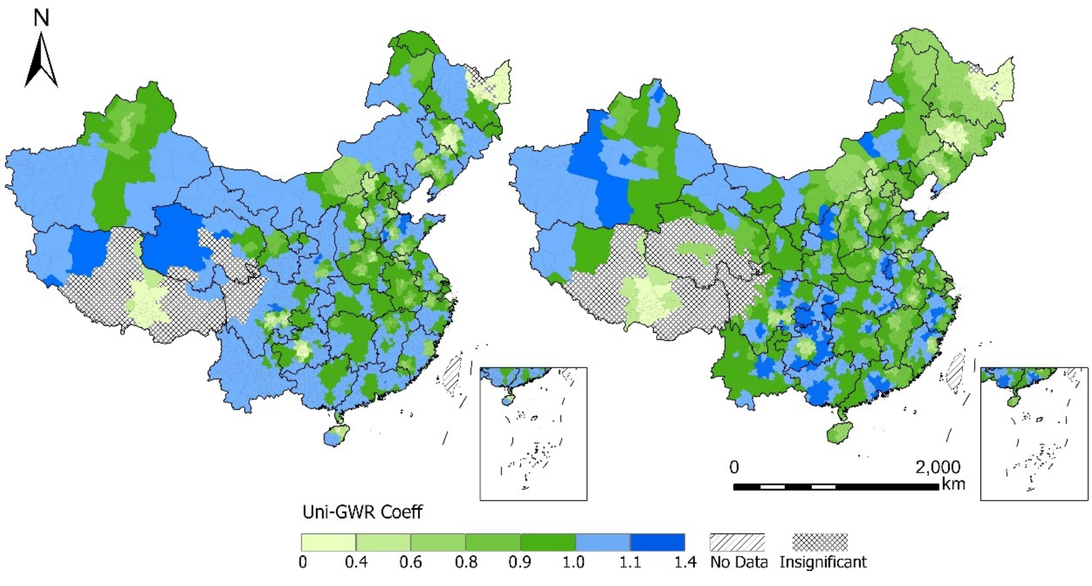


(a) NPC7-GPCD (b) NPC7-GPWv4

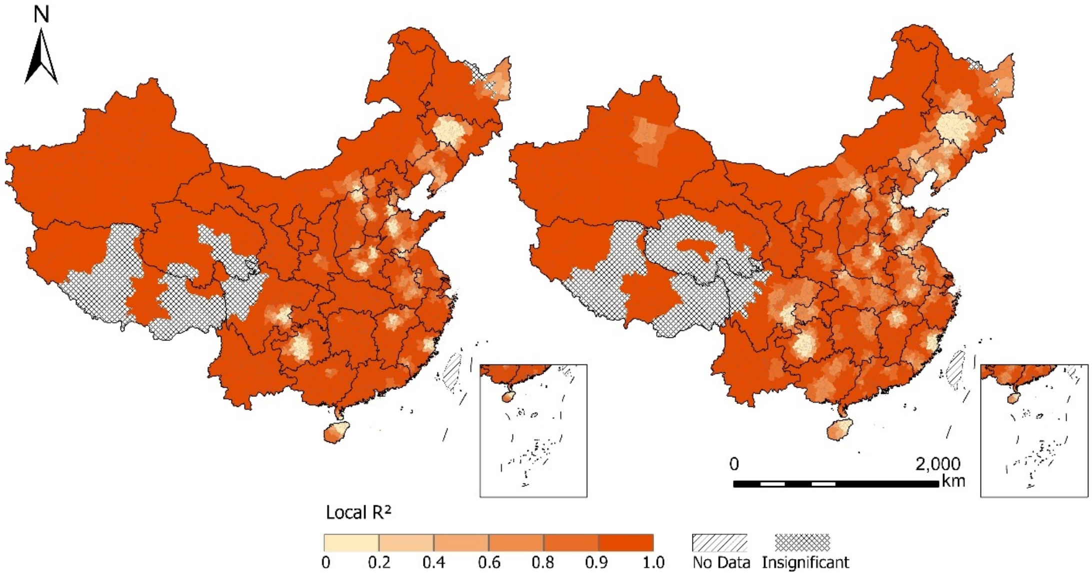
Figure 8 GW CC results with optimized bandwidth from univariate GWR

Figure 9(a-d) presents the spatial distribution of univariate GWR coefficients and the corresponding local R^2 values, with local t-test results used to show the significance of the coefficient estimates.

Compared to the ULR, the univariate GWR model provides insights into identifying regions where GPCD or GPWv4 either underestimates or overestimates population sizes compared to NPC7. Figure 9(a-b) reveals that in most regions, GPCD exhibits a tendency to overestimate population sizes while GPWv4 exhibits a certain degree of underestimation, which align with their respective trend observed in ULR. However, taking GPWv4 as an example, overestimation is also evident in Xinjiang, Inner Mongolia, Jiangxi, Sichuan, and Fujian, where the GWR coefficients exceed 1. Furthermore, the local R^2 values show the goodness of fit of each localized GWR calibration. In most regions, the local R^2 values exceed 0.8, signifying a good local fit and demonstrating that the univariate GWR model outperforms effectively by capturing the spatially varying relationships between the two datasets.



Coefficient estimates of the univariate GWR between NPC7, and (a) GPCD, (b) GPWv4



Local R^2 values of the univariate GWR between NPC7 and (c) GPCD, (d) GPWv4

Figure 9 Univariate GWR results with optimized bandwidth from univariate GWR

Overall, the local indicators offer more detailed and nuanced assessments of discrepancies and consistency by considering spatial variations in population sizes across different data sets. This is particularly valuable for identifying the underlying distributions and causes of discrepancies between NPC7 and GPCD, GPWv4, respectively. Understanding these regional differences aids in pinpointing potential sources of errors and guiding decisions on the selection of multi-source data sets for future studies.

3.2.2 Local indicators calculated with customized bandwidth

The bandwidth is crucial in determining how spatial weights decay with increasing distance. A small bandwidth causes weights to decline more rapidly, whereas a larger bandwidth results in a more gradual decrease (A. Comber, et al., 2023), leading to different variations observed in the calculated local indicators. To further assess the robustness of the proposed GW framework and to illustrate how validation outcomes vary across spatial scales, additional analyses were conducted using customized bandwidths. In this section, we applied two different bandwidths, i.e. 50 and 100 (number of nearest neighbors), to assess the sensitivity of validation results to spatial scale. Note, however, the selection of these bandwidths is intended to examine whether key spatial patterns of agreement and discrepancy persist under meaningful changes in different bandwidths, rather than to exhaustively enumerate all possible scales.

Figures 10- 13 illustrate the spatial distribution of GW ME, GW MAE, GW MRE and GW RMSE calculated with the two bandwidths. As the bandwidth increases, the local indicators become progressively smoother, and validation results begin to preserve local characteristics in a broader scale. A clear example of this pattern is observed in the north-eastern provinces, when comparing Figure 5(a) (optimal bandwidth: 26) with Figures 11(a) and 11(c) (customized bandwidths of 50 and 100, respectively). With the optimal bandwidth, substantial discrepancies are highly localized within Heilongjiang and Jilin. However, as the bandwidth increases to 50 and 100, the differences between the two datasets become much smoother, and the entire northeastern region begins to exhibit large discrepancies.

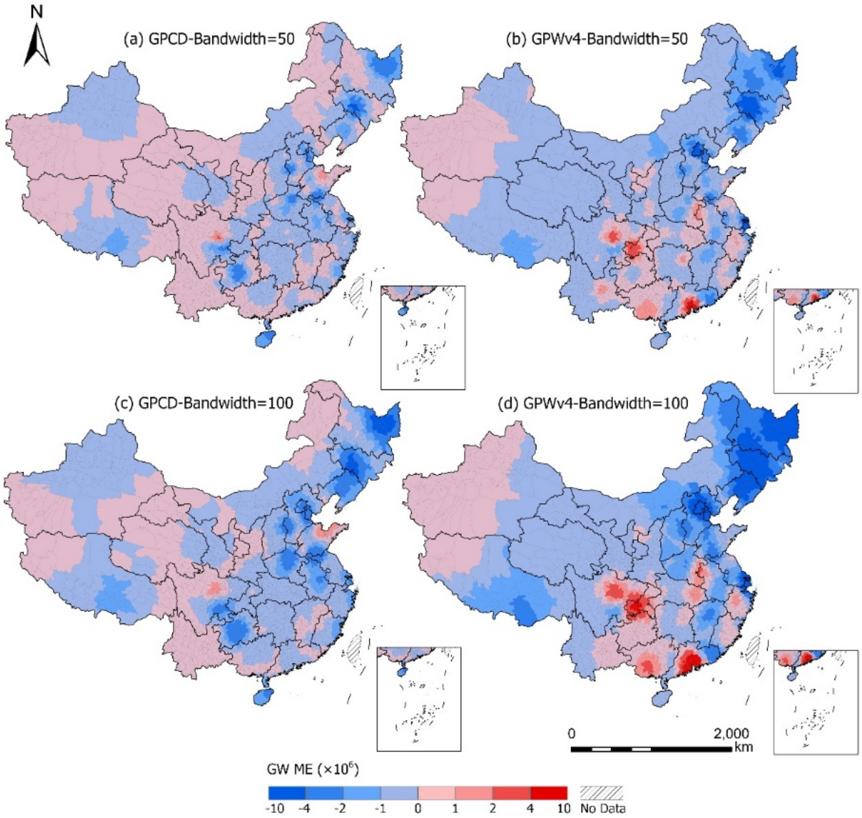


Figure 10 GW ME results with customized bandwidths

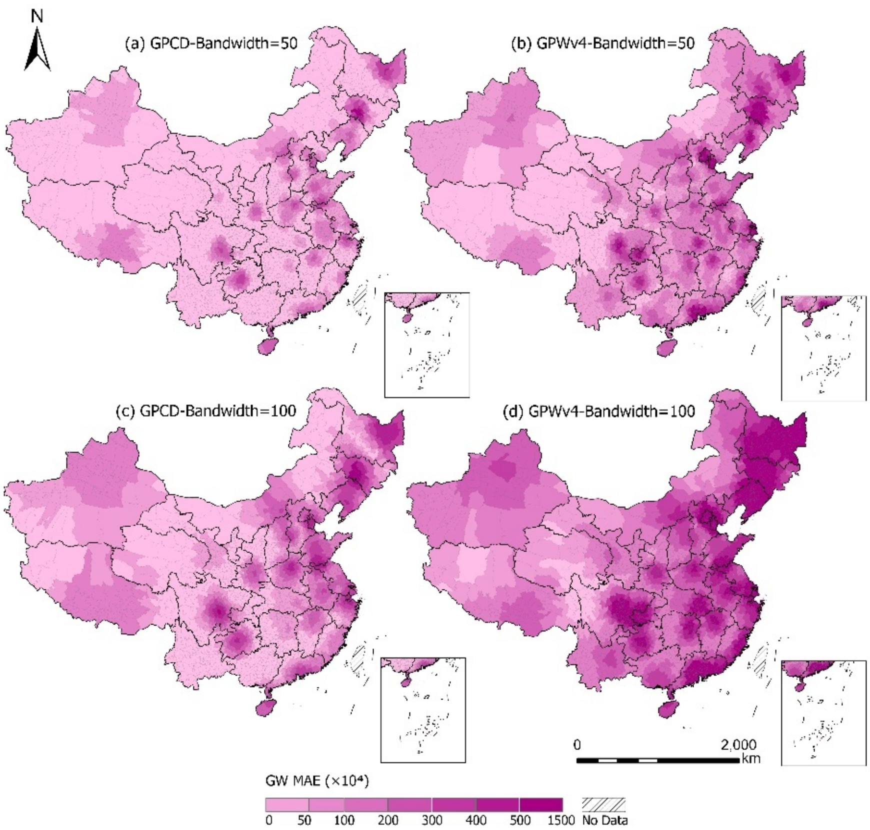


Figure 11 GW MAE results with customized bandwidths

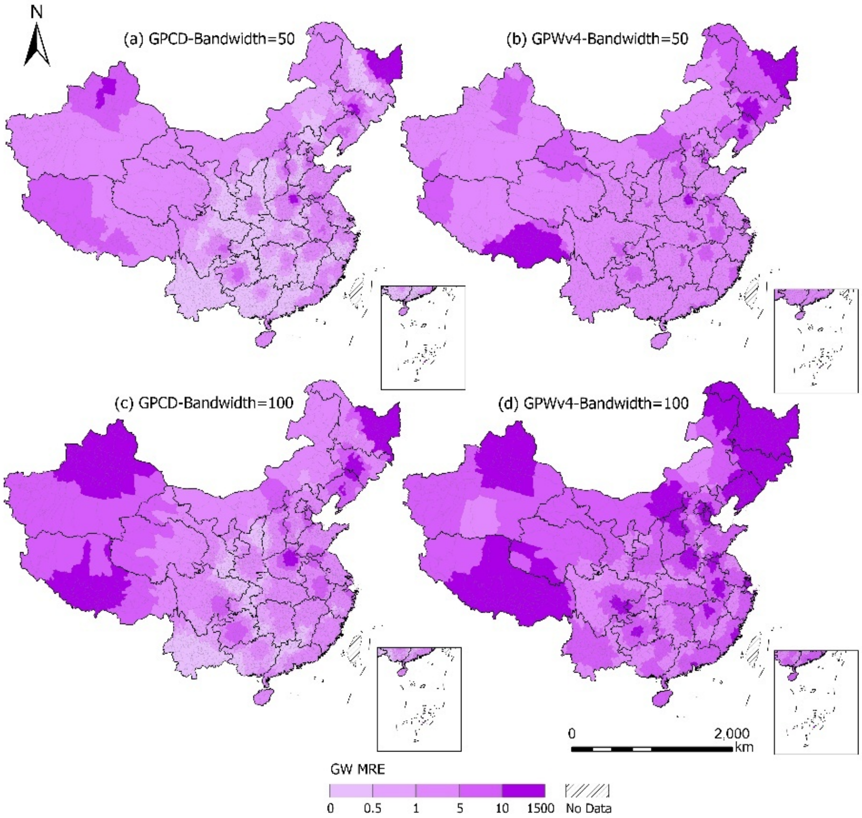


Figure 12 GW MRE results with customized bandwidths

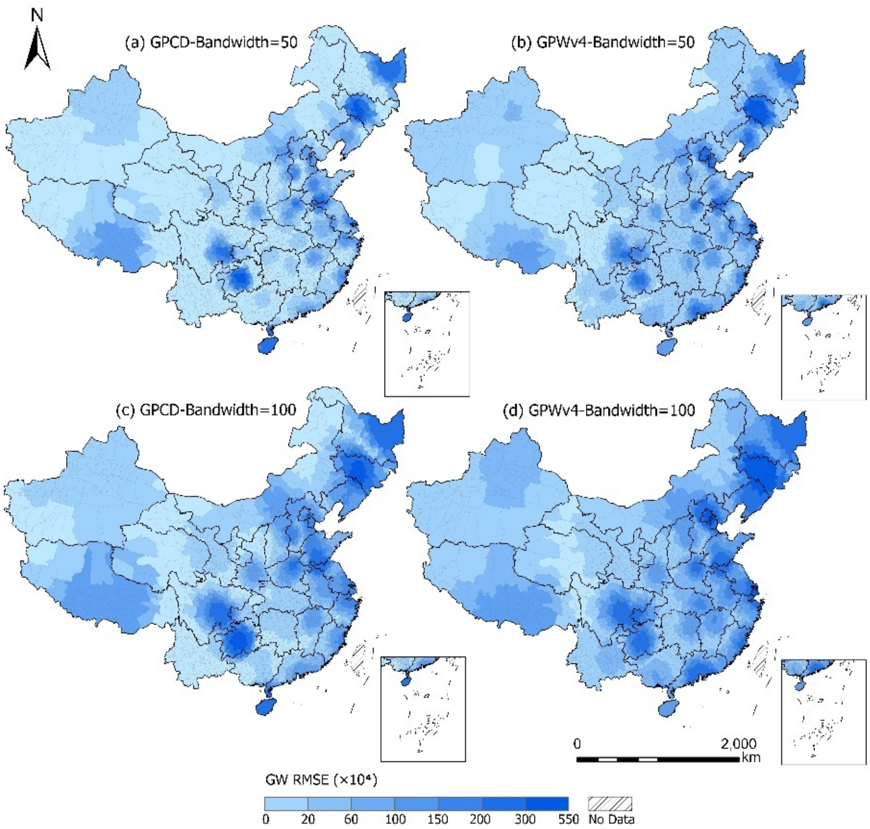


Figure 13 GW RMSE results with customized bandwidths

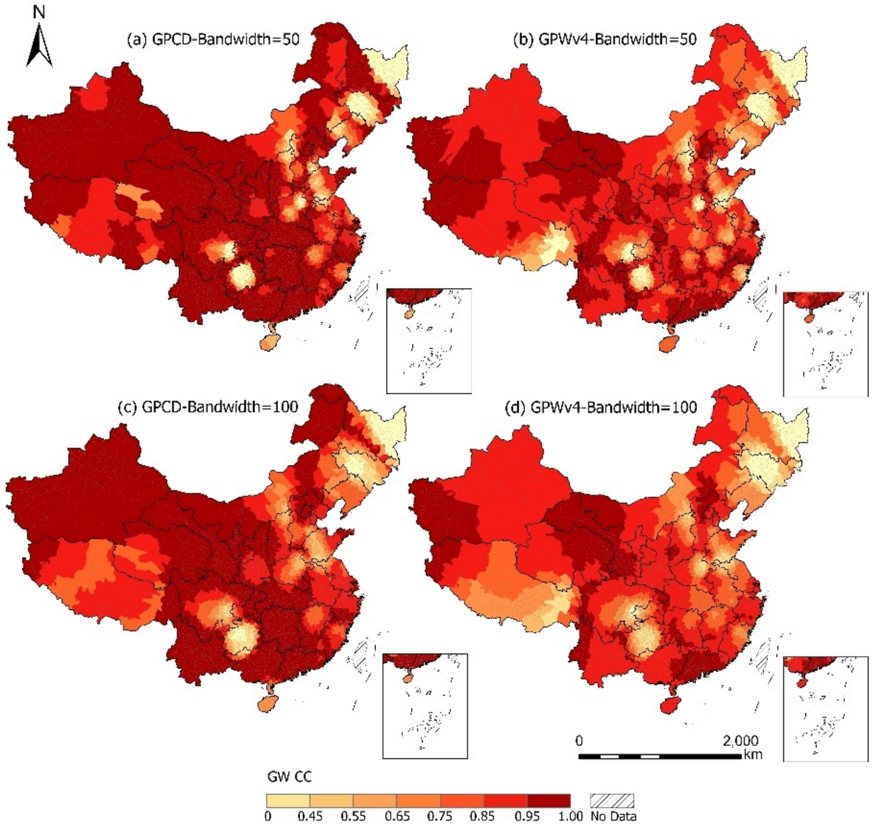


Figure 14 GW CC results with customized bandwidths

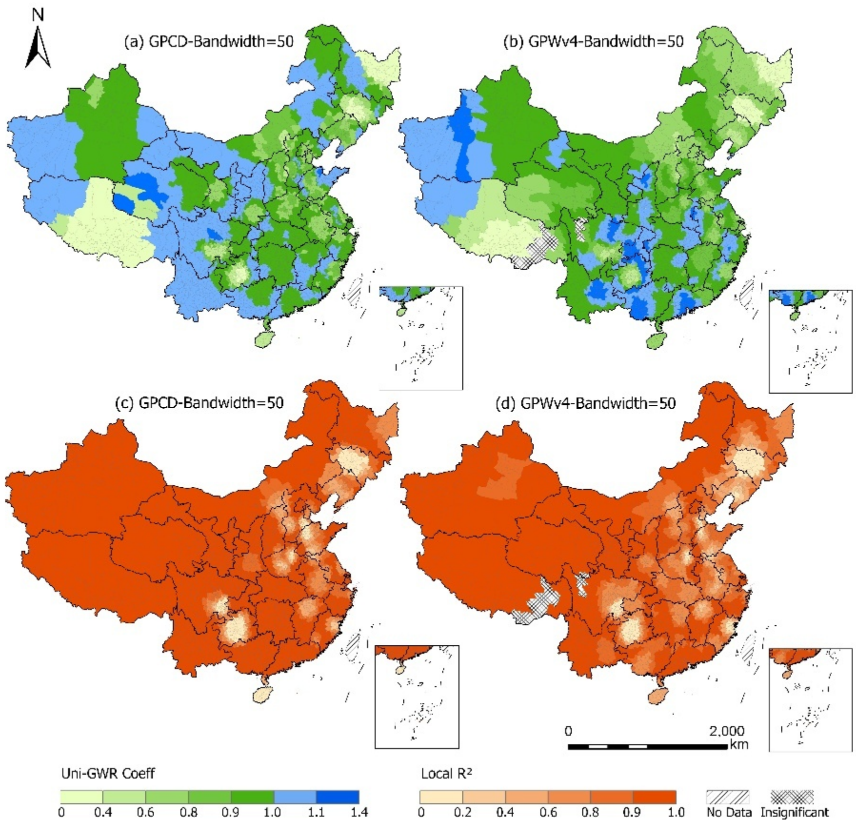


Figure 15 Univariate GWR results with customized bandwidth: 50

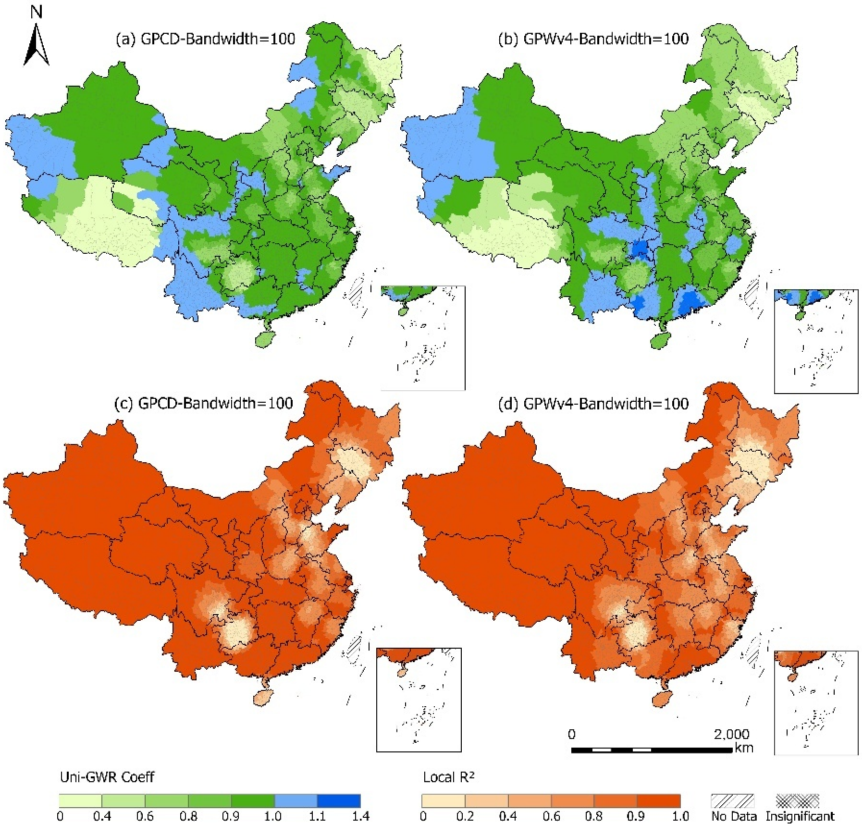


Figure 16 Univariate GWR results with customized bandwidth: 100

Figures 14 -16 illustrate the results of the GW CC and the univariate GWR with the customized bandwidths. The results further reveal distinctive patterns in how bandwidth influences local estimations of the GW CC and univariate GWR models. For GWCC, increasing the bandwidth leads to a smoother correlation estimation, gradually reducing localized weak correlations and masking the spatial heterogeneity of the correlations between these two data sets. A similar effect is observed in Figures 15 and 16, where larger bandwidths cause the coefficient estimates to stabilize, allowing for a better understanding of generalized overestimation or underestimation patterns at either a prefectural or provincial scale. For example, as shown in Figures 14-16, it is evident that, compared to NPC7, either GPCD or GPWv4 is seriously inconsistent in the northeastern provinces of China, where both of them exhibit a severe underestimation of population sizes. However, for regions such as Yunnan and Xinjiang, GPCD exhibits a high degree of consistency with NPC7, markedly outperforming GPWv4. These findings may suggest that both GPCD and GPWv4 are not equivalent or interchangeable if a choice must be made when conducting comparative studies in regions like Heilongjiang, whereas GPCD might be a better alternation for research in regions such as Yunnan and Xinjiang.

These findings highlight a fundamental trade-off in validating multi-source spatial data with local indicators: conducting large-scale comparability with a large bandwidth or obtaining a fine-scale validation with a small bandwidth. Consequently, by underscoring the importance of bandwidth, bandwidth selection should be tailored to the specific analytical objectives, striking a balance between preserving localized spatial characteristics and ensuring broader-scale generalization for comparative analysis.

4. Discussion

With the rapid collection and production of new spatial datasets, the consistencies and correspondences among multi-source data sets has become a critical challenge in spatial statistics and analysis (A. Comber, et al., 2017). It is usually essential to validate their quality and reliability before adopting them for further modeling and decision-making. Traditionally, global indicators, including ME, MAE, MRE, RMSE, CC, and the coefficient estimate from the ULR, have been widely used to assess overall error direction, absolute and relative errors, the presence of extreme discrepancies, global correlation, and linear relationships between the corresponding observations in pair-wise datasets.

While these global metrics offer an overview of the consistency and discrepancies among multi-source spatial data sets, they cannot pinpoint specific details of the spatially varying degrees of differences or correlations (A. Comber, et al., 2017; Tsutsumida, et al., 2019). To address this limitation, we introduced six local indicators, including GW ME, GW MAE, GW MRE, GW RMSE, GW CC, and the coefficient estimate from the univariate GWR. These local indicators provide spatially detailed insights into consistency and accuracy by calculating geographically varying metrics, effectively overcoming the constraints of global measures. Moreover, the results demonstrate that the proposed GW framework is not dependent on a single bandwidth choice, but rather provides a scale-aware validation perspective in which bandwidth can be adapted to specific analytical objectives. By leveraging flexible bandwidth optimization or selection, these local indicators enable multi-scale comparisons. This flexibility can be particularly valuable in data production processes, where it facilitates precise quality control and iterative refinement of datasets.

In this study, we took the population data for a case study to compare the correspondences among the three data set, i.e. GPWv4, GPCD and the NPC7 of China. The GW indicators provide more nuanced evaluations of consistency and discrepancies by accounting for spatial heterogeneity in population distributions, offering valuable insights into the underlying patterns and potential sources of divergence among them By capturing regional differences, this local approach not only facilitates the identification of error origins, but also informs the appropriate choice of multi-source population data, that different decision may lead divergent or even contradictory outcomes (e.g. Hanberry, 2022; Hierink, et al., 2022; Yin, et al., 2021). The local validation framework is always straightforward to be applied to other types of spatial data sets, like carbon emission data (Lu, Dong, Wang, Sun, &

Yao, 2024) and nighttime light dataset (Zheng, Seto, Zhou, You, & Weng, 2023). This adaptability underscores the broad applicability and potential impact of the GW techniques in advancing multi-source spatial data validation and integration across diverse research domains. Overall, the value of introducing GW techniques lies in transforming validation from an aspatial summary into a spatial diagnosis. By identifying where discrepancies are concentrated and how they vary across spatial scales, the proposed framework provides essential information for downstream urban and environmental applications, where local data reliability often matters more than global averages and where subsequent causal investigations can be more effectively targeted.

While the proposed local validation framework offers a novel approach for assessing multi-source spatial data consistency, several limitations should be critically addressed here. Before validation, multi-source datasets must be accurately aligned regarding spatial reference, temporal reference, and sampling scale, as these factors can largely influence validation outcomes. Consequently, as with most validation studies involving spatial aggregation, the results are subject to the modifiable areal unit problem (MAUP) (Openshaw, 1984) and support effects associated with spatial upscaling or downscaling. In this study, such effects are treated as an inherent constraint of administrative-unit based validation rather than a source of methodological bias, as all datasets are harmonized to the same spatial support prior to analysis. An important direction for future work is to investigate how GW indicators can be adapted to datasets defined on different spatial supports, which would necessitate explicit treatment of scale transformation and support related uncertainty. Another important consideration is the scalability of the framework for multi-dataset comparisons. These findings may suggest that both GPCD and GPWv4 are not equivalent or interchangeable if a choice must be made when conducting comparative studies in regions like Heilongjiang, whereas GPCD might be a better alternation for research in regions such as Yunnan and Xinjiang. However, as the number of datasets increases, the complexity of calculations for local indicators and comparisons will also grow rapidly, creating challenges in interpreting the results. Additionally, while this study discussed six pairs of global and local indicators, it does not explore how to integrate these metrics into a comprehensive evaluation of data discrepancies and consistency. Future research should establish systematic methods for combining multiple indicators to deliver a comprehensive assessment of dataset reliability. Furthermore, this study primarily addresses the validation of multi-source spatial datasets, but when dealing with spatio-temporal datasets, additional considerations become necessary. A key future direction is to incorporate spatio-temporally varying techniques, like geographically and temporally weighted regression (Huang, Wu, & Barry, 2010), to validate dynamic and time-sensitive datasets. Addressing these challenges will enhance the practical applicability of the proposed framework, making it more effective for multi-source data validation, quality control, and integration across diverse spatial and temporal contexts.

5. Concluding remarks

In this study, we proposed a local validation framework for assessing the consistency and accuracy of multi-source spatial datasets. By utilizing the GPCD-NPC7 and GPWv4-NPC7 comparisons as illustrative examples for validation, we compared traditional global indicators with GW indicators to more effectively capture spatially varying discrepancies. The findings highlight that while global indicators such as ME, MAE, MRE, RMSE, CC, and coefficient estimate from ULR, provide a broad assessment of error direction, magnitude, and correlation strength, they fail to reveal spatial details of discrepancies between these two data sets. In contrast, local indicators such as GW ME, GW MAE, GW MRE, GW RMSE, GW CC, and coefficient estimates from univariate GWR, address this limitation by offering detailed spatial insights into dataset consistency and differences. The study also emphasizes the crucial role of bandwidth selection in determining the scale of analysis, with large bandwidths smoothing local variations and small bandwidths preserving fine-scale details. Hopefully, these indicators will be integrated into **GWmodelS** (Lu & Dong, 2022; Lu, et al., 2023; Lu, Hu, et al., 2024), offering a handy and interactive toolkit for comparing and validating multi-source spatial data. Overall, this technical framework provides a systematic and scalable approach to validating multi-source spatial data, enhancing quality control, improving integration strategies, and facilitating more informed decision-making in spatial analysis and geographic studies.

Data availability statement

All the data and code that support the findings of this study are available at the private link: https://figshare.com/s/1ae48d90b0ceef0c4992 .

References

Alexander, M., Polimis, K., & Zagheni, E. (2022). Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States. *Population Research and Policy Review, 41*, 1-28.

Atluri, G., Karpatne, A., & Kumar, V. (2018). Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Comput. Surv., 51*, Article 83.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (2002). Geographically weighted summary statistics -- a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems, 26*, 501-524.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis, 28*, 281-298.

Brunsdon, C., Fotheringham, S., & Charlton, M. (2007). Geographically Weighted Discriminant Analysis. *Geographical Analysis, 39*, 376-396.

Center For International Earth Science Information Network-CIESIN-Columbia University (2017). Gridded Population of the World, Version 4 (GPWv4): Population Density. In N. S. D. a. A. C. (SEDAC) (Ed.). Palisades.

Chin, T., Johansson, M. A., Chowdhury, A., Chowdhury, S., Hosan, K., Quader, M. T., Buckee, C. O., & Mahmud, A. S. (2025). Bias in mobility datasets drives divergence in modeled outbreak dynamics. *Communications Medicine, 5*, 8.

Comber, A., & and Tsutsumida, N. (2023). Geographically weighted accuracy for hard and soft land cover classifications: 5 approaches with coded illustrations. *International Journal of Remote Sensing, 44*, 6233-6257.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., & Harris, P. (2023). A Route Map for Successful Applications of Geographically Weighted Regression. *Geographical Analysis, 55*, 155-178.

Comber, A., Chris, B., Martin, C., & and Harris, P. (2017). Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. *Remote Sensing Letters, 8*, 234-243.

Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment, 127*, 237-246.

Comber, A. J. (2013). Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. *Remote Sensing Letters, 4*, 373-380.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences, 111*, 15888-15893.

Dong, N., Yang, X., Cai, H., & Xu, F. (2017). Research on Grid Size Suitability of Gridded Population

Distribution in Urban Area: A Case Study in Urban Area of Xuanzhou District, China. *PloS one, 12*, e0170830.

Doyle, J., Hung, P., Farrell, R., & McLoone, S. (2014). Population Mobility Dynamics Estimated from Mobile Telephony Data. *Journal of Urban Technology, 21*, 109-132.

Ehrlich, D., Melchiorri, M., Florczyk, A. J., Pesaresi, M., Kemper, T., Corbane, C., Freire, S., Schiavina, M., & Siragusa, A. (2018). Remote Sensing Derived Built-Up Area and Population Density to Quantify Global Exposure to Five Natural Hazards over Time. *Remote Sensing*.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. *Journal of Statistical Software, 63*, 1-50.

Goodchild, M. F. (2004). The Validity and Usefulness of Laws in Geographic Information Science and Geography. *Annals of the Association of American Geographers, 94*, 300-303.

Hall, O., Stroh, E., & Paya, F. (2012). From Census to Grids: Comparing the Gridded Population of the World with Swedish Census Records. *The Open Geography Journal, 5*, 1-5.

Hamdi, A., Shaban, K., Erradi, A., Mohamed, A., Rumi, S. K., & Salim, F. D. (2022). Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review, 55*, 1441-1488.

Hanberry, B. B. (2022). Imposing consistent global definitions of urban populations with gridded population density models: Irreconcilable differences at the national scale. *Landscape and Urban Planning, 226*, 104493.

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science, 25*, 1717-1736.

Harris, P., Chris, B., & and Charlton, M. (2013). The comap as a diagnostic tool for non-stationary kriging models. *International Journal of Geographical Information Science, 27*, 511-541.

Harris, P., & Juggins, S. (2011). Estimating Freshwater Acidification Critical Load Exceedance Data for Great Britain Using Space-Varying Relationship Models. *Mathematical Geosciences, 43*, 265-292.

Hierink, F., Boo, G., Macharia, P. M., Ouma, P. O., Timoner, P., Levy, M., Tschirhart, K., Leyk, S., Oliphant, N., Tatem, A. J., & Ray, N. (2022). Differences between gridded population data impact measures of geographic access to healthcare in sub-Saharan Africa. *Communications*

Medicine, 2, 117.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*, 383-401.

Karimi, Y., Haghi Kashani, M., Akbari, M., & Mahdipour, E. (2021). Leveraging big data in smart cities: A systematic review. *Concurrency and Computation: Practice and Experience, 33*, e6379.

Lu, B., & Dong, G. (2022). GWmodelS: A High-Performance Computing Framework for Geographically Weighted Models. In H. Wu, Y. Liu, J. Li, X. Meng, Q. Guan, X. Song, G. Liao & G. Li (Eds.), *Spatial Data and Intelligence* pp. 154-161). Cham: Springer Nature Switzerland.

Lu, B., Dong, J., Wang, C., Sun, H., & Yao, H. (2024). High-resolution spatio-temporal estimation of CO2 emissions from China's civil aviation industry. *Applied Energy, 373*, 123907.

Lu, B., Dong, Z., Yue, P., & Qin, K. (2024). Spatial Analysis of the Aging Population and Socio-economic Factors of China: Global and Local Perspectives. *Journal of Geodesy & Geoinformation Science, 7*.

Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spatial Information Science, 17*, 85-101.

Lu, B., Hu, Y., Yang, D., Liu, Y., Liao, L., Yin, Z., Xia, T., Dong, Z., Harris, P., Brunsdon, C., Comber, L., & Dong, G. (2023). GWmodelS: A software for geographically weighted models. *SoftwareX, 21*, 101291.

Lu, B., Hu, Y., Yang, D., Liu, Y., Ou, G., Harris, P., Brunsdon, C., Comber, A., & Dong, G. (2024). GWmodelS: a standalone software to train geographically weighted models. *Geo-Spatial Information Science*, 1-23.

Lu, B., Shi, Y., Qin, S., Yue, P., Zheng, J., & Harris, P. (2024). Evaluating Urban Land Resource Carrying Capacity With Geographically Weighted Principal Component Analysis: A Case Study in Wuhan, China. *Transactions in GIS, n/a*.

Lü, G., Batty, M., Strobl, J., Lin, H., Zhu, A. X., & Chen, M. (2019). Reflections and speculations on the progress in Geographic Information Systems (GIS): a geographic perspective. *International Journal of Geographical Information Science, 33*, 346-367.

Luo, P., Chen, C., Gao, S., Zhang, X., Majok Chol, D., Yang, Z., & Meng, L. (2024). Understanding of the predictability and uncertainty in population distributions empowered by visual analytics. *International Journal of Geographical Information Science*, 1-31.

Münch, Z. (2020). Global and local patterns of landscape change accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing, 161*, 264-277.

Noardo, F. (2022). Multisource spatial data integration for use cases applications. *Transactions in GIS, 26*, 2874-2913.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Norwich, England: Geo Abstracts Ltd.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PloS one, 10*, e0107042.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography, 46*, 234-240.

Tsutsumida, N., Rodríguez-Veiga, P., Harris, P., Balzter, H., & Comber, A. (2019). Investigating spatial error structures in continuous raster data. *International Journal of Applied Earth Observation and Geoinformation, 74*, 259-268.

Uhl, J. H., & Leyk, S. (2022). A scale-sensitive framework for the spatially explicit accuracy assessment of binary built-up surface layers. *Remote Sensing of Environment, 279*, 113117.

Wang, S., & Li, W. (2021). GeoAI in terrain analysis: Enabling multi-source deep learning and data fusion for natural feature detection. *Computers, Environment and Urban Systems, 90*, 101715.

Wang, Z., Hale, S., Adelani, D. I., Grabowicz, P., Hartman, T., Flöck, F., & Jurgens, D. (2019). Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. *The World Wide Web Conference* pp. 2056–2067). San Francisco, CA, USA: Association for Computing Machinery.

Xie, Y., Weng, A., & Weng, Q. (2015). Population estimation of urban residential communities using remotely sensed morphologic data. *IEEE Geoscience and Remote Sensing Letters, 12*, 1111-1115.

Xu, X. (2020). Gridded Population of China Dataset. In D. R. a. P. S. o. R. a. E. S. D. C. o. C. A. o. Sciences (Ed.). Beijing: Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

Xu, Y., Ho, H. C., Knudby, A., & He, M. (2021). Comparative assessment of gridded population data sets for complex topography: a study of Southwest China. *Population and Environment, 42*, 360-378.

Yin, X., Li, P., Feng, Z., Yang, Y., You, Z., & Xiao, C. (2021). Which Gridded Population Data Product Is Better? Evidences from Mainland Southeast Asia (MSEA). *ISPRS International Journal of Geo-Information*.

Zhang, B., Liu, L., Zhang, Y., Wei, B., Gong, D., & Li, L. (2024). Spatial Consistency and Accuracy Analysis of Multi-Source Land Cover Products on the Southeastern Tibetan Plateau, China. *Remote Sensing*.

Zheng, Q., Seto, K. C., Zhou, Y., You, S., & Weng, Q. (2023). Nighttime light remote sensing for urban applications: Progress, challenges, and prospects. *ISPRS Journal of Photogrammetry and Remote Sensing, 202*, 125-141.