

Article

Population Mapping with Multisensor Remote Sensing Images and Point-Of-Interest Data

Xuchao Yang^{1,2}, Tingting Ye¹, Naizhuo Zhao³ , Qian Chen¹, Wenze Yue⁴ , Jiaguo Qi² ,
Biao Zeng^{5,*}  and Peng Jia^{6,7,*} 

¹ Ocean College, Zhejiang University, Zhoushan 316021, China; yangxuchao@zju.edu.cn (X.Y.), tingting.ye@zju.edu.cn (T.Y.), chenqian824@zju.edu.cn (Q.C.)

² Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48824, USA; qi@msu.edu

³ Department of Medicine, McGill University, Montreal, QC H3A 1A1, Canada; naizhuo.zhao@mail.mcgill.ca

⁴ Department of Land Management, Zhejiang University, Hangzhou 310058, China; wzyue@zju.edu.cn

⁵ College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China

⁶ Department of Earth Observation Science, Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, 7500 Enschede, The Netherlands

⁷ International Initiative on Spatial Lifecourse Epidemiology (ISLE)

* Correspondence: zengb@lzu.edu.cn (B.Z.); p.jia@utwente.nl (P.J.); Tel.: +86-15352318596 (B.Z.); +31-534894038 (P.J.)

Received: 6 February 2019; Accepted: 3 March 2019; Published: 8 March 2019



Abstract: Fine-resolution population distribution mapping is necessary for many purposes, which cannot be met by aggregated census data due to privacy. Many approaches utilize ancillary data that are related to population density, such as nighttime light imagery and land use, to redistribute the population from census to finer-scale units. However, most of the ancillary data used in the previous studies of population modeling are environmental data, which can only provide a limited capacity to aid population redistribution. Social sensing data with geographic information, such as point-of-interest (POI), are emerging as a new type of ancillary data for urban studies. This study, as a nascent attempt, combined POI and multisensor remote sensing data into new ancillary data to aid population redistribution from census to grid cells at a resolution of 250 m in Zhejiang, China. The accuracy of the results was assessed by comparing them with WorldPop. Results showed that our approach redistributed the population with fewer errors than WorldPop, especially at the extremes of population density. The approach developed in this study—incorporating POI with multisensor remotely sensed data in redistributing the population onto finer-scale spatial units—possessed considerable potential in the era of big data, where a substantial volume of social sensing data is increasingly being collected and becoming available.

Keywords: point-of-interest; remote sensing; nighttime light; population modeling

1. Introduction

High-resolution population distribution data are essential in addressing a wide range of critical issues, such as vulnerability assessment [1,2], urban planning [3,4], emergency management [5], and public health [6,7]. In most countries worldwide, commonly available information on population number and composition through the Census Bureau is aggregated over administrative units, such as provinces, counties, townships, census tracts, and block groups. The usefulness of these census data is limited due to the spatial heterogeneity of population distribution within administrative units [8]. Meanwhile, both the availability and quality of environmental data are increasing. Such an unmatched development of demographic and socioeconomic data and natural science data, especially at the fine

levels of granularity, has hindered the advancement of decision making in many aspects, such as resource allocation [9] and disease prevention [10], and, more broadly, the integration of natural and social sciences [11]. Therefore, the development of efficient methods for accurately modeling fine-scale population distribution is urgently needed.

A number of approaches have been developed to disaggregate census population data, the most reliable population data sources worldwide, onto fine-scale grids, with the value of each grid representing the population number within that grid. Examples include areal weighting interpolation [12], pycnophylactic interpolation [13], and dasymetric mapping [11,14] which outperforms previous approaches by utilizing high-quality ancillary data to redistribute the population over space [15]. Remote sensing data, such as land cover data, have been widely used as ancillary information on where people may live in dasymetric mapping approaches for population redistribution [11,16].

Since the late 1990s, satellite-derived nighttime light (NTL) data have been proven to be a reliable proxy for population distribution [17,18]. The NTL dataset from the US Air Force Defense Meteorological Satellite Program Operational Linescan System (DMSP/OLS) is a widely used product for dasymetric mapping [19–22]. Despite its benefits, the DMSP/OLS dataset has several limitations, such as its single spectral band and coarse spatial resolution (2.7 km), saturation in urban centers, and blooming effect [23–25]. For example, due to the blooming effect, the lit areas shown in the DMSP/OLS dataset are generally larger than actual urban areas [26,27]. Several studies have been conducted to overcome limitations above, such as combining NTL and land cover data to improve the representation of population distribution [25,28–30]. Using such data fusion approaches, a pixel-based elevation-adjusted human settlement index (EAHSI) has been produced on the basis of NTL, enhanced vegetation index (EVI), and the digital elevation model (DEM). The EAHSI was used as ancillary data to generate a population density map with the spatial resolution of 250 m for Zhejiang province, China [31]. However, some industrial areas with a high EAHSI value are actually less populated than expected, which has led to a considerable degree of population overestimation.

Social sensing data that are becoming extremely popular in the era of big data could be a potential solution to improve the accuracy of the population products generated purely on the basis of the environmental ancillary data. For example, the point-of-interest (POI) data are one of the most commonly used social sensing datasets in urban studies. Each POI with geographic coordinates generally represents a functionally built environmental feature. Certain POI types associated with more human activities may indicate better livability and higher population density than other types [32]. Recently, POI data have been used as ancillary data to enhance population estimation over relatively small areas [3,4]. However, these data have not been combined with other data sources in terms of high-resolution population modeling.

Here, we incorporated POIs with multisource remote sensing data to further improve the accuracy of the population modeling. The resulting population dataset was compared with a widely used global population product. This study has introduced the field of high-resolution population modeling by utilizing an innovative combination of remote sensing and social sensing data to refine population distribution. With both types of sensing data increasingly becoming available, the approach proposed in this study would lead to the development of better predictive tools for population estimation.

2. Methods

2.1. Study Area

The study area was Zhejiang, which is located in the southeastern coast of China (south of Shanghai), with a total land area of approximately 101,800 km² and a long (6484 km) coastline in the eastern part (Figure 1a). With approximately 54.4 million permanent residents at the end of 2010 (the latest census year in China), Zhejiang is the 10th most populated province and that with the 4th largest gross domestic product in China. Hills and mountains cover 70.4% of the land in Zhejiang, with only 23.2% of the land covered by plains and basins. The majority of the population in Zhejiang resides in

the northern plains and eastern coastal areas (Figure 1b). The hierarchy of the administrative units in Zhejiang from coarse to fine includes 11 cities, 90 counties, and 1520 townships.

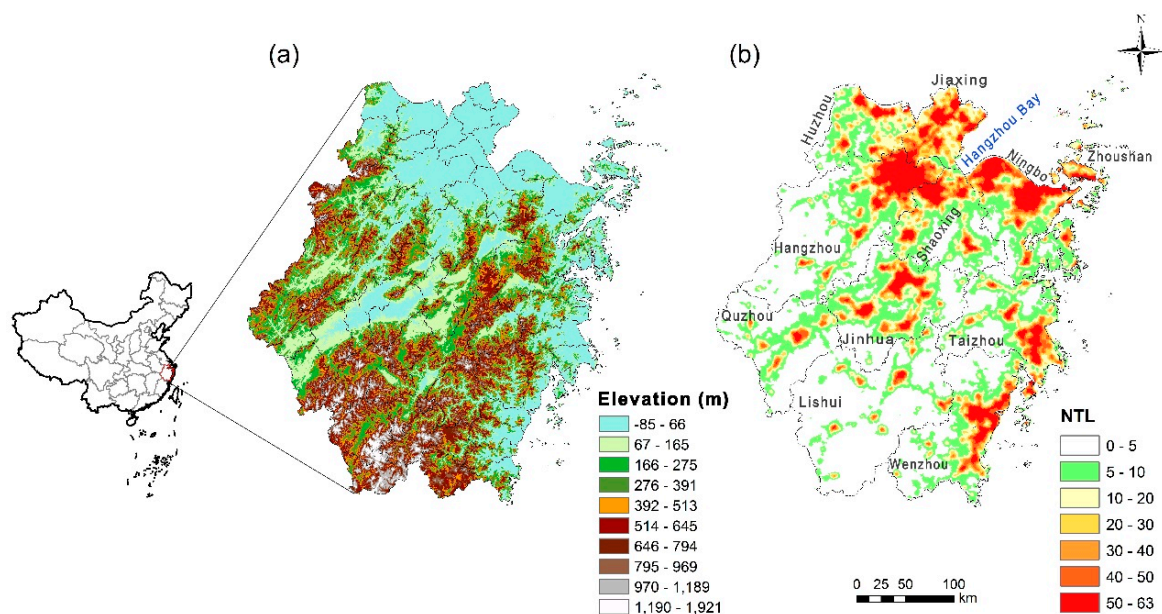


Figure 1. Elevation (a) and nighttime light (b) in Zhejiang in 2010.

2.2. Data Sources and Preprocessing

The seven datasets used in this study were all produced in 2010 and obtained from different sources (Table 1). The township-level (equivalent to level 4 of the Global Administrative Unit Layer defined by the Food and Agriculture Organization) population data and administrative unit boundaries were obtained and combined as the original population data source. For accuracy assessment, the Zhejiang part of a global gridded population dataset, that is, the WorldPop with a spatial resolution of 100 m [33], was used.

Table 1. List of datasets (all produced in 2010) used in this study.

Data Set	Format	Source & Citation
Census data	Table	Zhejiang Bureau of Statistics in 2010, China
Administrative boundary	Vector (polygon)	Zhejiang Administration of Surveying Mapping and Geoinformation, China Population Count, Revision 10. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PG1PPM . Dataset: CHN_ppp_v2c_2010.tif http://www.worldpop.org.uk/data/summary/?doi=10.5258/SOTON/WP00055
WorldPop-China	Raster (100 m)	Baidu Inc., China
Point-of-interest	Vector (point)	National Oceanic and Atmospheric Administration's National Geophysical Data Center, USA https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html Data set: F182010
DMSP/OLS data	Raster (1 km)	MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD13Q1.006 Land Processes Distributed Active Archive Center, USA
MODIS EVI	Raster (250 m)	ASTER Global Digital Elevation Model V002, DOI: 10.5067/ASTER/ASTGTM.002
GDEM	Raster (30 m)	

The POI data including 386,178 POIs located in Zhejiang falling within 20 categories (Table 2) were obtained from Baidu Map Services (<http://map.baidu.com>), which is the most widely used and also largest web map service provider in China [3].

The Moderate Resolution Imaging Spectroradiometer EVI products (MOD13Q1) at a spatial resolution of 250 m, which was available every 16 days in 2010, were downloaded from the US Geological Survey. Compared with the normalized difference vegetation index, which is a well-known conventional vegetation index, EVI is responsive to canopy structural variations [34]; therefore, it is likely to avoid saturation in the southern and western areas of Zhejiang with extremely dense

vegetation. To remove the cloud effects, the annual maximum EVI (EVI_{max}) was produced for each grid cell by implementing raster math calculations on 23 EVI images over a year [31]. The NTL data in 2010 was obtained from a DMSP/OLS stable light image composite at a spatial resolution of 1 km, which is produced by the National Oceanic and Atmospheric Administration's National Centers for Environmental Information. The digital number (DN) values in an NTL image varying from 0 to 63 represent the average brightness of NTL in 2010, except for 63, which was assigned to saturated pixels. The Advanced Spaceborne Thermal Emission and Reflection Radiometer Global DEM version 2 at a spatial resolution of 30 m was downloaded from the US Land Processes Distributed Active Archive Center. Both NTL and DEM data were resampled to 250 m through bilinear interpolation to spatially match the EVI data. All three remote sensing images above were reprojected to the Albers Conical Equal Area Projection and then clipped by the Zhejiang boundary.

2.3. Methodology

The objective of our study was to spatially disaggregate census data by the township level into each pixel to produce a population distribution map with a fine spatial resolution (i.e., 250 m × 250 m). We adopted an improved linear regression-based method that combined the multisource remote sensing images and POIs. The major steps of this improved method are shown by the flowchart (Figure 2).

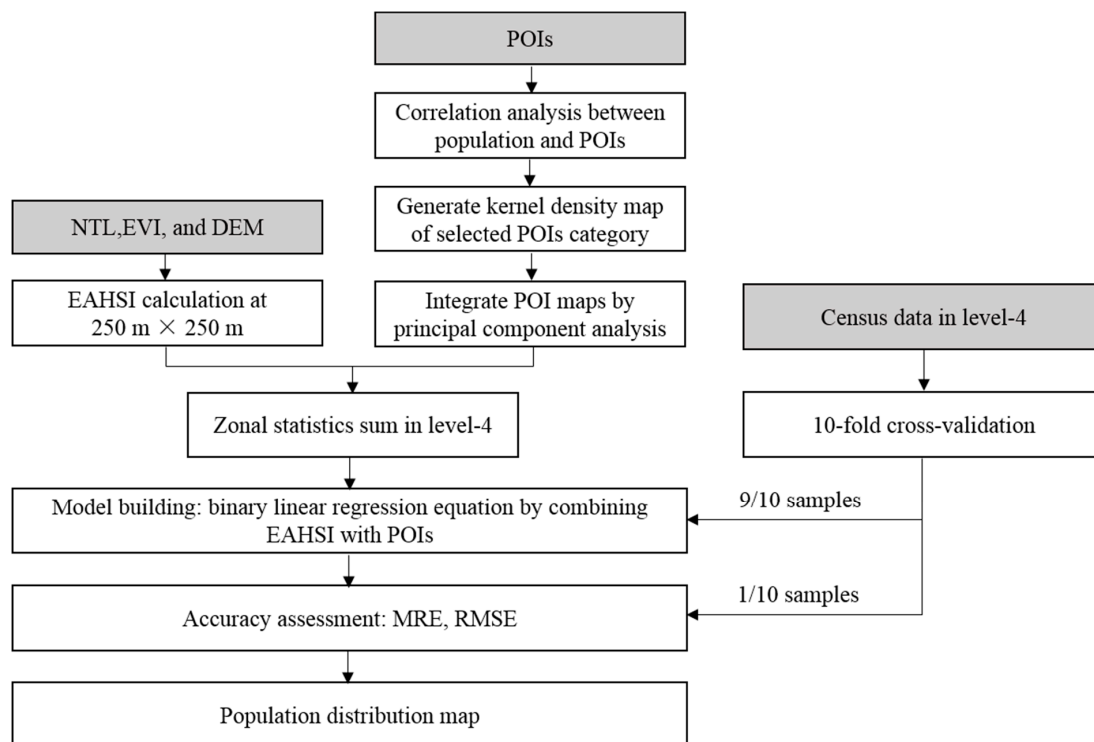


Figure 2. Flowchart of disaggregating census population data into the 250 m grid cells.

2.3.1. Generating an EAHSI Image

An EAHSI image covering Zhejiang for the year 2010, with a spatial resolution of 250m, was generated based on the EVI_{max} value and resampled NTL and DEM layers, as follows:

$$EAHSI = \frac{(1 - EVI_{max}) + NTL_{nor}}{(1 - NTL_{nor}) + EVI_{max} + NTL_{nor} \times EVI_{max}} \times e^{-0.003DEM} \quad (1)$$

where e is approximately equal to 2.71828, and NTL_{nor} is the normalized DN value of the NTL image, which was calculated as follows:

$$NTL_{nor} = (NTL - NTL_{min}) / (NTL_{max} - NTL_{min}) \quad (2)$$

where NTL_{max} and NTL_{min} are the maximum and minimum NTL values in the study area, respectively. Detailed information about EAHSI can be found in a previous study [31].

2.3.2. Generating a POI Density Layer

Spearman's correlation analysis was adopted to examine the relationship between the number of each POI category and the population at the township level (Table 2).

Table 2. Spearman's correlation coefficients between each point-of-interest (POI) category and census population at the township level (ranked in descending order).

POI Category	Correlation Coefficient	POI Count
Education facility	0.904	16,173
Hospital and clinic facility	0.880	24,173
Public service facility	0.872	26,887
Retail	0.871	89,192
Bank	0.869	19,596
Restaurant and entertainment	0.868	55,406
Company	0.863	75,233
Government agency	0.795	20,737
Residential communities	0.767	10,174
Factory	0.766	13,400
Auto service	0.765	11,675
Hotel	0.747	13,972
Gas station	0.727	2768
Commercial building	0.644	3426
Public transport station	0.417	450
Park	0.381	931
Service zone of Highway	0.315	1144
Toll station	0.312	764
Railway station	0.205	46
Airport	0.092	31

All correlation coefficients were significant at the 0.01 level.

Kernel density estimation (KDE) is a well-established method in analyzing the first-order properties of a point event distribution [35–37] and identifying hot spots [38–40]. KDE was used to convert each category of discrete POIs into a smooth and continuous density surface. Previous studies showed that the statistical results are insignificantly affected by the choice of the kernel function; hence, bandwidth is the main parameter for KDE [35,36]. The planar KDE with a quartic kernel function, which is one of the most commonly used functions [41], was implemented in this study. The township level boundary map was used to summarize each category of POI density to level 4 and train the linear correlation between the sum value of POI density and the population counts. We tested different bandwidths ranging from 500 m to 8000 m at an interval of 100 m. There was a slight fluctuation when the bandwidth was between 2000 m and 5000 m, and the correlation coefficients were relatively high. Finally, we determined a reasonable bandwidth for all categories of POI at a 3000 m bandwidth when the POI densities of most categories had the largest correlation coefficients with the population. Thus, we selected 3000 m as the bandwidth of the KDEs.

Principal component analysis, which is one of the most commonly used dimension-reducing techniques that can reduce a large number of correlated variables to a small number of uncorrelated

ones [42], was adopted to combine multiple POI kernel density maps into one composite POI density layer.

2.3.3. Mapping Population

Given that both POI density and EAHSI values linearly correlate to the population count at the township level (graphical abstract), a multiple linear regression model was built, as follows:

$$\text{POP} = a \times \text{POI} + b \times \text{EAHSI} \quad (3)$$

where POP represents the estimated population counts at township level, and the coefficients “a” and “b” are the average values of the 10 repeated trials of 10-fold cross-validation. A 10-fold cross-validation was adapted for randomly dividing township census data into 10 groups, and repeated 10 trials to determine steady coefficients “a” and “b”. Specifically, census data of nine groups of the townships were used to train the model, and the census data in the remaining group were used to evaluate the multiple linear regression model. To show the advantage of fusing POIs for population estimation, we also conducted similar cross-validations to build a linear regression model by using only EAHSI. Table 3 shows a summary of 100 groups’ repeated trials for EAHSI and POI-EAHSI. Finally, the gridded EAHSI and POI were used to disaggregate POP at the township level to predict pixel-level population counts.

Table 3. Summary of 10 repeated trials of 10-fold cross-validations for EAHSI and POI-EAHSI.

	EAHSI					POI-EAHSI				
	Training Group		Testing Group			Training Group		Testing Group		
	a	R ²	%RMSE	MRE (%)	a	b	R ²	%RMSE	MRE (%)	
Mean	83.29	0.55	73.58	67.34	52.61	25.61	0.78	48.85	30.46	
Standard Error	0.680	0.004	0.72	1.03	0.693	0.661	0.004	0.76	0.55	

2.3.4. Accuracy Assessment

To highlight the fact that the use of POI data can significantly increase the accuracy of population mapping, we compared our population map produced by POI and EAHSI (referred to as the POI-EAHSI population map) with the map produced only by EAHSI and the WorldPop gridded population products. For accuracy assessment, an average estimate of out-of-sample prediction was generalized for POI-EAHSI and EAHSI data sets and WorldPop population data was aggregated by townships and then compared with census data to evaluate the accuracy. Summary statistics, including the root mean square error (RMSE), the RMSE divided by the mean township population count (%RMSE), the mean absolute error, and mean relative error (MRE), were calculated for the three methods.

3. Results

3.1. Population Density

Table 2 shows that all categories of POIs were positively related to population counts. We combined 20 kernel POI density maps to one composite POI density layer. Only the first principal component image (Figure 3) was used because its contributing rate of the cumulative sums of squares reached 88.07%.

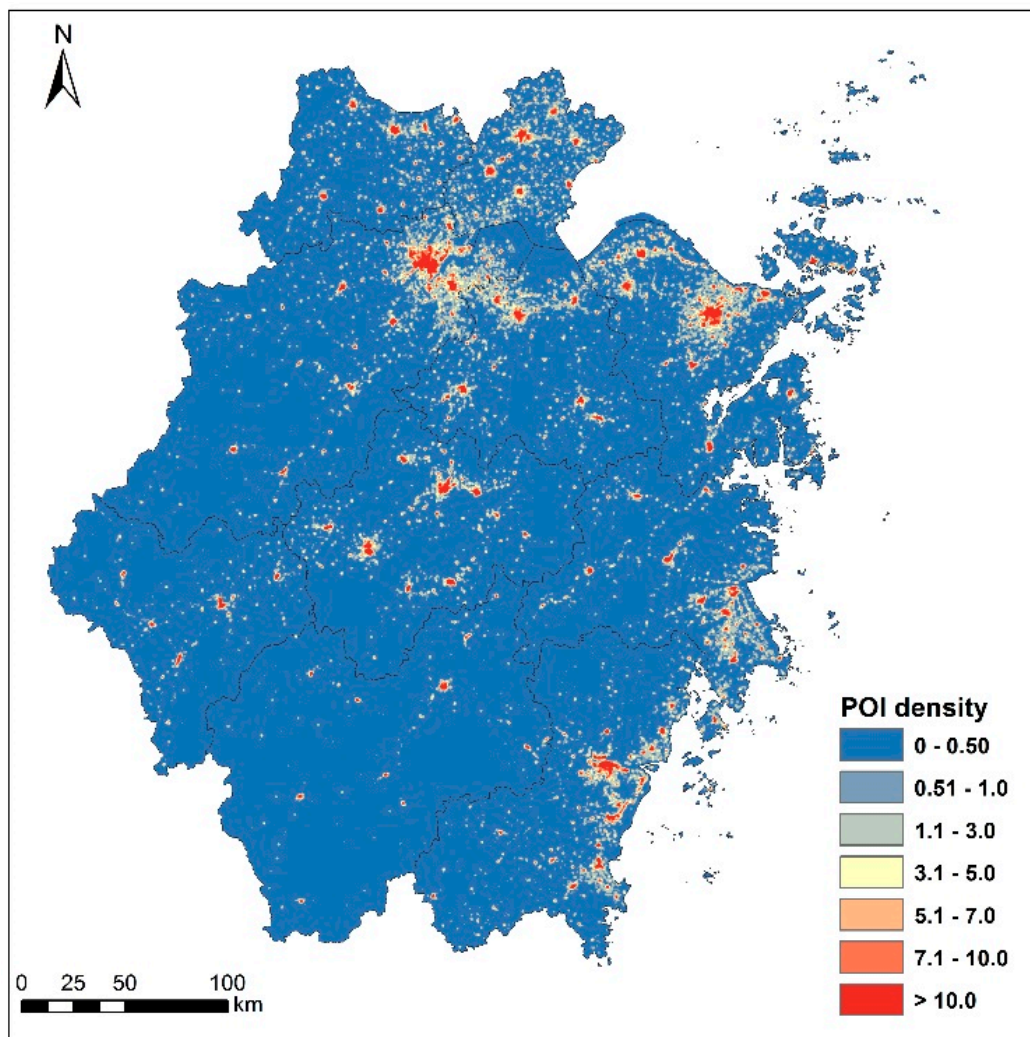


Figure 3. Kernel density map of the first principal component image.

Table 3 shows the summary of 10 repeated trials of 10-fold cross-validation, which indicated that the multiple linear regression model on the basis of fused POIs and EAHSI was credible to estimate population distribution, with a mean value of the coefficient of determination (R^2) of 0.78, while the mean R^2 of using only EAHSI was 0.55. The %RMSE and MRE of incorporating POIs were also significantly smaller than using only EAHSI.

According to Equation (1) with the a and b values of 52.61 and 25.61, respectively, the gridded population map in Zhejiang for 2010 at a spatial resolution of 250 m was generated. Most of the population lived in the urban agglomerations around Hangzhou Bay, Wenzhou–Taizhou coastal region, and Jinhua–Quzhou basin. The Hangzhou and Ningbo regions in northern Zhejiang were the most heavily populated regions (Figure 4). The spatial distribution of predicted population for Zhejiang was generally consistent with the results of a previous study [31]. However, population distribution in the present study revealed apparent spatial heterogeneity and rich information in urban centers due to the combination of POIs. The population density map can be widely used in numerous activities, such as demographic studies, decision making, spatial planning, and emergency response in Zhejiang.

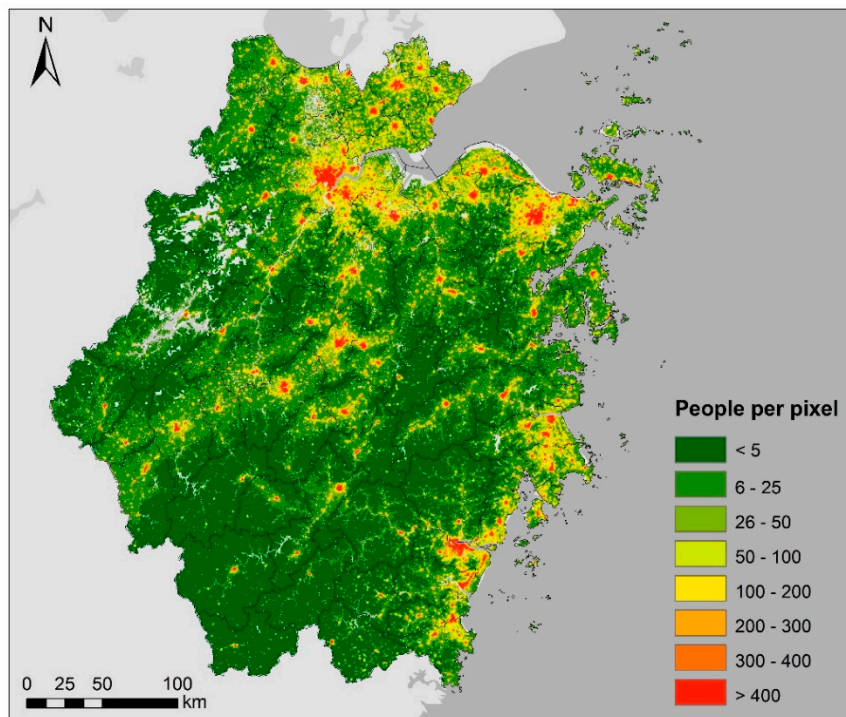


Figure 4. Gridded population map of Zhejiang province in 2010 (units—individuals per 0.0625 km²).

3.2. Accuracy Assessment

Figure 5 shows the relationship between estimated and census population counts at the township level for Zhejiang. Each point represented an estimated and actual population count within a township unit. The relationship between the predicted gridded estimates and the census population totals was substantially more linear for the POI-EAHSI method than the WorldPop method. The POI-EAHSI method also showed the highest correlation between estimated and census values ($R^2 = 0.88$) compared with the WorldPop ($R^2 = 0.79$) dataset. A significant decrease in MRE (30.46%) and RMSE (1.78) were attained to confirm the improved performance of our proposed method.

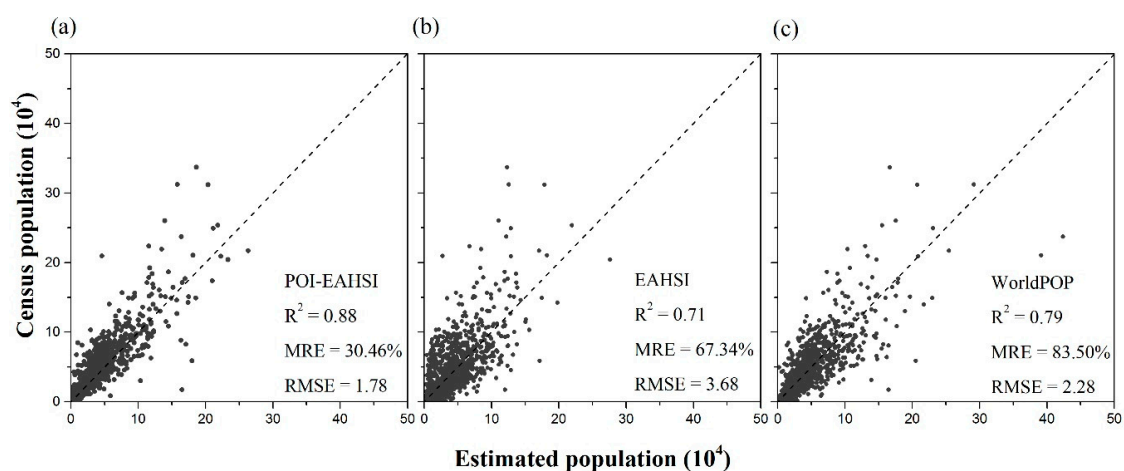


Figure 5. Scatterplots between the census population counts and (a) POI-EAHSI population estimates, (b) EAHSI-derived population estimates, and (c) WorldPop population at the township level for Zhejiang in 2010.

We compared the distribution of the residuals of population estimation by POI-EAHSI (Figure 6a) and EAHSI methods (Figure 6b). Population residual was calculated by subtracting the census data

from the predicted population of out-of-samples. A negative residual implied that the predicted value was an underestimation, and a positive residual indicated an overestimation. A same-color bar was adopted to compare the distribution of errors easily. In general, EAHSI caused population overestimation in most regions of Zhejiang (Figure 6b). The POI-EAHSI method can significantly decrease the errors and improve the model precision over the entire province, especially in the northern part of Zhejiang (Figure 6a). In the southeast coastal regions of Zhejiang, significant population underestimation was observed (Figure 6a,b). Extensive land reclamation in the coastal areas of Taizhou and Wenzhou for real estate development caused a mismatch between satellite images and administrative boundaries, thereby influencing the results of zonal statistics. The long coastline and the numerous islands in the southeastern coastal regions resulted in the discarding of pixels in zonal statistics, thereby contributing to population underestimation. Compared with WorldPop data, the improvement of the POI-EAHSI method was also apparent across most township units in Zhejiang (Figure 6a,c). A previous study suggested that the WorldPop mainland China dataset has high accuracy [43]. Therefore, we further compared the POI-EAHSI results with the WorldPop dataset.

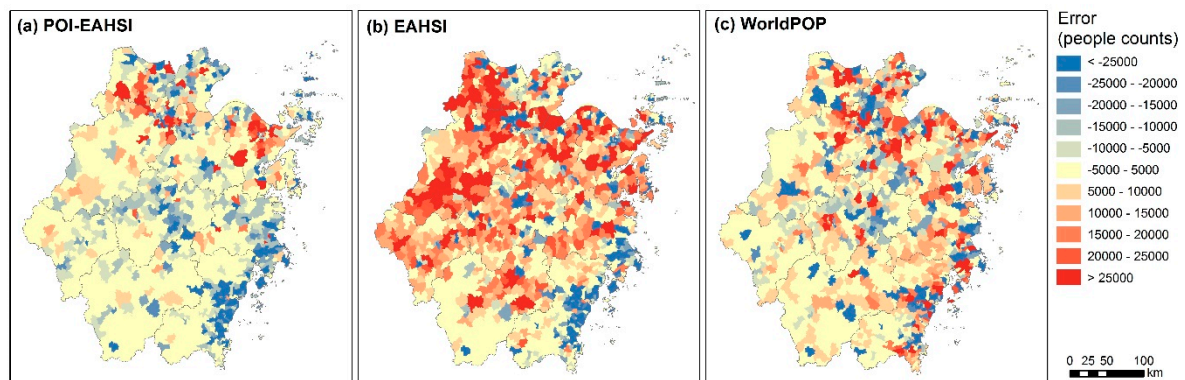


Figure 6. Distribution of the residuals of predicted population by (a) using POI-EAHSI, (b) EAHSI, and (c) WorldPop.

Figure 7 shows the model fit between the predicted population density of each township unit compared with the original census population density at the same census unit level for 2010 for POI-EAHSI results and the WorldPop dataset. According to the census population density, all the township units in Zhejiang were classified into three groups, namely, top 20%, medium 60%, and low 20% (red, green, and blue dots, respectively, in Figure 7). There was a good fit at medium population densities for both POI-EAHSI and WorldPop with a similar variable explanation ($R^2 = 0.72$ vs. $R^2 = 0.65$). However, there were larger errors at extreme population densities (Figure 7). At high population density, an underestimation of the original census data was observed, whereas significant overestimation was observed at extremely low population density, especially for WorldPop. These types of errors were also observed in previous dasymetric modeling studies [44–46]. However, POI-EAHSI showed significantly higher accuracy than WorldPop in both tails of population density, especially for the low tail ($R^2 = 0.57$ vs. $R^2 = 0.15$).

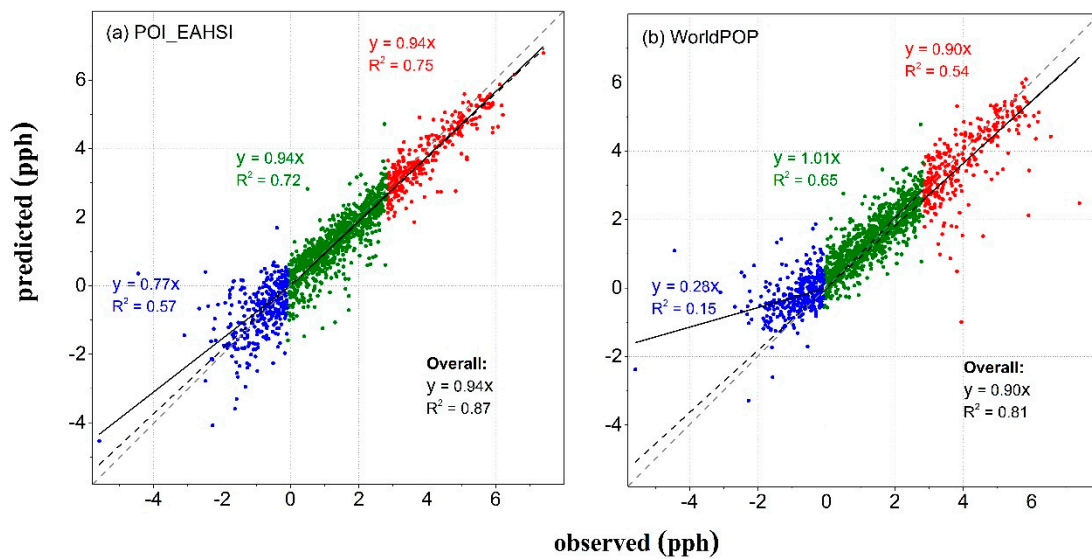


Figure 7. Scatterplots between the predicted population density on a \log_{10} – \log_{10} scale at the township unit and the original census population density at the same unit level. Red points are township units with a large population density for the top 20%, and blue points are units with low population density for low 20% tails. The comparison of the validation unit counts divided by unit area (population density) on an \ln – \ln scale with those estimated from maps produced using county census units.

4. Discussion

Spatially accurate data on human population distributions are vital for many applied and theoretical studies. Dasymetric mapping techniques using NTL data as the ancillary information have been increasingly used to disaggregate census population to a finer spatial level. However, the uncertainties in the relationships between NTL and human population distribution should be recognized. NTL emissions depend on affluence, culture, and economic structure [17,47]. In many cities of developed countries, commercial advertising, sports facilities, and security lighting often represent additional sources of NTL emissions. Therefore, NTL brightness does not directly or consistently reflect population distribution. A number of experiments demonstrated that remote sensing data, such as land use and NTL data, cannot be used to conduct accurate population estimation at a fine scale, especially in a complex urban environment [4]. The underestimation in high-population-density areas and the overestimation in low-population-density areas due to spatial nonstationarity is a frequently recurring problem in dasymetric mapping studies [44–46]. The derivation of global parameters in this method imposes an averaging effect on the disaggregation of the population that masks the intrinsic heterogeneity in population distribution characteristics [46].

Recently, social sensing data proved to be useful in population estimation. Previous studies used mobile phone data [48], Twitter [49], or OpenStreetMap data [50] to improve population mapping. However, volunteered geographic information (VGI)-based data such as OpenStreetMap data in China is far from being complete [51]. Mobile phone data is difficult to obtain for a large study area. Twitter cannot be used in China. Few studies started to use POIs to estimate population distributions at a fine spatial resolution on a small scale, such as a single city [4] or urban districts [3]. This study built a population model to disaggregate census data and obtained a high-precision population map at a fine spatial resolution of 250 m by fusing multisource remote sensing data and POIs. A case study for Zhejiang, China has been conducted. KDE and principal component analysis were used to generate a POI density map, which highly relates to human daily life and population distribution in urbanized areas. The results showed that POIs can be considered as useful ancillary data for population estimation even at the regional scale. Compared with WorldPop global population datasets, the method in this study that fuses information from multisource remote sensing data and POI data

can generate improved estimation performance in revealing the actual population distribution at a fine scale, especially in urbanized areas.

Human settlement index on the basis of NTL and vegetation index can effectively map human settlements [52] and impervious surface [53] while distinguishing commercial, residential, and industrial areas is difficult. POIs can supplement information to identify urban functional zones [54–57]. Compared with EAHSI, POIs that are mainly located in urban areas and are highly related to human daily life can represent an area with high population density and exclude industrial regions [58]. In addition, POI data possessed a simple data structure compared with other multidimensional data. Therefore, POIs can be easily used to refine population estimation, especially in urbanized areas. The incorporation of POIs facilitated the decrease in the weight in commercial and industrial areas, which certainly improved the population prediction.

The quality and the appropriateness of the ancillary data used influenced the accuracy of the population estimation. One of the uncertainties of our method is the quality of POI data because POI descriptions are generally provided by volunteers, and inaccurate descriptions are likely to occur. However, our POI data were obtained from a commercial navigation database and were collected by trained persons and subject to strict inspection. In addition, these POI data are used in the Baidu Map and navigation app. Therefore, the positional and thematic accuracy of Baidu POI data is reliable. Most POIs concentrate in urban areas, which most likely limits the improvement of population estimation of our method to urban regions. In rural areas and urban fringe areas, many POIs are unreported, and POI density is relatively low. Therefore, POI data may not be an effective measure of population density in non-urban areas. Moreover, the correlations between some categories of POI and population density may vary in different cities, since urban fabric patterns vary across regions. In this regard, reproductions of this study in other geographic areas/countries need to investigate spatial patterns of POIs that reflect population distribution. Finally, although POIs can identify the footprints of human activities, they cannot provide the extent of these activities. The lack of information on the volume of buildings may lead to population underestimation or overestimation [44].

5. Conclusions

Social sensing data, such as POI, directly reflect human activities and contain rich information on place semantics, and have significantly complemented traditional remote sensing data in the context of population estimation. Considering that social sensing and remote sensing data capture different aspects of human activities, integrating these two types of data is a promising research topic. Our approach took advantage of the information from POIs and multisource remote sensing data to obtain the detailed and accurate characteristics of the population distribution and subsequently improve population estimation. The POI-EAHSI model incorporating POI data overcame the systematic overestimation and underestimation issues in previous studies and produced the most accurate results, especially at the extremes of the population density. This paper provided a new approach for the rapid and accurate estimation of the human population at the regional scale. The integrated approach for population estimation has the potential to adopt more remote sensing data and new types of social sensing big data to estimate population in more flexible ways in the future, such as age-specific population estimation [59]. The values of multisource social sensing data in population estimation will be explored in future studies to further improve the accuracy of population mapping.

Author Contributions: Conceptualization, P.J. and B.Z.; methodology, X.Y. and N.Z.; formal analysis, T.Y. and Q.C.; resources, W.Y. and J.Q.; data curation, T.Y.; writing—original draft preparation, X.Y. and T.Y.; writing—review and editing, P.J., N.Z., B.Z., W.Y., and J.Q.; visualization, T.Y.; supervision, P.J. and B.Z.; project administration, X.Y., funding acquisition, X.Y. and P.J.

Funding: This work was supported by the National Natural Science Foundation of China (No. 41671035), the Fundamental Research Funds for the Central Universities, and the State Key Laboratory of Urban and Regional Ecology of China (No. SKLURE2018-2-5). Peng Jia, Director of the International Initiative on Spatial Lifecourse Epidemiology (ISLE), thanks Lorentz Center, the Netherlands Organization for Scientific Research, the Royal

Netherlands Academy of Arts and Sciences, the Chinese Center for Disease Control and Prevention, the West China School of Public Health in Sichuan University, for funding the ISLE and supporting ISLE's research activities.

Acknowledgments: The authors acknowledge the four anonymous reviewers and Editor for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Aubrecht, C.; Özceylan, D.; Steinnocher, K.; Freire, S. Multi-level geospatial modeling of human exposure patterns and vulnerability indicators. *Nat. Hazards* **2013**, *68*, 147–163. [[CrossRef](#)]
2. Zeng, J.; Zhu, Z.Y.; Zhang, J.L.; Ouyang, T.P.; Qiu, S.F.; Zou, Y.; Zeng, T. Social vulnerability assessment of natural hazards on county-scale using high spatial resolution satellite imagery: A case study in the Luogang district of Guangzhou, South China. *Environ. Earth Sci.* **2012**, *65*, 173–182. [[CrossRef](#)]
3. Yao, Y.; Liu, X.; Li, X.; Zhang, J.; Liang, Z.; Mai, K.; Zhang, Y. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1220–1244. [[CrossRef](#)]
4. Bakillah, M.; Liang, S.; Mobasher, A.; Jokar Arsanjani, J.; Zipf, A. Fine-resolution population mapping using OpenStreetMap points-of-interest. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1940–1963. [[CrossRef](#)]
5. Dobson, J.E.; Bright, E.A.; Coleman, P.R.; Durfee, R.C.; Worley, B.A. LandScan: A Global Population Database for Estimating Populations at Risk. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 849–857.
6. Jia, P.; Sankoh, O.; Tatem, A.J. Mapping the environmental and socioeconomic coverage of the INDEPTH international health and demographic surveillance system network. *Health Place* **2015**, *36*, 88–96. [[CrossRef](#)] [[PubMed](#)]
7. Hay, S.I.; Noor, A.M.; Nelson, A.; Tatem, A.J. The accuracy of human population maps for public health application. *Trop. Med. Int. Health* **2005**, *10*, 1073–1086. [[CrossRef](#)] [[PubMed](#)]
8. Bhaduri, B.; Bright, E.; Coleman, P.; Urban, M.L. LandScan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *GeoJournal* **2007**, *69*, 103–117. [[CrossRef](#)]
9. Jia, P.; Anderson, J.D.; Leitner, M.; Rheingans, R. High-Resolution Spatial Distribution and Estimation of Access to Improved Sanitation in Kenya. *PLoS ONE* **2016**, *11*, e0158490. [[CrossRef](#)] [[PubMed](#)]
10. Wang, J.; Jia, P.; Cuadros, D.F.; Xu, M.; Wang, X.; Guo, W.; Portnov, B.A.; Bao, Y.; Chang, Y.; Song, G.; et al. A Remote Sensing Data Based Artificial Neural Network Approach for Predicting Climate-Sensitive Infectious Disease Outbreaks: A Case Study of Human Brucellosis. *Remote Sens.* **2017**, *9*, 1018. [[CrossRef](#)]
11. Zandbergen, P.A.; Ignizio, D.A. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartogr. Geogr. Inf. Sci.* **2010**, *37*, 199–214. [[CrossRef](#)]
12. Tobler, W.; Deichmann, U.; Gottsegen, J.; Maloy, K. World population in a grid of spherical quadrilaterals. *Int. J. Popul. Geogr.* **1997**, *3*, 203–225. [[CrossRef](#)]
13. Tobler, W.R. Smooth pycnophylactic interpolation for geographical regions. *J. Am. Stat. Assoc.* **1979**, *74*, 519–530. [[CrossRef](#)] [[PubMed](#)]
14. Mennis, J. Generating Surface Models of Population Using Dasymetric Mapping. *Prof. Geogr.* **2003**, *55*, 31–42.
15. Jia, P.; Qiu, Y.; Gaughan, A.E. A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida. *Appl. Geogr.* **2014**, *50* (Suppl. C), 99–107. [[CrossRef](#)]
16. Jia, P.; Gaughan, A.E. Dasymetric modeling: A hybrid approach using land cover and tax parcel data for mapping population in Alachua County, Florida. *Appl. Geogr.* **2016**, *66*, 100–108. [[CrossRef](#)]
17. Sutton, P.; Roberts, D.; Elvidge, C.; Baugh, K. Census from heaven: An estimate of the global human population using night-time satellite imagery. *Int. J. Remote Sens.* **2001**, *22*, 3061–3076. [[CrossRef](#)]
18. Sutton, P. Modeling population density with night-time satellite imagery and GIS. *Comput. Environ. Urban Syst.* **1997**, *21*, 227–244. [[CrossRef](#)]
19. Sutton, P.; Roberts, D.; Elvidge, C.; Meij, H. A comparison of nighttime satellite imagery and population density for the continental United States. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 1303–1313.

20. Sutton, P.C.; Elvidge, C.D.; Obremski, T. Building and evaluating models to estimate ambient population density. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 545–553. [[CrossRef](#)]
21. Zhuo, L.; Ichinose, T.; Zheng, J.; Chen, J.; Shi, P.J.; Li, X. Modelling the population density of China at the pixel level based on DMSP/OLS non-radiance-calibrated night-time light images. *Int. J. Remote Sens.* **2009**, *30*, 1003–1018. [[CrossRef](#)]
22. Amaral, S.; Monteiro, A.M.V.; Camara, G.; Quintanilha, J.A. DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. *Int. J. Remote Sens.* **2006**, *27*, 855–870. [[CrossRef](#)]
23. Elvidge, C.D.; Cinzano, P.; Pettit, D.R.; Arvesen, J.; Sutton, P.; Small, C.; Nemani, R.; Longcore, T.; Rich, C.; Safran, J.; et al. The Nightsat mission concept. *Int. J. Remote Sens.* **2007**, *28*, 2645–2670. [[CrossRef](#)]
24. Levin, N.; Duke, Y. High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sens. Environ.* **2012**, *119*, 1–10. [[CrossRef](#)]
25. Tan, M.; Li, X.; Li, S.; Xin, L.; Wang, X.; Li, Q.; Li, W.; Li, Y.; Xiang, W. Modeling population density based on nighttime light images and land use data in China. *Appl. Geogr.* **2018**, *90*, 239–247. [[CrossRef](#)]
26. Small, C.; Pozzi, F.; Elvidge, C. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sens. Environ.* **2005**, *96*, 277–291. [[CrossRef](#)]
27. Liu, Y.; Delahunty, T.; Zhao, N.; Cao, G. These lit areas are undeveloped: Delimiting China's urban extents from thresholded nighttime light imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 39–50. [[CrossRef](#)]
28. Briggs, D.J.; Gulliver, J.; Fecht, D.; Vienneau, D.M. Dasyetric modelling of small-area population distribution using land cover and light emissions data. *Remote Sens. Environ.* **2007**, *108*, 451–466. [[CrossRef](#)]
29. Zeng, C.; Zhou, Y.; Wang, S.; Yan, F.; Zhao, Q. Population spatialization in China based on night-time imagery and land use data. *Int. J. Remote Sens.* **2011**, *32*, 9599–9620. [[CrossRef](#)]
30. Wang, L.; Wang, S.; Zhou, Y.; Liu, W.; Hou, Y.; Zhu, J.; Wang, F. Mapping population density in China between 1990 and 2010 using remote sensing. *Remote Sens. Environ.* **2018**, *210*, 269–281. [[CrossRef](#)]
31. Yang, X.; Yue, W.; Gao, D. Spatial improvement of human population distribution based on multi-sensor remote-sensing data: An input for exposure assessment. *Int. J. Remote Sens.* **2013**, *34*, 5569–5583. [[CrossRef](#)]
32. Zhang, C.; Qiu, F. A Point-Based Intelligent Approach to Areal Interpolation. *Prof. Geogr.* **2011**, *63*, 262–276. [[CrossRef](#)]
33. Tatem, A.J.; Gaughan, A.E.; Stevens, F.R.; Patel, N.N.; Jia, P.; Pandey, A.; Linard, C. Quantifying the effects of using detailed spatial demographic data on health metrics: A systematic analysis for the AfriPop, AsiaPop, and AmeriPop projects. *Lancet* **2013**, *381*, S142. [[CrossRef](#)]
34. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
35. Gatrell, A.C.; Bailey, T.C.; Diggle, P.J.; Rowlingson, B.S. Spatial point pattern analysis and its application in geographical epidemiology. *Trans. Inst. Br. Geogr.* **1996**, 256–274. [[CrossRef](#)]
36. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman & Hall: London, UK; CRC Press: Boca Raton, FL, USA, 1986; Volume 26.
37. Bailey, T.C.; Gatrell, A.C. *Interactive Spatial Data Analysis*; Longman: Essex, UK, 1995.
38. Chainey, S. Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *Bull. Geogr. Soc. Liege* **2013**, *60*, 7–19.
39. Fotheringham, A.S.; Brunsdon, C.; Charlton, M. *Quantitative geography: Perspectives on Spatial Data Analysis*; SAGE: Thousand Oaks, CA, USA, 2000.
40. Lin, Y.-P.; Chu, H.-J.; Wu, C.-F.; Chang, T.-K.; Chen, C.-Y. Hotspot Analysis of Spatial Environmental Pollutants Using Kernel Density Estimation and Geostatistical Techniques. *Int. J. Environ. Res. Public Health* **2011**, *8*, 75–88. [[CrossRef](#)]
41. Yu, W.; Ai, T.; Shao, S. The analysis and delimitation of Central Business District using network kernel density estimation. *J. Transp. Geogr.* **2015**, *45*, 32–47. [[CrossRef](#)]
42. Demšar, U.; Harris, P.; Brunsdon, C.; Fotheringham, A.S.; McLoone, S. Principal Component Analysis on Spatial Data: An Overview. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 106–128. [[CrossRef](#)]
43. Bai, Z.; Wang, J.; Wang, M.; Gao, M.; Sun, J. Accuracy Assessment of Multi-Source Gridded Population Distribution Datasets in China. *Sustainability* **2018**, *10*, 1363. [[CrossRef](#)]
44. Harvey, J. Estimating census district populations from satellite imagery: Some approaches and limitations. *Int. J. Remote Sens.* **2002**, *23*, 2071–2095. [[CrossRef](#)]

45. Langford, M. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Comput. Environ. Urban Syst.* **2006**, *30*, 161–180. [[CrossRef](#)]
46. Cockx, K.; Canters, F. Incorporating spatial non-stationarity to improve dasymetric mapping of population. *Appl. Geogr.* **2015**, *63*, 220–230. [[CrossRef](#)]
47. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R.; Davis, C.W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **1997**, *18*, 1373–1379. [[CrossRef](#)]
48. Deville, P.; Linard, C.; Martin, S.; Gilbert, M.; Stevens, F.R.; Gaughan, A.E.; Blondel, V.D.; Tatem, A.J. Dynamic population mapping using mobile phone data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15888. [[CrossRef](#)] [[PubMed](#)]
49. Patel, N.N.; Stevens, F.R.; Huang, Z.; Gaughan, A.E.; Elyazar, I.; Tatem, A.J. Improving Large Area Population Mapping Using Geotweet Densities. *Trans. Gis* **2017**, *21*, 317–331. [[CrossRef](#)] [[PubMed](#)]
50. Stevens, F.R.; Gaughan, A.E.; Linard, C.; Tatem, A.J. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS ONE* **2015**, *10*, 107042. [[CrossRef](#)] [[PubMed](#)]
51. Tian, Y.; Zhou, Q.; Fu, X. An Analysis of the Evolution, Completeness and Spatial Patterns of OpenStreetMap Building Data in China. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 35. [[CrossRef](#)]
52. Lu, D.; Tian, H.; Zhou, G.; Ge, H. Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. *Remote Sens. Environ.* **2008**, *112*, 3668–3679. [[CrossRef](#)]
53. Kuang, W.; Liu, J.Y.; Zhang, X.; Lu, D.; Xiang, B. Spatiotemporal dynamics of impervious surface areas across China during the early 21st century. *Chin. Sci. Bull.* **2013**, *58*, 1691–1701. [[CrossRef](#)]
54. Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira Jr, J.; Pereira, F.C. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36–46. [[CrossRef](#)]
55. Liu, X.; He, J.; Yao, Y.; Zhang, J.; Liang, H.; Wang, H.; Hong, Y. Classifying urban land use by integrating remote sensing and social media data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1675–1696. [[CrossRef](#)]
56. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [[CrossRef](#)]
57. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sens.* **2016**, *8*, 151. [[CrossRef](#)]
58. Cai, J.; Huang, B.; Song, Y. Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sens. Environ.* **2017**, *202*, 210–221. [[CrossRef](#)]
59. Jia, P. Integrating kindergartener-specific questionnaires with citizen science to improve child health. *Front. Public Health.* **2018**, *6*, 236. [[CrossRef](#)] [[PubMed](#)]

