Article

Grid-Scale Poverty Assessment by Integrating High-Resolution Nighttime Light and Spatial Big Data—A Case Study in the Pearl River Delta

Minying Li, Jinyao Lin *, Zhengnan Ji, Kexin Chen and Jingxi Liu

School of Geography and Remote Sensing, Guangzhou University, Guangzhou 510006, China; 2001400019@e.gzhu.edu.cn (M.L.); 2001400041@e.gzhu.edu.cn (Z.J.); 2001400020@e.gzhu.edu.cn (K.C.); 32101400012@e.gzhu.edu.cn (J.L.)

* Correspondence: ljy2012@gzhu.edu.cn

Abstract: Poverty is a social issue of global concern. Although socioeconomic indicators can easily reflect poverty status, the coarse statistical scales and poor timeliness have limited their applications. While spatial big data with reasonable timeliness, easy access, and wide coverage can overcome such limitations, the integration of high-resolution nighttime light and spatial big data for assessing relative poverty is still limited. More importantly, few studies have provided poverty assessment results at a grid scale. Therefore, this study takes the Pearl River Delta, where there is a large disparity between the rich and the poor, as an example. We integrated Luojia 1-01, points of interest, and housing prices to construct a big data poverty index (BDPI). To evaluate the performance of the BDPI, we compared this new index with the traditional multidimensional poverty index (MPI), which builds upon socioeconomic indicators. The results show that the impoverished counties identified by the BDPI are highly similar to those identified by the MPI. In addition, both the BDPI and MPI gradually decrease from the center to the fringe of the study area. These two methods indicate that impoverished counties were mainly distributed in ZhaoQing, JiangMen and HuiZhou Cities, while there were also several impoverished parts in rapidly developing cities, such as CongHua and HuaDu Counties in GuangZhou City. The difference between the two poverty assessment results suggests that the MPI can effectively reveal the poverty status in old urban areas with convenient but obsolete infrastructures, whereas the BDPI is suitable for emerging-development areas that are rapidly developing but still lagging behind. Although BDPI and MPI share similar calculation procedures, there are substantial differences in the meaning and suitability of the methodology. Therefore, in areas lacking accurate socioeconomic statistics, the BDPI can effectively replace the MPI to achieve timely and fine-scale poverty assessment. Our proposed method could provide a reliable reference for formulating targeted poverty-alleviation policies.

Keywords: poverty assessment; spatial big data; nighttime light; remote sensing; targeted poverty alleviation

1. Introduction

Poverty is a global social problem, and poverty eradication is the first of the 17 sustainable development goals proposed by the United Nations [1–5]. The issue of poverty is particularly prominent in China. Although all previous national-level impoverished counties in China were lifted out of poverty in 2020, China is still facing a serious problem of relative poverty and a huge disparity between the rich and the poor [6–9]. Unlike absolute poverty (e.g., food poverty), relative poverty refers to individuals whose income can ensure food supplies only and cannot meet other basic living needs [10–13]. Therefore, during the relative poverty stage with multidimensional poverty as the main feature, it is of great importance to conduct poverty assessments to ensure the effectiveness and sustainability of targeted poverty alleviation.
Traditional poverty assessments rely mainly on official statistics or field surveys [14–18]. Since such methods cannot comprehensively reflect poverty status, many studies have gradually constructed different poverty assessment systems by considering various factors (e.g., personal income and consumption level) [19–24]. These attempts include the human poverty index, multidimensional poverty index (MPI), and multiple deprivation index established by the United Nations Development Programme considering three dimensions of health, education, and living standards. For example, Li et al. [25] used multidimensional statistical data to analyze the poverty status at county scales in China from 2000 to 2010 and revealed the spatial distribution characteristics of poverty over this period. Previous studies have demonstrated the effectiveness of various poverty indices from the perspectives of multidimensional poverty and relative poverty. Although the traditional statistical data at the county level in China are complete and accessible, the coarse statistical scales and poor timeliness have limited their applications. For example, the complete statistical data of 2023 will be available only one or two years later. In addition, there exist some inconsistencies caused by the change in administrative units [26–31].

To overcome these limitations, remote sensing data have been increasingly used to assess socioeconomic conditions and poverty status [32–40]. For example, Yong et al. [41] combined DMSP-OLS and NPP-VIIRS data to assess poverty in Southwest China from 2000 to 2019. Liu et al. [42] used NPP-VIIRS data to identify relative poverty in the surrounding areas of Beijing and Tianjin during the period of 2012–2020. Yu et al. [43] constructed the average light index using NPP-VIIRS data and then analyzed the correlation between the average light index and poverty status. Pan et al. [44] used nighttime light data to build a multidimensional poverty model and verified the validity of the average light index using linear regression. However, as mentioned above, poverty is reflected in not only the economic dimension but also other aspects, such as health, education, and environment. Since nighttime light data reflect only nighttime economic activities, it is difficult to comprehensively measure poverty based on nighttime light data alone [45–53]. For example, Pokhriyal et al. [54] pointed out that it is necessary to integrate multidimensional data so that poverty status can be reflected more accurately.

To this end, spatial big data have been gradually considered to overcome the limitations of nighttime light data. In particular, a number of previous studies have shown that POIs can effectively reflect socio-economic activities and urban vitality [51,55], while housing price is also an important indicator of economic and residential conditions [56,57]. Therefore, it is feasible to combine these various different data sources for fine-scale poverty estimation. For example, Ni et al. [58] indicated that combining nighttime lights with daytime remote sensing data can improve the accuracy of poverty prediction. Shi et al. [59] combined topography, vegetation index, points of interest (POIs), and nighttime light data to identify poverty in Chongqing, China. Lin et al. [60] found that POI data can reflect urban poverty to a certain extent, and thus, a combined use of POI and nighttime light data can improve the accuracy of poverty assessment. Although spatial big data have been increasingly used in poverty assessments, previous studies have relied greatly on the coarse-resolution DMSP-OLS and NPP-VIIRS data, and most of them have focused on regions with absolute poverty [61–66]. More importantly, few studies have provided poverty assessment results at a grid scale. Therefore, it is still necessary to analyze the performance of multisource spatial big data in assessing regions with relative poverty [67–71].

In fact, housing price is an important indicator for measuring the level of regional economy and living conditions [56,57]. Housing expenditure varies greatly among different income groups [72]. For example, Delang and Ho [73] analyzed the influence of public housing policy on the phenomenon of poverty concentration and found that poor people tend to concentrate in public housing areas that do not require much expenditure. Nevertheless, traditional housing price data were usually collected based on official statistics, censuses, and surveys, which cannot provide accurate fine-scale information [72,74]. For example, Yi and Huang [75] stressed the lack of housing price data in China. In this regard, online house price data can offer the possibility to measure poverty degrees at the micro
scale [74,76]. Previous studies have also demonstrated that advanced machine learning techniques can be used to estimate fine-scale housing prices based on proxy variables in regions where offline housing transaction data are not available [76,77]. Therefore, this research employed online housing price data as one of the poverty indicators, which can compensate for the deficiency of nighttime light data and measure relative poverty more comprehensively.

In summary, this study aims to apply multisource spatial big data to relative poverty assessment. A big data poverty index (BDPI) was constructed by integrating high-resolution Luoja 1-01 data, housing prices, and POIs. In addition, we also constructed the traditional MPI by selecting 17 indicators from six dimensions (human capital, natural capital, financial capital, physical capital, social capital, and environmental vulnerability) according to the vulnerability–sustainable livelihood analysis framework established by the Department for International Development (DFID). We compared the poverty assessment results of the BDPI and MPI to analyze the performance of these two methods. We took the Pearl River Delta in China, where the disparity between the rich and the poor is huge, as a case study area. The conclusions could provide support for local governments to formulate poverty-alleviation policies and implement development plans for underdeveloped areas.

2. Case Study and Data

2.1. Study Area

In this study, county-level administrative districts were regarded as the research unit. For clarity, we referred to all county-level administrative districts as “counties” in this manuscript. The study area includes 50 counties in the nine cities of the Pearl River Delta, of which Dongguan and Zhongshan were special cities with no counties (Figure 1). Located on the southern coast of China, the Pearl River Delta is one of the most populated areas in the Asia-Pacific region [78]. As an economically developed area, the Pearl River Delta generates 85% of Guangdong Province’s GDP and accounts for 70% of this province’s population. However, some counties in the Pearl River Delta have witnessed different degrees of relative poverty. Since it is extremely important to accurately identify relative poverty within developed regions, taking the Pearl River Delta as a case study is of universal significance [79–81].

Figure 1. Administrative divisions of the Pearl River Delta.
2.2. Data

This study adopted data that are easy to obtain and have reasonable timeliness, including: (1) high-resolution LuoJia 1-01 nighttime light data (with a spatial resolution of 130 m) in 2018 from Wuhan University (http://www.hbeos.org.cn/, accessed on 6 May 2023) [82–84]. This high-quality dataset has been pre-processed by constructing the geometric processing model of the whole link network for image-to-ground target positioning, and has been geo-referenced based on the ground control points participating in the processing of regional network adjustment. Therefore, the original spatial positioning errors have been adequately corrected. Lastly, the background noise for LuoJia 1-01 nighttime light data was removed by using the same period of NPP/VIIRS image as the mask data; (2) POI data in 2018 from AutoNavi Map (https://www.amap.com/, accessed on 6 May 2023), including corporates, scenic spots, scientific, educational and cultural services, transportation facilities, living facilities, and sports and leisure services. The attributes of the POIs include the specific geographic location information and detailed information on service facilities; (3) housing price data in 2021 and 2022 from Anjuke (https://www.anjuke.com/, accessed on 6 May 2023) and Loupan (https://www.loupan.com/, accessed on 6 May 2023), two of the most influential real estate platforms in China. All the housing price data were captured through the web crawler, and each data record includes name, area, price, address, and coordinate. A total of 64,750 valid records were obtained; (4) data on social, economic, population and other basic indicators from the 2019 County-level Statistical Yearbook and the Seventh Population Census in China, including population aged 0–14 years and over 65 years old, rural population, ethnic minority population, proportion of population with primary education, proportion of population with secondary education, proportion of population with college or higher education, illiterate population over 15 years old, average years of education completed, number of professional technicians, number of people engaged in agriculture, forestry, animal husbandry and fishery, number of tertiary sectors, income in local government budget, and proportion of jobholders; (5) digital elevation model (DEM) (with a spatial resolution of 30 m) and precipitation data (with a spatial resolution of 1 km) in 2018 from the China Earth System Science Data Center (http://www.geodata.cn/, accessed on 6 May 2023); and (6) road network data in 2018 from OpenStreetMap (https://download.geofabrik.de/, accessed on 6 May 2023), including main roads and other roads.

It should be noted that there is only one high-quality pre-processed dataset of LuoJia 1-01 nighttime light remote sensing image, from the year 2018. To reduce the negative influence caused by the inconsistent data year, we collected data from the same year as much as possible. However, the housing price platforms only provide the data of the latest year; thus, we cannot obtain the housing price data of the study area in 2018. At the same time, the most recent population census data of China come from the seventh population census. Therefore, we have to adopt the house price data of 2021–2022 as well as the census data of 2019.

2.2.1. LuoJia 1-01

The radiance of LuoJia 1-01 nighttime light data was converted using the following formula given on the official website [85–88]:

\[ L = DN^\frac{3}{2} \cdot 10^{-5} \]  

(1)

where \( L \) is the corrected radiance (W/ (m\(^2\)·sr·µm)), and \( DN \) is the original digital number value of the data. The mean, maximum, standard deviation, sum, and Moran’s \( I \) Index of the converted nighttime light data for each county were regarded as the nighttime light-related indicators (Table 1).
Table 1. Indicators for measuring nighttime light data.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average value</td>
<td>( L/G, ) ( L ) means the total values of all land use grids, ( G ) means the number of grids</td>
</tr>
<tr>
<td>Average light index</td>
<td>( L/G_0, ) ( G_0 ) means the number of land use grids in lighting parts</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}, \ x_i ) means the value for the ( i )th grid, ( \bar{x} ) means the average score for every grid</td>
</tr>
<tr>
<td>Maximum value</td>
<td>Maximum value for every land use grid</td>
</tr>
<tr>
<td>Moran’s I Index</td>
<td>Relationship between pixel light values for each county, which represents spatial correlation</td>
</tr>
</tbody>
</table>

2.2.2. POI

POI can effectively reflect socio-economic activities and urban vitality. The POI data were preprocessed to eliminate incomplete and inconsistent data. Then, we calculated the POI densities of various categories for each county.

2.2.3. Housing Price

Housing prices can comprehensively reflect the built environment, such as housing conditions, public facility configurations, traffic conditions, as well as the economic conditions of residents [89,90]. While online housing price data are available at the community level, traditional statistical data are usually available at the county level. In order to compare the performance of the MPI and BDPI, it is necessary to rescale the housing price data to the county level. Therefore, we collected housing price data and then calculated the average housing price in each county.

3. Method

As shown in Figure 2, this study includes five major steps: (1) dimensionless processing of all data; (2) reducing the dimensionality of 11 indicators related to nighttime light data, housing prices, and POIs by principal component analysis (PCA) to obtain the BDPI scores; (3) reducing the dimensionality of 17 indicators related to socioeconomic statistics by factor analysis to obtain the MPI scores; (4) comparing the spatial characteristics of these two poverty assessment results; (5) using random forest to estimate grid-scale housing price data according to road network and POI; (6) constructing grid-scale BDPI by combining grid-scale house price data, POI, and nighttime light; and (7) using the BDPI to measure grid-scale poverty conditions in the study area.

3.1. Big Data Poverty Index (BDPI)

This study integrated three types of indicators (nighttime light, housing price, POI) to construct the BDPI. First, the dimensionality of the original data was reduced by using PCA to eliminate duplicate information. Second, the dimensionally reduced data were assessed by using Kaiser–Meyer–Olkin (KMO) and Bartlett’s sphericity tests. A KMO score greater than 0.6 indicates a good performance of PCA (Table 2). Finally, the weight of each principal component was assigned based on the ratio of the variance contribution rate to the cumulative contribution rate of the principal component. Note that the sum of the variance contribution rates of all principal components is 100%. The formula for calculating the BDPI is given as follows:

\[
BDPI_i = \sum_{k=1}^{n} (a_k \times a_{ik})
\]

where \( a_{ik} \) is the score of indicator \( k \) for the \( i \)-th county; and \( a_k \) is the variance contribution rate of indicator \( k \). A lower value of BDPI suggests a higher degree of poverty.
Figure 2. Poverty index construction and poverty assessment based on multisource spatial big data.

Table 2. KMO and Bartlett’s test for the BDPI.

<table>
<thead>
<tr>
<th></th>
<th>KMO</th>
<th>Degree of freedom</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett’s test</td>
<td>0.635</td>
<td>55</td>
<td>0.000</td>
</tr>
</tbody>
</table>

PCA was performed on multisource spatial data to generate the BDPI. The indicators with a factor loading greater than 0.6 were extracted as the main contribution indicators of the principal component. The results show that the first three principal components should be preserved, as they can explain approximately 79.96% of the original information (Table 3).

Table 3. BDPI indicators, principal factor eigenvalues, explanatory contribution, and load matrix.

<table>
<thead>
<tr>
<th>Category</th>
<th>Indicator</th>
<th>Principal Component First</th>
<th>Second</th>
<th>Third</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nighttime light</td>
<td>Mean</td>
<td>0.952</td>
<td>0.639</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.847</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Moran’s I Index</td>
<td>-0.603</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POI</td>
<td>Corporates</td>
<td>0.875</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scenic spots</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scientific, educational and cultural services</td>
<td>0.689</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transportation facilities</td>
<td>0.938</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Living and leisure services</td>
<td>0.930</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing price</td>
<td>Average housing price</td>
<td>5.690</td>
<td>1.812</td>
<td>1.294</td>
</tr>
<tr>
<td></td>
<td>Eigenvalue of the principal component (%)</td>
<td>51.729</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contribution rate of the principal component (%)</td>
<td>31.729</td>
<td>46.18</td>
<td>11.763</td>
</tr>
<tr>
<td></td>
<td>Cumulative contribution rate of the principal components (%)</td>
<td>31.729</td>
<td>68.202</td>
<td>79.964</td>
</tr>
</tbody>
</table>
3.2. Multidimensional Poverty Index (MPI)

Based on the vulnerability-sustainable livelihood analysis framework and data availability, we adopted 17 indicators from 6 dimensions (human capital, natural capital, financial capital, physical capital, social capital, and environmental vulnerability). The dimensionality of the data was reduced by factor analysis, and the dimensionally reduced data were assessed by KMO and Bartlett’s tests. A KMO value greater than 0.6 indicates a good performance of the factor analysis (Table 4).

Table 4. KMO and Bartlett’s tests for the MPI.

<table>
<thead>
<tr>
<th></th>
<th>KMO</th>
<th>Degree of freedom</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett’s test</td>
<td>0.794</td>
<td>136</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The weight of each principal component was assigned based on the ratio of the variance contribution rate to the cumulative contribution rate of the component, and the sum of the variance contribution rates of all principal components is 100%. Finally, the MPI value of each county was calculated as follows:

\[ F_i = \sum_j W_j Y_{ij} \]  \hspace{1cm} (3)

where \( F_i \) is the MPI value of the \( i \)-th county, reflecting its poverty degree; \( W_j \) is the weight of the \( j \)-th common factor, which can be represented by the variance contribution rate of each common factor; and \( Y_{ij} \) is the single-item score of the \( j \)-th common factor in the \( i \)-th county, which can be calculated from the factor score coefficient matrix and the observed value of the variable. A lower value of MPI indicates a higher degree of poverty.

Factor analysis was performed on traditional statistical data to generate MPI (Table 5). The indicators with a factor loading greater than 0.9 were extracted as the main contribution indicators of each principal component. The maximum variance method was used to obtain the rotated component matrix. The cumulative explained variance after rotation reaches 89.51%, which indicates that the information of the original data can be well preserved.

Table 5. MPI indicators, principal factor eigenvalues, explanatory contribution, and load matrix.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
<th>Fifth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population aged 0–14 years and over 65 years old</td>
<td>0.917</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural population</td>
<td>0.894</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnic minority population</td>
<td>-0.595</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of population with primary education</td>
<td>-0.861</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of population with secondary education</td>
<td>-0.908</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of population with college or higher education</td>
<td>0.800</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illiterate population over 15 years old</td>
<td>0.829</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average years of education completed</td>
<td>0.737</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of professional technicians</td>
<td>0.764</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of people engaged in agriculture, forestry, animal husbandry and fishery</td>
<td>0.914</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3. Housing Price Estimation Based on Machine Learning

Since the valid housing price data did not completely cover the study area, we need to estimate the missing data based on advanced machine learning methods [91,92]. Random forest (RF) is one of the most effective and efficient machine learning methods and has been commonly used in various research fields [93,94]. Compared with other machine learning methods, RF can effectively handle high-dimensional data and missing values [95,96]. RF regression is an algorithm based on ensemble learning that completes the regression task by building multiple decision trees and integrating their prediction results.

In RF, each decision tree is trained independently using randomly selected subsamples, which effectively reduces the risk of overfitting. For each node of the decision tree, only part of the features is considered when selecting the best partition features to improve the robustness of the model. RF offers the final regression result by averaging the prediction results of multiple decision trees. In this study, the housing price estimation model was constructed based on RF by taking points of interest, road networks as independent variables and housing price data as dependent variable, according to previous findings [74,76,77].

Specifically, the distance to subway stations, distance to bus stops, distance to main roads, distance to ordinary roads, distance to medical facilities, distance to living facilities, and distance to school were used as independent variables, and the existing house price data were used as dependent variable. Then, the model trained by the RF was further used to estimate the unknown housing price data at the grid scale. Finally, the housing price data covering all the grids of the study area were obtained, and thus the grid-scale BDPI can be further constructed.

4. Results

4.1. BDPI Results

Figure 3 presents the results of the first, second, and third principal components and the associated BDPI. These three principal components can reveal the poverty status from different dimensions. Although the main components from the PCA may only have statistical meaning, researchers can interpret the substantive meaning of every principal component based on specialized knowledge [97]. Specifically, since the first component includes five high loading factors: mean, standard deviation, scenic spots, transportation facilities, and living and leisure services, it primarily represents the relative poverty caused by the lack of social capital. In addition, the second component exhibits high loadings on sum, which primarily indicates the relative poverty caused by the lack of economic

### Table 5. Cont.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
<th>Fifth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tertiary sectors</td>
<td>0.873</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income in local government budget</td>
<td>0.807</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of jobholders</td>
<td>0.587</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average elevation</td>
<td>0.948</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average topographic relief</td>
<td>0.926</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of areas with slopes greater than 15°</td>
<td>0.901</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average precipitation</td>
<td></td>
<td>0.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvalue of the principal component (%)</td>
<td>7.905</td>
<td>2.852</td>
<td>2.283</td>
<td>1.199</td>
<td>0.978</td>
</tr>
<tr>
<td>Rotated variance contribution rate of the principal component (%)</td>
<td>33.996</td>
<td>25.051</td>
<td>16.420</td>
<td>7.075</td>
<td>6.965</td>
</tr>
<tr>
<td>Cumulative variance contribution rate of the principal component (%)</td>
<td>33.996</td>
<td>59.047</td>
<td>75.466</td>
<td>82.541</td>
<td>89.506</td>
</tr>
</tbody>
</table>

Remote Sens. 2023, 15, 4618
capital. Similarly, the third component primarily reflects the relative poverty caused by the living environment.

Social capital (Figure 3a) can reflect the basic resources (e.g., transportation, education, and medical conditions) within the study area and is a key indicator of poverty status. We found that YueXiu County of GuangZhou City and FuTian County of ShenZhen City had higher social capital scores and belonged to highly urbanized areas. The counties with moderate to high social capital scores were clustered in ShenZhen City and the junction of GuangZhou and FoShan Cities. By comparison, the counties located on the fringe of the Pearl River Delta generally suffer a shortage of social capital, and most of them are in ZhaoQing, JiangMen and HuiZhou Cities.

Economic capital can reflect the degree of economic development within the study area. Figure 3b indicates that the central and eastern regions of the Pearl River Delta had higher economic capital scores, forming agglomeration centers with high economic potential. Most of the counties outside agglomeration centers had lower economic capital scores and relatively low degrees of economic development.

For the dimension of the living environment (Figure 3c), most counties had moderate to high scores, which suggests that the overall living environment in the Pearl River Delta is desirable. However, there still exist some counties with a relatively poor living environment, such as HuiCheng County of HuiZhou City and DeQing County of ZhaoQing City.

Remote Sens. 2023, 15, x FOR PEER REVIEW ...distributed in GuangZhou, ShenZhen, and ZhuHai Cities, all of which belong to the pilot zones of the Pearl River Delta.

Figure 3. Results of the (a) first, (b) second, (c) third principal components, and (d) BDPI.
Overall, the BDPI score generally gradually decreases from the center to the fringe of the study area, exhibiting an obvious spatial agglomeration pattern. The regions in the western Pearl River Delta were relatively poor since the counties with higher degrees of poverty were distributed mainly in ZhaoQing, JiangMen, and HuiZhou Cities. Nevertheless, rapidly growing cities also contain some impoverished counties, such as the CongHua and ZengCheng Counties of GuangZhou City.

4.2. MPI Results

In general, the MPI score in the Pearl River Delta also gradually decreases from the center to the fringe, and the further away from the center, the higher the poverty level (Figure 4). Relatively impoverished counties were mainly distributed in ZhaoQing, JiangMen, and HuiZhou Cities. By comparison, most well-developed counties were distributed in GuangZhou, ShenZhen, and ZhuHai Cities, all of which belong to the pilot zones of the Pearl River Delta.

4.3. Comparison of Two Poverty Assessment Results

The results of the MPI and BDPI were further compared to evaluate the rationality of the BDPI. Due to differences in the principles and calculation processes of these two methods, analyzing the ranking of the index scores (i.e., ranking according to the index score, and a lower ranking means a more severe poverty status) is more in line with the definition of relative poverty. Therefore, we quantified the consistency of these two results using the Wilcoxon signed-rank test and presented the visualization of the z-score in Figure 5. The significance level of the test is 0.898 (greater than 0.05), which indicates that there is no significant difference in the score ranking of the two poverty assessment
results. According to Figure 5, we also found that these two results show similar overall trends. In summary, the above validation has indicated a strong agreement between the poverty assessment results from the BDPI and MPI.

In addition, the top 20% and top 30% impoverished counties in these two results share coincidence degrees of 70% and 87%, respectively, which also demonstrates that the impoverished counties identified by the two methods are similar (Table 6). Nevertheless, we noticed that the impoverished counties identified by the MPI are concentrated in the fringe areas of the Pearl River Delta, while the BDPI can identify underdeveloped counties in some cities with rapid economic development, such as Guangzhou and Foshan. The difference between these two poverty assessment results suggests that the MPI can effectively reveal poverty status in old urban areas with convenient infrastructure but lagging behind, whereas the BDPI is suitable for emerging-development areas that are rapidly developing but still lagging behind.

Table 6. Consistency and difference between the results from BDPI and MPI.

<table>
<thead>
<tr>
<th>Top N%</th>
<th>City</th>
<th>Number of Impoverished County</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BDPI</td>
</tr>
<tr>
<td>Top 20%</td>
<td>Guangzhou</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Dongguan</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Zhongshan</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Foshan</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Huizhou</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Jiangmen</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Shenzhen</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Zhuhai</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ZhaoQing</td>
<td>5</td>
</tr>
<tr>
<td>Top 30%</td>
<td>Guangzhou</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Dongguan</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Zhongshan</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Foshan</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Huizhou</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Jiangmen</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Shenzhen</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Zhuhai</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>ZhaoQing</td>
<td>6</td>
</tr>
</tbody>
</table>
The proposed BDPI and traditional MPI have different applicabilities for identifying different types of impoverished counties. To reveal the differences between the two poverty assessment results, the MPI ranking of every county was subtracted from its BDPI ranking. The BDPI poverty degree of a county is more severe than its MPI poverty degree if the ranking difference of the county is positive. In contrast, the MPI poverty degree of the county is more severe than its BDPI poverty degree if the ranking difference of this county is negative. The result in Figure 6 shows that the counties with positive ranking differences were mainly located in FoShan, ZhongShan, DongGuan, and HuiZhou Cities, whereas the counties with negative ranking differences were generally located in ZhuHai, JiangMen, GuangZhou, and ShenZhen Cities.

![Figure 6. Differences between the two poverty assessment results.](image)

Among the counties with positive ranking differences, DingHu County, DongGuan City, and ZhongShan City had relatively larger differences. After the implementation of reform and opening-up policies, these areas rapidly promoted industries and handicrafts, and the supporting facilities gradually improved. They have considerable economic capital and can easily attract a large number of migrant workers. However, these areas contain a large number of urban villages with poor living spaces and low densities of basic living facilities. Therefore, the BDPI poverty degrees of these areas were more severe than the corresponding MPI poverty degrees.

Among the counties with negative ranking differences, JinWan County, GaoMing County, and YanTian County had relatively larger differences. Most of these areas were emerging-development zones with relatively complete basic living facilities, desirable living environments, and convenient transportation systems. However, they were still economically underdeveloped compared with the developed urban districts. Therefore, the MPI poverty degrees of these counties were more severe than the corresponding BDPI poverty degrees.
4.4. Grid-Scale BDPI-Based Poverty Assessment

The above comparisons have demonstrated that the proposed BDPI is feasible and effective for poverty assessment. Therefore, this study further estimated the poverty conditions at the grid scale by taking Guangzhou and Foshan as examples. These regions were divided by a grid of 500 m × 500 m. All the independent variables and the resultant grid-scale housing price data are shown in Figures 7 and 8, respectively. The training and testing accuracies of the RF model reached 0.9590 and 0.9035, respectively. This phenomenon has indicated that the RF model can effectively estimate missing housing price data at the grid scale.

Figure 7. Independent variables for housing price estimation.

Figure 8. Grid-scale housing price data.
Finally, the grid-scale BDPI-based poverty assessment result is displayed in Figure 9. This result was classified into five levels, and a lower level indicates a more impoverished region. It is found that the center parts are well developed. With the increase in distance from the urban center, an increasing number of counties are lagging behind to varying degrees, especially CongHua, ZengCheng, SanShui, and GaoMing Counties. For example, the southern part of the boundaries between BaiYun, LiWan, YueXiu, and TianHe Counties are in a better development condition with higher BDPI levels. By comparison, most of the eastern part of BaiYun County and NanHai and ShunDe Counties are relatively lagging behind. We also noticed that level 1 regions also exist around the regions with higher BDPI levels, such as the central and eastern parts of ShunDe County. In summary, the application of the BDPI can effectively identify relative poverty at a finer scale.

Figure 9. Grid-scale BDPI-based poverty assessment result.

5. Discussion and Conclusions

5.1. Advantages and Disadvantages of the BDPI

The above analysis shows that compared with the MPI, the BDPI proposed by this study has three major advantages. First, the BDPI was constructed by integrating multi-source spatial big data, including high-resolution Luojia 1-01 nighttime light, housing price, and POI data. Our experiments have suggested that the proposed BDPI can identify areas that are relatively lagging behind in emerging-development cities. Second, the BDPI can reduce the cost of poverty assessment and enhance the timeliness of results due to the use of easily accessible, low-cost, and rapidly updated data. Third, the BDPI can effectively reflect social, economic, living environment, and other conditions. Therefore, it provides a new approach to poverty assessment for developing countries lacking accurate socioeconomic statistics.

It should also be noted that the people in some less developed counties are more inclined to complete housing transactions offline, and these data cannot be acquired online. Fortunately, the difference between online and offline data is not significant in many regions. In addition, as mentioned above, previous studies have demonstrated that advanced
machine learning techniques can be used to estimate fine-scale housing prices based on proxy variables in regions where offline data are not available. Although there are uncertainties and limitations in data preparation, BDPI still has distinct advantages in fine-scale poverty assessment. Compared with MPI, which relies on dated indicators and periodic data collection, BDPI can capture changes in poverty status instantly and provide more timely support for policy making. In addition, the use of spatial data can reflect regional differences at a grid scale.

Notwithstanding these advantages, the BDPI still has the following disadvantages. First, online housing price data may not be available in some underdeveloped areas. In that case, advanced machine learning techniques and offline data, such as local government statistics and survey data, should be used to supplement the BDPI's data sources. Second, the indicators used to construct the BDPI also have some limitations. Poverty is a multidimensional phenomenon involving not only social, economic, and natural aspects but also humanities, policies, and other factors. Therefore, although the BDPI can be used as an indirect indicator, the associated assessment results may overestimate or underestimate the actual poverty status, and the BDPI needs to be further enhanced. Third, it is difficult for the BDPI to assess the poverty status from a long time ago due to the unavailability of historical data. The inconsistency of data years may affect the poverty assessment results to some degree. In future research, we will adopt long-term time-series data for poverty assessment to further test the applicability of the BDPI. Fourth, we will consider more spatial big data and higher-resolution nighttime light data, compare the poverty assessment results of the Pearl River Delta with other regions, and analyze the differences in poverty issues in those regions.

5.2. Policy Recommendations

In this research, the internal development of a city is unbalanced if this city contains both counties with positive ranking differences and counties with negative ranking differences, such as DingHu County and DuanZhou County in ZhaoQing City. DingHu County has an average living environment and a relatively remote location, but as a new developing zone of ZhaoQing City, this county has desirable social and economic resources. As the central urban area of ZhaoQing City, DuanZhou County has a satisfactory living environment and relatively complete supporting facilities. However, there is basically no more space for urban expansion since the amount of land available for development is decreasing gradually. Therefore, the urban development costs are relatively higher in DuanZhou County.

Overall, the poverty assessment results obtained in this study can offer the following policy recommendations for poverty alleviation. The huge disparity between the rich and the poor in the Pearl River Delta shows that poverty-alleviation operations require not only raising social income but also reducing inequality in regional development. For areas with a shortage of social resources, it is necessary to improve the diversity of their resources, for example, enhancing infrastructures such as those for medical care, public transportation, and education, to improve urban vitality. For the fringe areas of the Pearl River Delta, which are greatly affected by natural conditions, the development strategy should be carefully designed according to local characteristics. In relatively underdeveloped areas of rapidly developing cities, the government needs to offer more support, carry out detailed strategic planning, and allocate more socioeconomic resources to areas with unbalanced development to reduce the disparity between the rich and the poor. In addition, the government should further upgrade the poverty assessment method in the context of big data and improve the accuracies and pertinence of the assessment results. Taking the Pearl River Delta region as an example, this study has demonstrated that the proposed BDPI could provide technical support for poverty assessment in many other regions.
5.3. Main Conclusions

The significance and novelty of this research are twofold. Firstly, high-resolution nighttime light and spatial big data have been integrated to construct the novel big data poverty index (BDPI) for relative poverty assessment. Secondly, a poverty assessment has been conducted in the regions with relative poverty at a grid scale after the validation of the BDPI. Both the traditional MPI and BDPI were used to assess the poverty status of the Pearl River Delta, where there is a considerable disparity between the rich and the poor. We analyzed the differences in the poverty assessment results obtained by the two methods. These two methods generated similar assessment results, which verify the effectiveness of the BDPI. Overall, the results of the two poverty assessment methods show that the poverty index gradually decreases from the center to the fringe of the study area. The counties in the western Pearl River Delta were relatively impoverished, and even fast-growing cities also contained some impoverished counties.

The statistical-based MPI is more suitable for old urban areas with convenient but obsolete infrastructures, while the BDPI is more suitable for emerging-development areas that are rapidly developing but still lagging behind. Therefore, combining these two types of indices for poverty assessment can yield optimal results if the data are well-prepared. The BDPI can also successfully replace the traditional MPI if the statistical data are not updated in time. The BDPI proposed by this study can effectively assess regional poverty status and provide a new method of poverty assessment for developing countries lacking accurate socioeconomic statistics. The results of this study could help local governments to fully understand the multidimensional characteristics of poverty and could provide decision support for formulating targeted poverty-alleviation policies.

Author Contributions: Conceptualization, J.L. (Jinyao Lin); methodology, M.L.; validation, M.L.; formal analysis, Z.J. and K.C.; investigation, Z.J. and K.C.; data curation, J.L. (Jingxi Liu); writing—original draft preparation, M.L.; visualization, M.L. and J.L. (Jingxi Liu); supervision, J.L. (Jinyao Lin); project administration, J.L. (Jinyao Lin); funding acquisition, J.L. (Jinyao Lin). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangdong Philosophy and Social Science Foundation (Grant No. GD23XSH11), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023A1515030300), the National Natural Science Foundation of China (Grant No. 41801307), and the Guangzhou Municipal Science and Technology Project (Grant No. 202201010289).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References


Tselios, V.; Stathakis, D.; Faraslis, I. Concentration of populations and economic activities, growth, and convergence in Europe using satellite-observed lighting. *Geoarrow Int.* **2020**, *35*, 1527–1552. [CrossRef]


Han, Y.; Liu, L.; Sui, Q.; Zhou, J. Big Data Spatio-Temporal Correlation Analysis and LRM Model Based Targeted Poverty Alleviation through Education. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 837. [CrossRef]


He, C.; Ma, Q.; Liu, Z.; Zhang, Q. Modeling the spatiotemporal dynamics of electric power consumption in Mainland China using saturation-corrected DMSP/OLS nighttime stable light data. *Int. J. Digit. Earth* **2014**, *7*, 993–1014. [CrossRef]

17 of 20


47. Gibson, J.; Olivia, S.; Boe-Gibson, G.; Li, C. Which night lights data should we use in economics, and where? J. Dev. Econ. 2021, 149, 102602. [CrossRef]


50. Puttanapong, N.; Martinez, A.; Bulan, J.A.N.; Addawe, M.; Durante, R.L.; Marillan, M. Predicting Poverty Using Geospatial Data in Thailand. ISPRS Int. J. Geo-Inf. 2022, 11, 293. [CrossRef]


60. Lin, J.; Luo, S.; Huang, Y. Poverty estimation at the county level by combining Luojia1-01 nighttime light data and points of interest. Geocarto Int. 2022, 37, 3590–3606. [CrossRef]


66. Xu, J.; Song, J.; Li, B.; Liu, D.; Cao, X. Combining night time lights in prediction of poverty incidence at the county level. Appl. Geogr. 2021, 135, 102552. [CrossRef]


68. Xu, Y.; Mo, Y.; Zhu, S. Poverty Mapping in the Dian-Gui-Qian Contiguous Extremely Poor Area of Southwest China Based on Multi-Source Geospatial Data. Sustainability 2021, 13, 8717. [CrossRef]


72. Krafft, N.J.; Wurm, M.; Taubenböck, H. Housing forms of poverty in Europe—A categorization based on literature research and satellite imagery. Appl. Geogr. 2022, 149, 102820. [CrossRef]

73. Delang, C.O.; Lung, H.C. Public Housing and Poverty Concentration in Urban Neighbourhoods: The Case of Hong Kong in the 1990s. Urban Stud. 2010, 47, 1391–1413. [CrossRef]


76. Zhou, X.; Tong, W.; Li, D. Modeling Housing Rent in the Atlanta Metropolitan Area Using Textual Information and Deep Learning. ISPRS Int. J. Geo-Inf. 2019, 8, 349. [CrossRef]

77. Yao, Y.; Zhang, J.; Hong, Y.; Liang, H.; He, J. Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. Trans. GIS 2018, 22, 561–581. [CrossRef]


79. Yuan, Y.; Wu, F. The development of the index of multiple deprivations from small-area population census in the city of Guangzhou, PRC. Habitat Int. 2014, 41, 142–149. [CrossRef]


81. Lin, J.; Zhang, W.; Wen, Y.; Qiu, S. Evaluating the association between morphological characteristics of urban land and pluvial floods using machine learning methods. Sustain. Cities Soc. 2023, 104891. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.