ISEIS

Journal of Environmental Informatics 34(1) 15-27 (2019)

Journal of Environmental Informatics

www.iseis.org/jei

China Population Distributions at Multiple Geographical Scales and Their Correlates

X. Liu[1], J. F. Wang[2]*, G. Christakos[3], and Y. L. Liao[2]

[1]*Curtin University Sustainability Policy Institute (CUSP), Curtin University, Bentley WA 6102, Australia*
[2]*The State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Nature Resources Research, Chinese Academy of Sciences, Beijing 100101, China*
[3]*Institute of Island & Coastal Ecosystems, Ocean College, Zhejiang University, Hangzhou, 310058, China; and Department of Geography, San Diego State University, California, USA*

Received 30 March 2016; revised 5 June 2017; accepted 25 November 2018; published online June 23 2019

ABSTRACT. Most population distribution studies have focused on a single spatial resolution scale, thus, leading to a limited representation of the "real-world". The present work, instead, proposes a "zoom lens" approach to detect and remove localized variation while retaining the general population distribution trends at multiple resolution scales. Different "focal length optics" are able to eliminate the unnecessary spatial details and filter out the underlined trend at the specified resolution. As spatial resolution scale decreases, the general trend and local variation of population distribution can be identified from the fine resolution to the national level. On the basis of small-scale analysis, it was shown that high-density population in China is roughly aligned with an oblique trend line along the Heihe–Tengchong Line, which provides a mathematical foundation for it. It was also found that the positive relationship between correlates and population distribution became more significant at the national scale. Topographic elevation has the largest negative impact on population distribution at the country level, whereas water accessibility has the largest effect on population distribution at any resolution. Furthermore, by combining the "zoom lens" approach with geographic weighted regression, the population distribution correlates (main roads, railways, live green vegetation, elevation, relief amplitude, rivers and lakes) were studied. A significant deterioration of accessibility to main roads and water in certain areas was identified at the national scale, which was not detected without "zoom lens" approach. Therefore, this study demonstrated that correlation or any other relationship may vary at different spatial scales of study.

Keywords: China population distributions, correlates of population distribution, multiple geographical scales, spatial analysis, spatial filter

1. Introduction

Accurate determination of the population distribution is important in human population studies – such as public health, urban and regional development, environmental planning, homeland security, and even intergeneration relationship. Both scientific analysis and decision making are based on accurate population estimation (Rogerson and Kim, 2005; Bhaduri et al., 2007; Badescu, 2008; Huang et al., 2009; Tong et al., 2012). For example, Yin et al. (2005) pointed out that population distribution can serve as a good urbanization indicator. As a result, a number of studies have been conducted to determine the spatial distribution of human population (Small and Naumann, 2001; Yue et al., 2005; Maantay et al., 2007).

However, representing real-world population distributions has always been a challenging affair (Bracken, 1993; Martin and Higgs, 1996; Ingram, 1998; Gaughan et al., 2015). This is mainly due to estimation heterogeneities at different study resolution scales (Linard et al., 2012). Visualizing the population distribution is not only for descriptive purposes – density and geographic extent of population – but also for assessing potential risks and refining population policy (Gregory, 2000; Moon and Farmer, 2001; Poulsen and Kennedy, 2004; Sleeter, 2004; Maantay et al., 2007). The simplest and most intuitive way is the de facto standard, which is to evenly distribute the head count within the administrative census tracts (Lutz and Samir, 2010). However, in reality the internal distribution of population is spatially heterogeneous (Langford et al., 2008). Normally, residential housing is concentrated in towns and cities, whereas remote areas are essentially devoid of population.

Dasymetric mapping can be adopted to avoid this problem, since it is "…a technique that involves estimating the distribution of aggregated data within the units of analysis, by adding additional information that provides insights on how these data are potentially distributed" (Poulsen and Kennedy, 2004). The most significant advantage of this technique is to maximize the faithful representation of within-zone uniformity. Maantay et al. (2007) tried to improve the estimation of total population depiction on a dasymetric map by considering the ancillary information, such as cadastral data, land-use filters,

* Corresponding author. Tel.: +86 10 64888965; fax: +86 10 64889630.
E-mail address: wangjf@lreis.ac.cn (J. F. Wang).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201900414

routines, and other data. These methods can be categorized as surface modeling.

Yin et al. (2005) established the relationship between built up area and population density pattern, and then measured the population distribution by using remote sensing technology. Luo and Wei (2006) advanced Yin et al. (2005)'s study by considering one more factor – urban land parcel – to achieve accurate estimation of the population density pattern. Li and Weng (2005) applied stepwise regression analysis to examine factors extracted from remote sensing images that may be related to population density. This method is a possible way to identify suitable variables in a population estimation model. However, determining the structure and model coefficients prior to model development is not an easy task.

With the development of remote sensing technologies, population density can be estimated from the remote sensing images. Sutton et al. (2001) tried to calculate the population using night time satellite images by analyzing the relationship among areal extent, population and light area on the images. Building height has also considered a population distribution indicator determined from the digital terrain model (DTM) and the digital surface model (DSM) (Ural et al., 2011). Wesolowski et al. (2012) quantified the spatial variation of human distribution and malaria mobility using mobile phone data at regional spatial scales.

Recently, McKee et al. (2015) presented a spatially explicit population projection method. However, this method focuses only at a single resolution scale and relies overwhelmingly on the accurate determination of components that affect population distribution. Therefore, although different surface modeling methods were used to improve the determination accuracy, it is necessary that they can be tested in different real-world situations and at different spatial resolutions.

The spatial extent of these previous studies ranged from the city level (Luo and Wei, 2006) to the global level (Balk and Yetman, 2004), but without considering correlations between the different levels. Also, most traditional "multiple spatial scales" studies can be described as the research extents variation. In this work, a method is presented that can analyze the spatial population distribution at multiple resolution levels. In particular, while previous studies applied "prime lens" to determine population distribution, the present work adopts a "zoom lens" approach to study the spatial distribution of a population. An advantage of this approach is that a variety of population distribution scenarios can be estimated using different "focal length optics". They are able to eliminate the unnecessary spatial details and filter out the underlined trend at the specified resolution.

Correlation analysis between population distribution and its potential factors can also benefit from the "zoom lens" approach. However, this is rarely discussed in previous studies. To assess the impact of its correlates on a population distribution, Cohen and Small (1998) quantified the relationship between population distribution and the global elevation above mean sea level. Feng et al. (2008) analyzed the correlation between population distribution and relief degree of land surface.

Cincotta et al. (2000) estimated the relationship between world population density and the biodiversity hotspots. Population density was identified as a risk factor in areas rich in endemic species. The correlation between human presence and species richness at a coarse scales has also been reported (Chown et al., 2003). Pautasso (2007) improved the results of the previous studies by determining resolution dependency on the relationship between population density and live green vegetation, but ignored the impact of other factors, such as transport and water resource, that might contribute to population density.

The connection between population distribution patterns and transport network development has also been described (Li et al., 2011a; Ural et al., 2011) and estimated (Li et al., 2011b; Linard et al., 2012). Bhaduri et al. (2007) simulated the population distributions at an hourly time interval by establishing the correlation between transport model and population data. Vörösmarty et al. (2000) emphasized the impacts of population change and the relationship between water supply and demand. Prior work occurred over two decades ago, when Falkenmark and Widstrand (1992) studied the problems between growing populations and limited local water resources.

The present work considers all the above factors in a single model but under different spatial analysis resolution scales. Rationalization is based on the fact that spatial patterns and the relationship between population distribution and its correlates can change with spatial resolution variation. We start with population surface modelling, followed by areal interpolation of the population distribution correlates. It has been suggested in the literature that efforts to assess the correlation between human population density and environmental factors on the basis of geographical determinism should be encouraged (Small and Cohen, 2004). Yet, since factors related to population density were not fully discussed with consideration to scale variation, global scale correlates may not be applicable at the urban level.

Regarding the outline of this paper, the introductory section provides an overview of previous studies of population distribution. The following Section 1 discusses two different categories of determination methods, the impacts of spatial scale on population distribution and the potential correlates of this distribution. Section 2 describes in detail the study area, the available data sets, and the methods used in the present work. Section 3 discusses our results and findings regarding population distribution at different spatial resolutions and their potential correlates in the study area. Our conclusions are summarized in Section 4, including a discussion of the limitations of the proposed analysis and future research directions.

2. Methods

2.1. Data Description

The population data used in this study were obtained from the 5th national population census of 2000 (National Bureau of Statistics of China, 2000), since 2010 data required by this study (such as road and rail networks) are not yet available. Population information was aggregated at administrative boundaries – census tracts. This means that the spatially continuous population distribution was discretized in terms of the administrative unit boundaries of the 3,404 counties included in the census, with 850 counties having no population records (Figure 1(a)). The county sizes in China range from 56 km^2 (in Shandong province) to 270,000 km^2 (in Tibet), with an average size of 2,895.22 km^2. The data cluster method used to determine the best population arrangement into different classes is Jenks natural breaks optimization. It aims to simultaneously reduce the variance within classes and maximize the variance between classes (Jenks, 1967; Campbell, 2001).

Spatial infrastructure data (such as main roads and railways) and the main rivers and lakes were derived from the China paper maps at the 1:4,000,000 scale (Wang and Feng, 2000). This small scaled dataset keeps the key information, and will avoid the "salt and pepper" effect during the analysis under coarser study resolution. Due to the large area to be covered by these data, only the main roads and railways, and the major rivers and lakes were considered in the present analysis. The resolution of the digital elevation model (DEM) used in this study is ~97 m (Figure 1(b)). Both elevation and relief amplitude information were used in this study. Relief amplitude information was derived from DEM using ArcGIS Desktop (ESRI, 2015). Its resolution is the same as in DEM (Figure 1(c)). The normalized difference vegetation index (NDVI) data were derived from satellite images with ~1 km resolution (Figure 1 (d)) (Wang and Feng, 2000). Consistent with the publica-tion date of the national population census, all datasets used in this work were published in 2000.

2.2. Rushton Circles

Using census population data aggregated within the administrative boundary ignores the local-spatial autocorrelation structure of population distribution and may exaggerate the boundary effects (a phenomenon known as the "modifiable areal unit problem" (Luo and Wei, 2006). Although the population within an administrative boundary is normally consid-

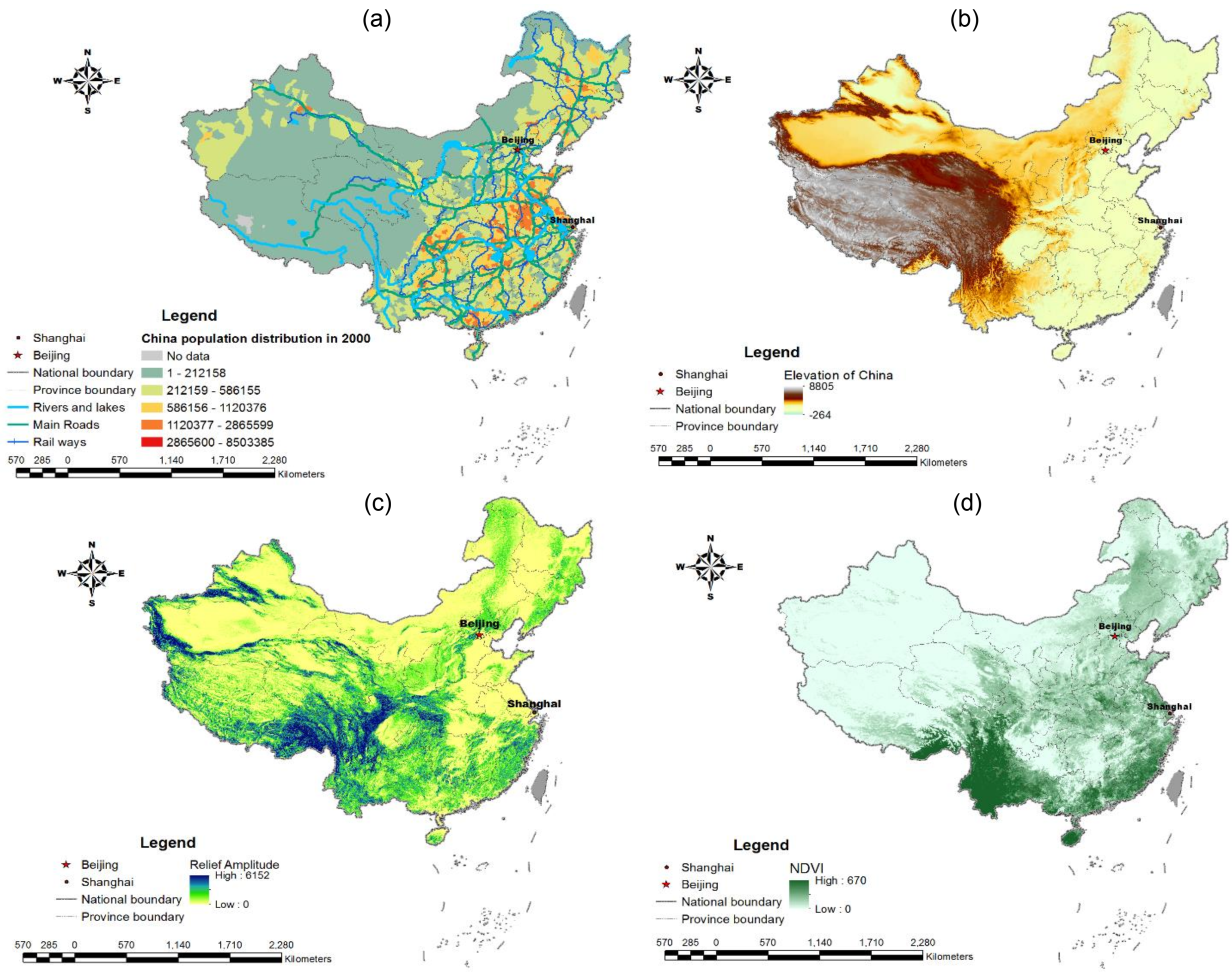

Figure 1. (a) China population distribution in the Census 2000 data at the county level (National Bureau of Statistics of China, 2000); (b) The geographic distribution of topographic elevation in China; (c) The geographic distribution the relief amplitude in China; (d) The geographic distribution of live green vegetation in China in 1999.

ered as evenly distributed, the actual population distribution is much more heterogeneous. This is especially valid for populations near the administrative boundaries and neighboring counties (Wu et al., 2005; Maantay et al., 2007).

To reduce boundary effects, Rushton circles (Rushton and Lolonis, 1996) were used in this study. First, 2,577 points were evenly sampled over the entire China, with south-north spacing of ~72 kilometers and east-west spacing of ~57 kilometers (Figure 2). The spacing distance of sampling points was determined by the range in empirical semivariogram model. Empirical semivariogram model represents the relationship between the averaged semivariogram values and the distance between two points. The distance where the model first levels out is the range. Point locations closer than the range are spatially autocorrelated, while point locations separated by distances larger than the range are not. This study shows that ranges along the south-north and east-west directions are ~72 kilometers and ~57 kilometers, respectively.

These sampling points are the center points of the Rushton circles, the radius of the Rushton circles being 50 kilometers,

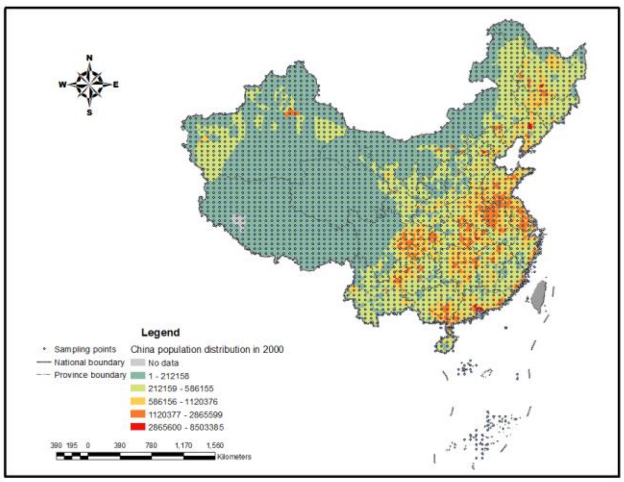

Figure 2. Sampling points distributed in China.

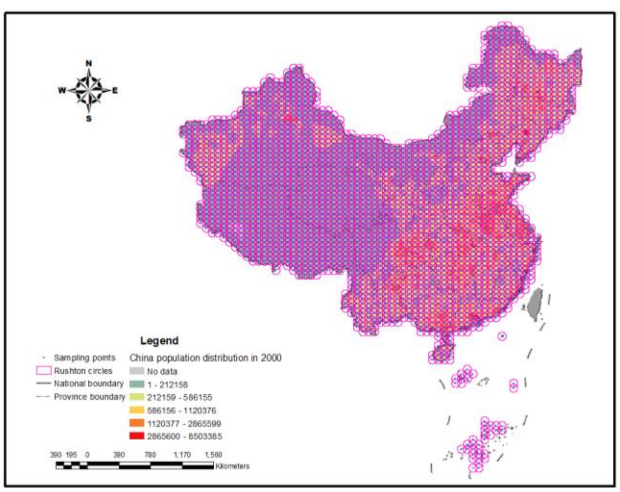

Figure 3. Rushton circles on each sampling points.

which means that the population totals of the sampling points were recalculated by averaging the population totals of the neighbor counties within a 50 kilometers distance (Figures 3 and 4). This radius length guaranteed that each circle includes enough neighboring information while maintaining relative independence. Finally, the involvement of Rushton circle can eliminate side effects due to inconsistent county sizes throughout the study area, when the "zoom lens" approach is applied.

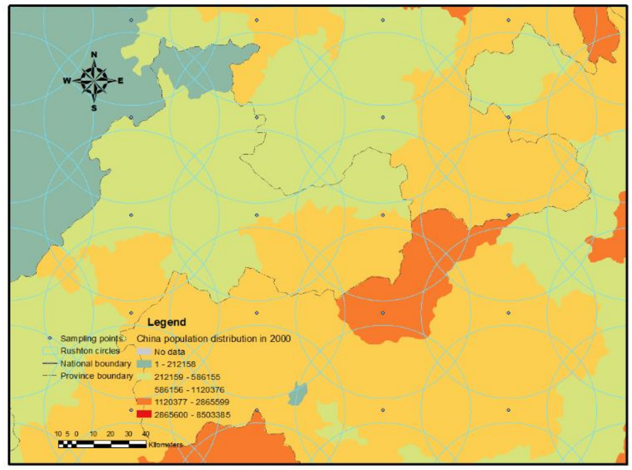

Figure 4. Detailed view of the Rushton circles.

2.3. The "Zoom Lens" Approach of Population Distribution

Being methodologically different from the traditional "multiple spatial scales" analysis, the proposed population distribution approach adopts a perspective involving the simultaneous operation of multiple-lens covering the entire study area. The "unnecessary" information (linked to certain scales; working like noises in an image) is ignored by applying these multiple-lens. As demonstrated by Jentsch et al. (2002), spatial patterns can only emerge and be identified at certain spatial scale. The patterns could be either diffuse (noise) at large spatial scale, or not differentiating at small spatial scale. For example, at certain large spatial scale, variability could be noise to forest dynamic analysis.

In this way, the approach focuses on the general trend at a scale specified according to the study objectives. The motivation underlying this approach is the concept of "wavelets" (Mallat, 1989). Although a number of image processing studies have been conducted using wavelets, there is no previous work on a combined wavelet-population distribution analysis. In this work, it is assumed that the population distribution is composed of two parts: a general distribution (V_j) component and a detailed local variation (W_j), component, so that:

$$V_{j+1} = V_j \oplus W_j \tag{1}$$

where $V_j \subset V_{j+1}$, and the index $j \in Z^+$, denotes the time of initial measure resolution. Z and Z^+ represent the sets of integers and positive integers, respectively. The larger the j, the more information is contained in the space V_j. Let (x, y) be the spatial coordinates of a sampling point. The $\{\varphi_{jk}(x)=2^{j/2}\varphi(2^jx-k)\}_{k\in Z}$ is the standard orthogonal basis in the general space $\{V_j\}_{j\in Z}$, and $\{\psi_{jk}(x)=2^{j/2}\psi(2^jx-k)\}_{k\in Z}$ is the standard orthogonal basis in the details space $\{W_j\}_{j\in Z}$. k defines the shift along different directions. The coefficients $D_j=\{d_{j,k}\}_{k\in Z}$ indicate the detailed local variations, while $C_j=\{c_{j,k}\}_{k\in Z}$ represent the general trend of population distribution with evaluation resolution times j. Giv-en that the population distribution is a function of (x, y), the corresponding basis $\psi_{jk}(x)$ constraints include the following: 1) the function must have a zero mean, and 2) it has to be localized in both two dimensional geographical space and frequency space (Farge, 1992).

In view of the above considerations, the population distribution takes the form:

$$f_j(x,y)=\sum_{k\in Z}c_{j-1,k}\varphi_{j-1,k}(x,y)+\sum_{k\in Z}d_{j-1,k}\psi_{j-1,k}(x,y) \qquad (2)$$

where $k\in Z^+$ indicates the length of displacement, $c_{j-1,k}=\{f_j, \varphi_{j-1,k}\}$ and $d_{j-1,k}=\{f_j, \psi_{j-1,k}\}$, and:

$$\begin{aligned}\varphi_{jk}(x,y)&=\varphi_{jk}(x)\ \varphi_{jk}(y)\\ \psi_{jk}(x,y)&=\psi_{jk}(x)\ \varphi_{jk}(y)\\ \psi_{jk}(x,y)&=\varphi_{jk}(x)\ \psi_{jk}(y)\\ \psi_{jk}(x,y)&=\psi_{jk}(x)\ \psi_{jk}(y)\end{aligned} \qquad (3)$$

These four functions define the filters used to detect the general structure and the detailed local variations of the population distribution along three directions in the study area: east-west, south-north, and northeast-southwest.

The variation of index j indicates the size change of the spatial filter used to detect the general trend and the detailed local variations of the population distribution. Relatively smaller spatial filters detect and remove high frequency spatial variations. The representation of the general population trend increases with filter size. Therefore, compared with the traditional global regression fitting methods, the above method reveals the spatial trend by locally removing unnecessary details. In this work, only the general structure analysis was considered, characterizing the overall trend of China population distribution at multiple geographical resolutions.

2.4. Correlates of Population Distribution

A number of studies have been conducted to evaluate the impacts of key factors on population distribution (Vörösmarty et al., 2000; Chown et al., 2003; Langford et al., 2008; Linard et al., 2012). Accordingly, the present work improves the analysis of the earlier studies by evaluating the impacts of these factors on population distribution at multiple spatial resolutions. As indicated by Small and Cohen (2004), within the category of natural environment factors, the physiographic factors have a more significant impact on population distribution than the climatic ones. Hence, our analysis focused only on physiographic factors, such as water resources, live green vegetation, topographic elevation and its variation.

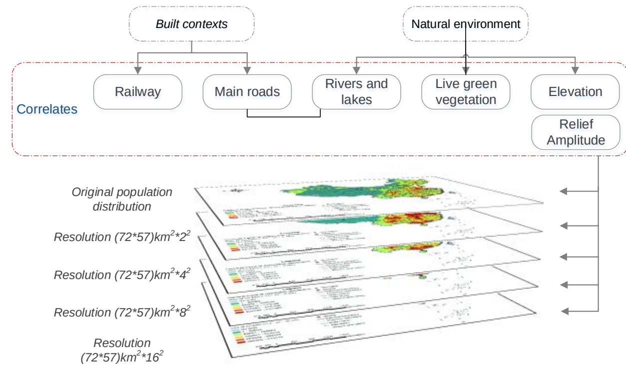

Figure 5. Conceptual framework of the regression analysis of population distribution correlates at different resolution scales.

Moreover, in this work linear regression was used to examine the impacts of potential correlates on the general trend of population distribution under different resolutions. Six correlates were considered, namely, shortest distance to railway ($RW_j(x, y)$), shortest distance to main roads ($MR_j(x, y)$), shortest distance to rivers and lakes ($RL_j(x, y)$), indicator of live green vegetation indicator ($NDVI_j(x, y)$), digital elevation model ($DEM_j(x, y)$) and Relief amplitude ($RA_j(x, y)$). These correlates were classified into two categories: built contexts and natural environment (Figure 5). The flowchart in Figure 6 displays the two main steps of population distribution correlate identification at different analysis resolutions.

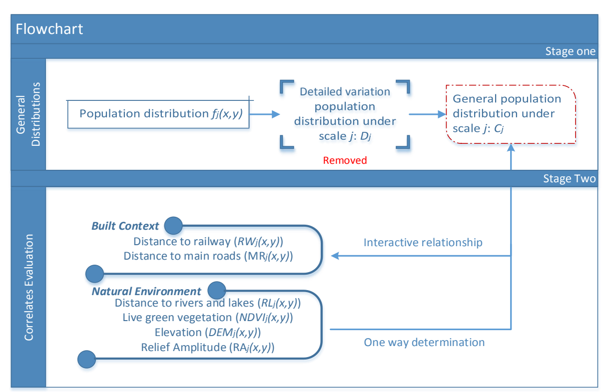

Figure 6. Flowchart of population distribution identification at different resolutions and correlate evaluation.

The global linear regression model is able to evaluate regional impacts based on the administrative boundary, whereas the Geographic Weighted Regression (GWR) (Fotheringham et al., 2003) model can also evaluate the localized impacts of different correlates. It determines the spatially varying independent variables, and local models instead of a global one are generated (Seppänen and Virrantaus, 2010; Lieske and Bender, 2011). GWR involves a more adaptive bandwidth or study extent to assess the spatial impacts on population distribution (Huang et al., 2010; Harris et al., 2011). Note that the correlate weights of the GWR model are functions of distance. Normally, the correlate weight (or influence) on a population distribution gets smaller as distance increases.

As noted earlier, in this work the GWR model was applied to evaluate the correlate impacts on both (a) the original population distribution of interest and (b) the distribution under the specified zoom lens ($72 \times 57 \times 16^2$ km^2). Linear regression at multiple resolutions enabled the investigation of population distribution correlates at different resolutions, whereas GWR advanced the analysis of the previous studies by refining the analysis extent. Therefore, this work studies the population distribution at multiple geographical scales, in terms of both resolution and extent.

3. Results and Discussions

3.1. China Population Distributions at Different "Focal Lengths"

Previous studies of population distribution were mainly conducted using the "prime lens" approach. Therefore, the results of estimation of one area always turn out to be almost identical and without successive spatial variations. In the present work, the successive variations of the population trends at

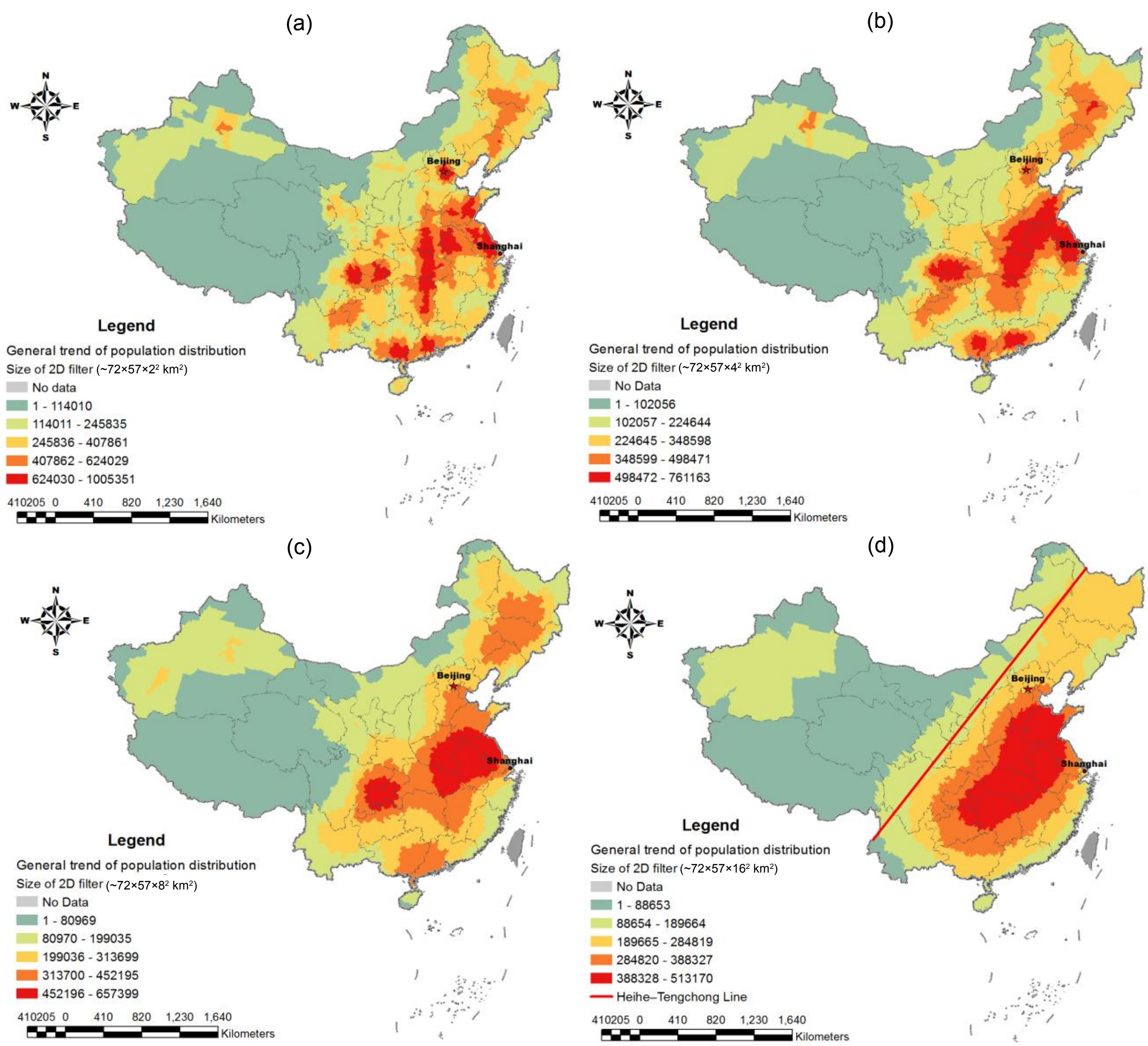

Figure 7. General trend of population distribution at resolution scales: (a) $72 \times 57 \times 2^2$, (b) $72 \times 57 \times 4^2$, (c) $72 \times 57 \times 8^2$, (d) $72 \times 57 \times 16^2$ (km^2).

Table 1. Impacts of Correlates on the Spatial Population Distribution at Different "Focal Lengths" Levels

	Overall test	Relative coefficients and significance						
Resolutions (km^2)	Adjusted R^2 and significance	Railway	Main roads	Rivers and lakes	Live green vegetation	Elevation	Relief amplitude	Constant
Original	.13*	.03*	.02	.08*	.03*	-.04*	.07*	.07*
$72×57×2^2$ (16,416)	.32*	.20*	–.06	.50*	.14*	–.31*	–.23*	.44*
$72×57×4^2$ (65,664)	.39*	.18*	–.06	.58*	.10*	–.42*	–.21*	.53*
$72×57×8^2$ (262,656)	.66*	.30*	–.22*	.72*	–.02	–.50*	–.10*	.58*
$72×57×16^2$ (1,050,624)	.68*	.43*	–.36*	.96*	–.13*	–.66*	.05*	.71*

*Significant at 0.05.

different spatial scales were identified.

By way of illustration, Figures 7(a) ~ (d) show maps of the China population distribution at different resolutions, ranging from ~$72 × 57 × 2^2$ to ~$72 × 57 × 16^2$ km. Note that the "focal lengths" become smaller as the resolution increases. Also, high frequency local variations were identified and removed (filtered) from the city to the national level, which allowed a realistic representation of the general population trend at these levels. Before the "zoom lens" was applied, population distribution comprises spatial variations at different spatial scales. By applying "zoom lens", distribution patterns at different "focal lengths" can be differentiated.

Figures 7(a) and (b) illustrate the regional hotspots of population (North East of China, Beijing-Tianjin, Middle East of China, Sichuan province, Guangdong Province), whereas the distribution of hotspots in Figure 7(b) is more continuous. In contrast, Figure 7(d) presents the national population distribution trend, and the distribution in Figure 7(c) works as the intermediate state between the discovered phenomenon in Figures 7(b) and (d). Figure 7(d) shows that the high-density population in China is roughly aligned with an oblique trend line along the Heihe–Tengchong Line, which is also known as "Hu line". This "geo-demographic demarcation line" was imagined by Chinese population geographer Hu Huanyong in 1935 (Yue et al., 2003). The southeastern side of the Heihe–Tengchong Line occupied 36% of the total area of China, however, it had 96% of the total population of China in 1935 (Yue et al., 2003). The result from this study directly and quantitatively supports the rationale of Heihe–Tengchong Line and provides a mathematical foundation for it. This trend shown in Figure 7(d) is also roughly aligned with the general distribution of China elevation. People concentrate in low areas, whereas smaller populations live at high altitudes. Further quantitative analysis will be presented in a subsequent section.

In light of the above considerations, the geographical "zoom lens" approach used in the present work constitutes progress over earlier studies based on the composite study of global population trend and spatial variations. They are controlled by the time of initial measure resolution j and the length of the displacement k, respectively. The j controls the variation of spatial analysis resolution, while the k provides sufficient support for the identification of local variations.

3.2. Relationship between Population Distributions and Their Correlates at Different "Focal Lengths"

It is generally recognized that in order to understand the spatial distribution of a population, it is necessary to quantify the relationship between it and its potential correlates (Small and Cohen, 2004). A gradation in population trend was identified from the eastern to the western and middle region, which is determined mainly by transportation infrastructure and environmental factors, such as net primary productivity and topographic elevation (Yue et al., 2005). However, this conclusion was drawn without considering the varying impacts of these factors at different study resolutions.

Previous studies have argued that there are a large number of urban developments in water areas, such as rivers, lakes, and sea coastal zones, and at lower elevations. This is due to coastal and marine food sources (Sachs, 1997) as well as to the economic and strategic advantages offered by coastlines (Turner, 1990). The present work is in agreement with the findings of previous studies and traditional understanding concerning the water's impact on population distribution. Furthermore, such an impact becomes significantly high at the national scale (Table 1), being much stronger than the impact of other potential correlates. On the other hand, water resources accessibility is not a serious problem, especially at the national scale, assuming that the current network of rivers and lakes in China does not dry up.

Pautasso (2007) has pointed out that correlation between population density and live green vegetation is inverse at very fine resolution (large spatial scale), and becomes positive as the spatial scale decreases (> 10,000 km^2). The present work partially confirms Pautasso's findings at small resolutions from a different viewpoint, and it also improves these findings by accounting for variations in live green vegetation impacts at multiple resolutions (> 10,000 km^2). The significant positive correlation drops to a non-significant one as the analysis scale reaches to ~260,000 km^2; and it changes to a negative correlation at fine resolution. Yet, compared with other correlates, the influence of live green vegetation is much less significant.

Furthermore, this work confirmed Cohen and Small (1998) conclusions, and it is in agreement with common sense that smaller populations live at high altitudes. Another finding of this work was that the negative correlation does not change significantly between the spatial resolution scales of ~65,000 and ~260,000 (km^2). Among the correlates evaluated in this work, topographic elevation was the second most important factor that had an impact on population distribution. Yu et al. (2013) have pointed out the negative relationship between population density and relief amplitude. The present work extends this observation by considering the study resolution. Specifically, we found that the observed negative relationship is more significant at a larger study resolution. The China population distribution trend is not significantly aligned with relief amplitude distribution. This happens because in China the relief amplitude distribution is not like the distribution of the topographic elevation distribution that has clear-boundaries of 1st-, 2nd- and 3rd-step geomorphic variations (according to population distribution nation-wide).

Compared to the natural environment, the relationship between population distribution and built context is more complicated and interactive. Built context has an impact on population distribution, whereas, in turn, population distribution influences the built context. In this work, it was found that the positive correlation between railway and population distribution increases, and becomes significant as the study resolution decrease. However, the relationship between main roads and population distribution only becomes significant negative when the study resolution is smaller than ~260,000 km^2. A possible reason for this phenomenon is the unbalanced accessi-bility to main roads at national level, and the accessibility at local or regional level is still acceptable. The "zoom lens" at smallest focal length (small scale view) identifies the general trend of population distribution, and the global regression determines the mismatch between population distribution and main roads construction at national level. Relationship may vary as study scale changes.

The consideration of the built context in quantitative analysis should also account for transportation accessibility at different resolutions of geographical scale. Compared to main roads development, it was found that railway construction took into consideration transportation accessibility to a satisfactory degree, especially at the national scale (with positive correlation coefficient 0.43). In contrast, main road accessibility is either non-significant or negative at different study resolutions. This indicates an uneven national distribution of main roads accessibility in 2000. This work provided useful indicators for future improvements of main road networks. For example, the higher-level road network should be aligned with the general population distribution shown in Figure 7(d), whereas the lower-level road network could be built in the hot spots neighborhoods shown in Figure 7(c). This has been supported by the fast road network development in these areas between 2000 and 2010 (Li and Fang, 2014). Accordingly, this work offered a new way to maximize the marginal benefit of built context improvement.

Generally, the impacts (positive or negative) of potential correlates became larger and more significant as the study resolution scale decreases. Also, the confidence of the overall analysis improved when the local variations were removed and only the general population trend was observed. Since original population distribution comprises many localized variations at different spatial scales, it is not easy to use a global regression to completely explain it and therefore a low R^2 value (0.13) was obtained. However, as unnecessary information was removed (filtered) at lower "focal lengths", distribution pattern can be easily and better explained (with R^2 value at about 0.67), even by a simple global regression. To better understand the relationship between original population distribution and its potential correlates, a further localized analysis is needed.

3.3. Localized Analysis of Population Distribution Correlates

The global investigation of correlates impact presented in the previous section may ignore the spatial heterogeneity of the relationships. In this section, therefore, population distribution correlates at different study resolutions were further investigated by means of localized analysis (Figures 8 and 9). Specifically, the analysis showed that these correlates can have a considerable impact on population distribution only in certain local areas. So, a correlate that has a non-significant impact at the global scale can have a significant impact at local scales. The localized analysis further improved the "zoom lens" approach by allowing for spatial variations. Furthermore, the localized relationship between correlates and population distribution changed after the "zoom lens" approach was implemented. A comparison between Figures 8 and 9 shows the variation.

Based on Table 1, positive correlations between transportation accessibility and the original population distribution were observed, whereas localized analysis further improved our findings by localizing these positive correlations. Previous studies have concluded that people are more likely to be concentrated to areas with better accessibility to transportation services (Liao et al., 2010). Strong positive relationships between railways and population distribution were identified only in the eastern part of China (Figure 8(a)), whereas the main roads networks have a significant positive influence in the middle-eastern and southern regions of China (Figure 8(b)).

A positive relationship between live green vegetation and population distribution was also revealed along the Jing-Hang canal and the southern part of China (Figure 8(c)), where sufficient river systems and lakes exist now and in the past. On the other hand, people in this area are not expected to have trouble accessing water, provided the rivers and lakes of the area do not dry up. Significant correlations, positive (east of China) and negative (part of Yunnan and Tibet), were identified between population distribution and live green vegetation (Figure 8(d)). The negative impact of elevation was spatially refined in the middle

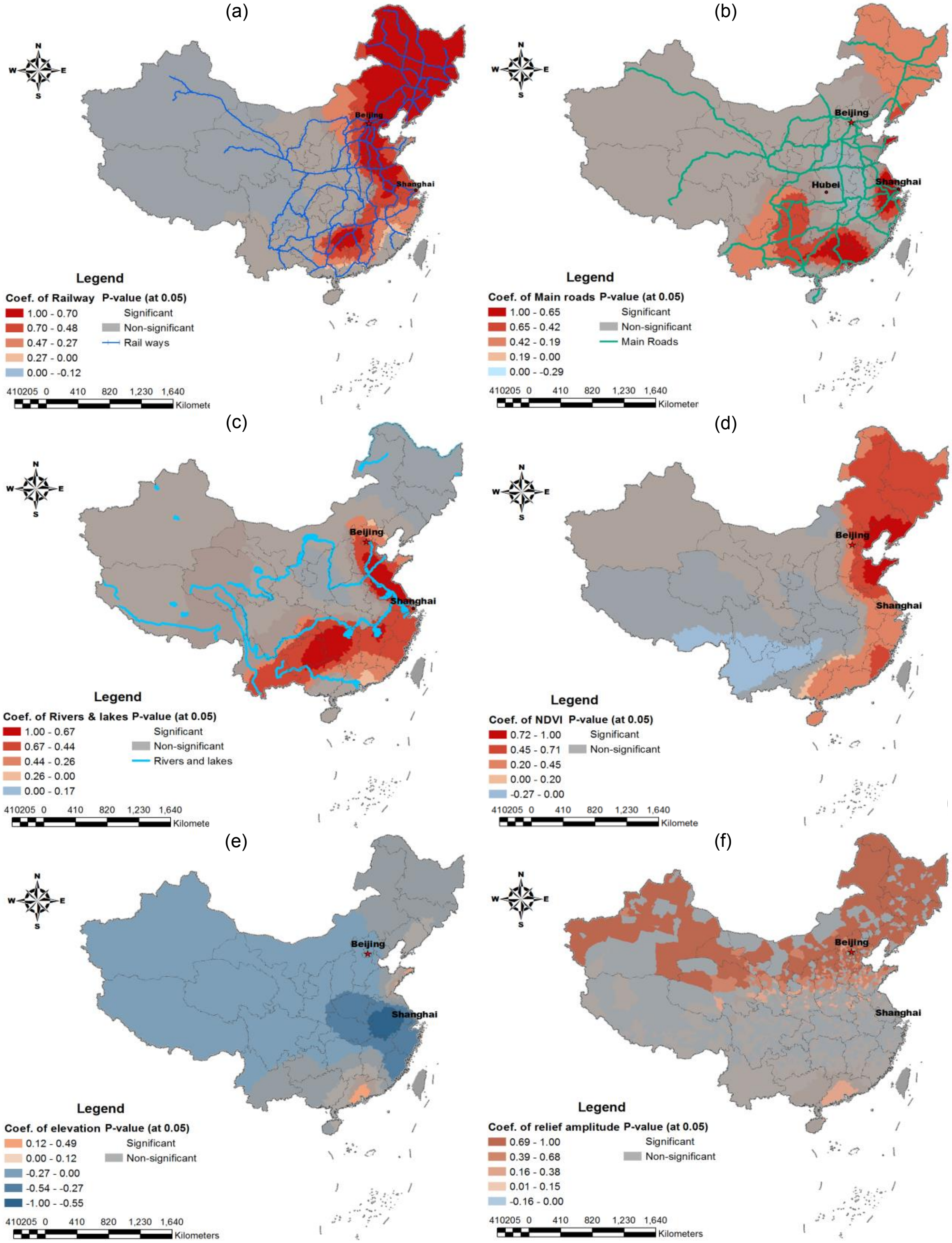

Figure 8. Localized analysis results of original population distribution correlates in China: (a) Railway, (b) Main roads, (c) Rivers and lakes, (d) Live green vegetation, (e) Elevation, (f) Relief amplitude.

of China, especially in areas around Shanghai city where low elevation areas with high density of population exist (Figure 8(e)). A significant high-high and low-low relationships between population density and relief amplitude were determined in the north of China (Figure 8(f)). This indicates that the population distribution in most areas of northern China exhibits a higher correlation with local topographic variability than in other regions of the country.

After the “zoom lens” approach (global view) was applied in the study of population distribution in this work, more insightful relationships were revealed by localized analysis. This implied that the spatial scale used in the study of correlations could change in terms of both resolution and extent. Furthermore, more complex and diversified relationships were identified for certain correlates, like transportation and water accessibility. For example, the center of Hubei province (which is the traditional traffic gateway in central China) displays a slight disadvantage concerning railway accessibility (Figure 9(a)).

A significant negative correlation between population distribution and main road accessibility was found after “zoom lens” was applied. Localized analysis furthered this finding by identifying areas lacking main road accessibility (coastal areas between Beijing and Shanghai) and, at the same time, pointing-out the nation-wide unevenness concerning main road accessibility. A plausible suggestion is that, a maximum marginal benefit could be achieved by building main roads along these areas (Figure 9(b)).

The proposed approach (based on the combination of “zoom lens” and localized analysis) revealed that the population in the He’nan and south of Shanxi provinces were lacking access to water (Figure 9(c)). This is caused mainly by the high-density population and the relatively smaller river network and number of lakes in the area. At the nation-wide scale, areas exhibiting strong correlations with topographic variability become smaller and less significant (Figure 9(d)). In contrast, a negative relationship was identified in eastern China, where large populations live in conditions of low relief amplitude. Overall, the implementation of the “zoom lens” approach leads to a definite change of emphasis concerning the relationship between population distribution and its correlates by focusing more on the nation-wide study of this relationship (larger significant areas were presented for most correlates nation-wide).

4. Conclusions

Methodologically, this work presented a new way of viewing geographical distributions of phenomena occurring on the earth’s surface. It proposed a decomposition of a 2D distributed population into component trends at distinct geographical scales, and, at the same time, it retained localities at different decomposition levels. The present work also improved the results of previous studies concerning the evaluation of correlates using a refined spatial study extent and an arbitrary spatial study resolution.

In this work, four different general population trends were identified at different spatial resolution scales. Underlying the proposed method is the same mechanism as that used by “zoom lens”. The method removed unnecessary details at multiple spatial resolutions with different “focal length optics”. Based on small resolution scale analysis, an oblique trend line of population distribution was determined roughly along the Heihe–Tengchong Line, which directly and quantitatively support the rationale of Heihe–Tengchong Line and provides a mathematical foundation for it.

The correlates were categorized into two different groups: built context and natural environment. From a global analysis viewpoint, railway construction is considered well-developed at any spatial resolution scale, in terms of population accessibility. It is noteworthy that the road construction problem was concealed in the traditional viewpoint, where a positive correlation between population distribution and main road accessibility was observed. However, the “zoom lens” approach used in this work identified a significant negative relationship at the national scale. Therefore, the results of this study demonstrated that correlation or any other relationship may vary at different spatial scales of study. The localized analysis with different “focal length optics” further proves this statement.

Furthermore, localized analysis showed that people in most areas of eastern China and in the middle-eastern and the southern China had good accessibility to railways and main roads, respectively. However, when the combined “zoom lens-localized analysis” was used, the unevenness of main roads and railways accessibility was identified nation-wide. Also, it was suggested that the maximum marginal benefit could be achieved in the future if transportation infrastructure was developed in the deterioration areas.

The accurate assessment of the links between population distribution and natural environment, such as population along the rivers, lakes and live green vegetation, would have a significant impact on coastal planning and policy-making in terms of coastal hazard, water pollution and disease control. From a global analysis perspective, it was found that the two most important factors that influence population distribution were topographic elevation and water (river, lake) accessibility, characterized by negative and positive correlations, respectively. These correlations became more significant as the study scale decrease.

From a localized analysis perspective, water was shown to be a significant population distribution factor in eastern and southern China. Although currently people in China have good water accessibility, assuming that the existing lakes and rivers do not dry up, the combined spatial “zoom lens-localized analysis” identified certain water accessibility problems for the populations in the He’nan and south of Shanxi regions, to which more attention should be paid in the future.

The negative correlation of topographic elevation and population is significant in most areas of China. This will not vary by changing the study resolution scale. However, the variation of study resolution scale does have an impact on the relationship between topographic variation and population distribution. From global view, a negative impact of topographic varia-

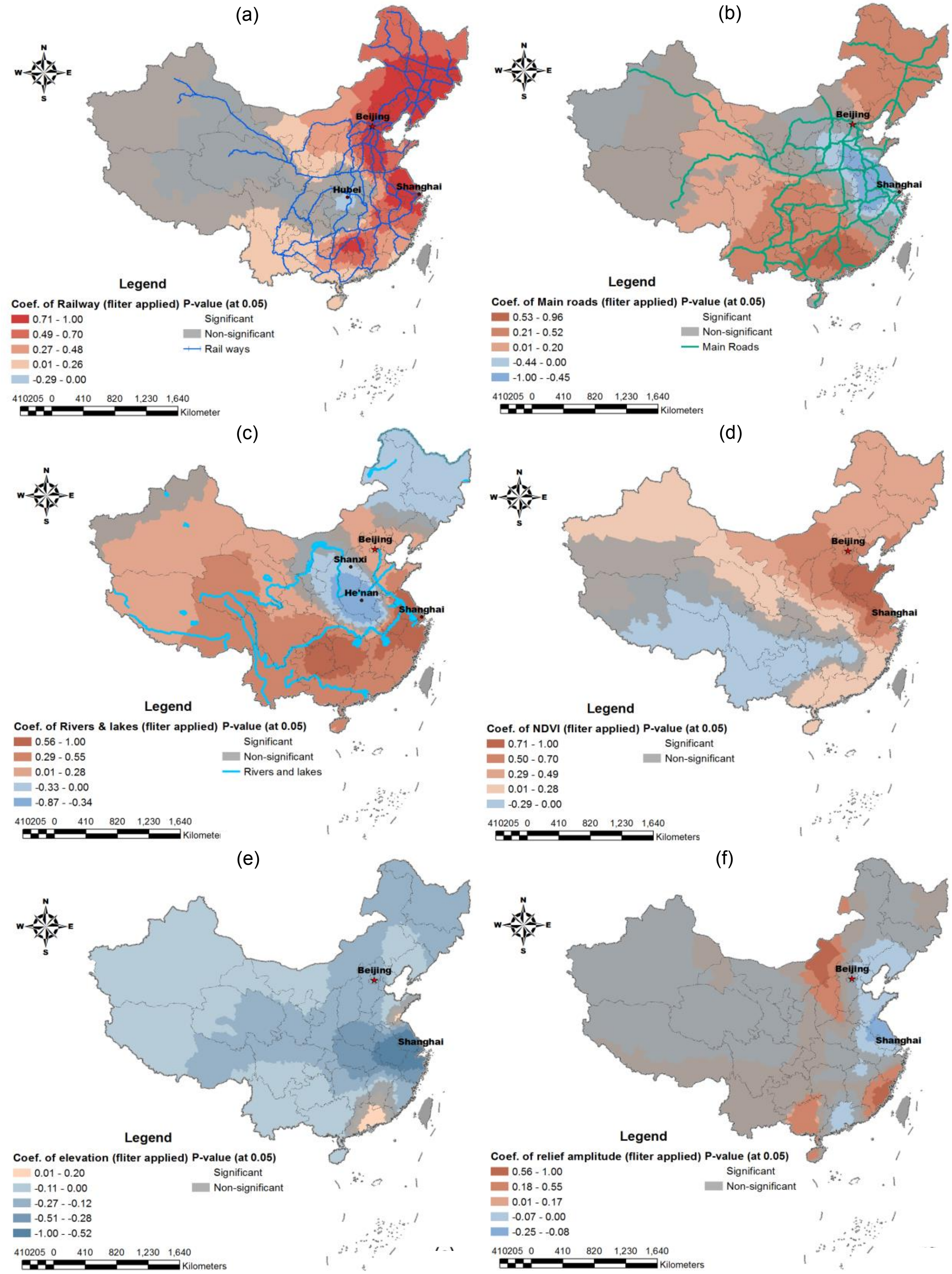

Figure 9. Localized analysis results of the "zoom lens (national scale)" population distribution correlates in China: (a) Railway, (b) Main roads, (c) Rivers and lakes, (d) Live green vegetation, (e) Elevation, (f) Relief amplitude.

tion can be identified in the coastal areas of eastern China characterized by high population density and small topographic variation.

The live green vegetation impact on population distribution is complex. It changes from slight positive to slight negative as the study resolution downscales from city to national level. On the basis of localized analysis, it was found that live green vegetation was a positive factor in eastern China, but it showed a negative impact around the Yunnan province.

Generally, the "zoom lens" approach implemented in this work illustrated the significance of the impact of certain correlates on population distribution based on both the global and local perspectives. Some areas that had been previously characterized as satisfactory with respect to transportation infrastructure and natural resource accessibilities were found to be defective under the light of the proposed combined "zoom lens-localized analysis".

Lastly, using the proposed method, it would be interesting to investigate the Chinese population distribution and its correlates when the more recent dataset become available. Then, more informative results could be derived by comparing the temporal variation at different resolutions, and obtain some valuable insight concerning population migration trend. In the same context, this work's conclusions could be further evaluated by investigating the world population distribution when the relevant data become available.

Acknowledgments. NSFC (4127404) and MOST (2012CB955503). The authors thank Zhoupeng Ren for his interest and for providing a number of suggestions in this research.

References

Badescu, V. (2008). Risks for poisoning of coastal and inland population due to asteroid impacts in Southern regions of Black Sea. *Stochastic Environ. Res. Risk Assess.,* 22, 461-476. https://doi.org/10.1007/s00477-007-0146-x

Balk, D. and Yetman, G. (2004). The global distribution of population: evaluating the gains in resolution refinement. *New York: Center for International Earth Science Information Network (CIESIN),* Columbia University.

Bhaduri, B., Bright, E., Coleman, P., and Urban, M.L. (2007). Land-Scan USA: A high-resolution geospatial and temporal modeling approach for population distribution and dynamics. *Geo-Journal,* 69, 103-117. https://doi.org/10.1007/s10708-007-9105-9

Bracken, I. (1993). An extensive surface model database for population-related information: concept and application. *Environ. Plann. B*, 20, 13-27. https://doi.org/10.1068/b200013

Campbell, J. (2001). *Map Use & Analysis*, McGraw-Hill New York.

Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S., and van Jaarsveld, A.S. (2003). Energy, species richness, and human population size: conservation implications at a national scale. *Ecol. Appl.,* 13, 1233-1241. https://doi.org/10.1890/02-5105

Cincotta, R.P., Wisnewski, J., and Engelman, R. (2000). Human population in the biodiversity hotspots. *Nature,* 404, 990-992. https://doi.org/10.1038/35010105

Cohen, J.E. and Small, C. (1998). Hypsographic demography: The distribution of human population by altitude. *Proc. Natl. Acad. Sci.,* 95, 14009-14014. https://doi.org/10.1073/pnas.95.24.14009

ESRI (2015). *ArcGIS Desktop, 10.2 ed.* Environmental Systems Research Institute, Redlands, CA.

Falkenmark, M. and Widstrand, C. (1992). Population and water resources: a delicate balance. *Popul. Bull.,* 47, 1-36.

Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.,* 24, 395-458. https://doi.org/10.1146/annurev.fl.24.010192.002143

Feng, Z., Tang, Y., Yang, Y., and Zhang, D. (2008). Relief degree of land surface and its influence on population distribution in China. *J. Geogr. Sci.,* 18,237-246.https://doi.org/10.1007/s11442-008-0237-8

Fotheringham, A.S., Brunsdon, C., and Charlton, M. (2003). *Geographically Weighted Regression,* John Wiley & Sons, Limited.

Gaughan, A.E., Stevens, F.R., Linard, C., Patel, N.N., and Tatem, A.J. (2015). Exploring nationally and regionally defined models for large area population mapping. *Int. J. Digit. Earth,* 8, 989-1006. https://doi.org/10.1080/17538947.2014.965761

Gregory, I.N. (2000). An evaluation of the accuracy of the areal interpolation of data for the analysis of long-term change in England and Wales, *Proceedings of the 5th International Conference on GeoComputation University of Greenwich*, United Kingdom, pp. 23-25.

Harris, P., Brunsdon, C., and Fotheringham, A.S. (2011). Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. Stochastic *Environ. Res. Risk Assess.,* 25, 123-138. https://doi.org/10.1007/s00477-010-0444-6

Huang, B., Wu, B., and Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.,* 24, 383-401. https://doi.org/10.1080/13658810802672469

Huang, B., Xie, C., Tay, R., and Wu, B. (2009). Land-use-change modeling using unbalanced support-vector machines. *Environ. Plann. B,* 36, 398-416. https://doi.org/10.1068/b33047

Ingram, G.K. (1998). Patterns of metropolitan development: what have we learned? *Urban stud.,* 35, 1019-1035. https://doi.org/10.1080/0042098984466

Jenks, G.F. (1967). The data model concept in statistical mapping. *Int. Yearbook Cartogr.,* 7, 186-190.

Jentsch, A., Beierkuhnlein, C., and White, P.S. (2002). Scale, the dynamic stability of forest ecosystems, and the persistence of biodiversity. *Silva Fenn.,* 36(1), 393-400. https://doi.org/10.14214/sf.570

Langford, M., Higgs, G., Radcliffe, J., and White, S. (2008). Urban population distribution models and service accessibility estimation. *Comput., Environ. Urban Syst.,* 32, 66-80. https://doi.org/10.1016/j.compenvurbsys.2007.06.001

Li, G. and Fang, C. (2014). Analyzing the multi-mechanism of regional inequality in China. *Ann. Reg. Sci.,* 52(1), 155-182. https://doi.org/10.1007/s00168-013-0580-2

Li, G. and Weng, Q. (2005). Using Landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. *Photogramm. Eng. Remote Sensing,* 71, 947-958. https://doi.org/10.14358/PERS.71.8.947

Li, X., Chen, Y., Liu, X., Li, D., and He, J. (2011a). Concepts, methodologies, and tools of an integrated geographical simulation and optimization system. *Int. J. Geogr. Inf. Sci.,* 25, 633-655. https://doi.org/10.1080/13658816.2010.496370

Li, X., Shi, X., He, J., and Liu, X. (2011b). Coupling simulation and optimization to solve planning problems in a fast-developing area. *Ann. Assoc. Am. Geogr.,* 101, 1032-1048. https://doi.org/10.1080/00045608.2011.577366

Liao, Y., Wang, J., Meng, B., and Li, X. (2010). Integration of GP and GA for mapping population distribution. *Int. J. Geogr. Inf. Sci.,* 24, 47-67. https://doi.org/10.1080/13658810802186874

Lieske, D. and Bender, D. (2011). A robust test of spatial predictive models: geographic cross-validation. *J. Environ. Inf.,* 17. https://doi.org/10.3808/jei.201100191

Linard, C., Gilbert, M., Snow, R.W., Noor, A.M., and Tatem, A.J. (2012). Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS ONE*, 7, e31743. https://doi.org/10.1371/journal.pone.0031743

Luo, J. and Wei, Y.D. (2006). Population distribution and spatial structure in transitional Chinese cities: A study of Nanjing. *Eurasian Geogr. Econ.,* 47, 585-603. https://doi.org/10.2747/1538-7216.47.5.585

Lutz, W. and Samir, K. C. (2010). Dimensions of global population projections: what do we know about future population trends and structures? *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.,* 365(1554), 2779-2791. https://doi.org/10.1098/rstb.2010.0133

Maantay, J.A., Maroko, A.R., and Herrmann, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartogr. Geogr. Inf. Sci.,* 34, 77-102. https://doi.org/10.1559/152304007781002190

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.,* 11(7), 674-693. https://doi.org/10.1109/34.192463

Martin, D. and Higgs, G. (1996). The visualization of socio-economic GIS data using virtual reality tools. *Trans. GIS,* 1, 255-266. https://doi.org/10.1111/j.1467-9671.1996.tb00050.x

McKee, J.J., Rose, A.N., Bright, E.A., Huynh, T., and Bhaduri, B.L. (2015). Locally adaptive, spatially explicit projection of US population for 2030 and 2050. *Proc. Natl. Acad. Sci.,* 112, 1344-1349. https://doi.org/10.1073/pnas.1405713112

Moon, Z.K. and Farmer, F.L. (2001). Population density surface: A new approach to an old problem. *Soc. Nat. Resour.,* 14, 39-51. https://doi.org/10.1080/089419201300199545

National Bureau of Statistics of China (2000). *National Population Census*, Department of Population and Employment Statistics, Beijing.

Pautasso, M. (2007). Scale dependence of the correlation between human population presence and vertebrate and plant species richness. *Ecol. Lett.,* 10, 16-24. https://doi.org/10.1111/j.1461-0248.2006.00993.x

Poulsen, E. and Kennedy, L.W. (2004). Using dasymetric mapping for spatially aggregated crime data. *J. Quant. Criminol.,* 20, 243-262. https://doi.org/10.1023/B:JOQC.0000037733.74321.14

Rogerson, P.A. and Kim, D. (2005). Population distribution and redistribution of the baby-boom cohort in the United States: Recent trends and implications. *Proc. Natl. Acad. Sci.,* 102, 15319-15324. https://doi.org/10.1073/pnas.0507318102

Rushton, G. and Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Stat. Med.,* 15, 717-726. https://doi.org/10.1002/(SICI)1097-0258(19960415)15:7/9<717::AID-SIM243>3.0.CO;2-0

Sachs, J. (1997). The limits of convergence. Nature nurture and growth. *Economist*, 19-22.

Seppänen, H. and Virrantaus, K. (2010). The role of GI-supported methods in crisis management. *Int. J. Digit. Earth,* 3, 340-354. https://doi.org/10.1080/17538947.2010.491560

Sleeter, R. (2004). Dasymetric mapping techniques for the San Francisco Bay region, California, *Urban and Regional Information Systems Association Annual Conference Proceedings*, pp. 7-10.

Small, C. and Cohen, J.E. (2004). Continental physiography, climate, and the global distribution of human population. *Curr. Anthropol.,* 45, 269-277. https://doi.org/10.1086/382255

Small, C. and Naumann, T. (2001). The global distribution of human population and recent volcanism. *Global Environ. Change B: Environ. Hazards,* 3, 93-109. https://doi.org/10.1016/S1464-2867(02)00002-5

Sutton, P., Roberts, D., Elvidge, C., and Baugh, K. (2001). Census from Heaven: an estimate of the global human population using night-time satellite imagery. *Int. J. Remote Sens.,* 22, 3061-3076. https:// doi.org/10.1080/01431160010007015

Tong, S., Sun, Y., and Yang, Y. (2012). Generating a Future Land Use Change Scenario with a Modified Population-Coupled Markov Cellular Automata Model. *J. Environ. Inf.,* 19.

Turner, B.L. (1990). *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, CUP Archive.

Ural, S., Hussain, E. and Shan, J. (2011). Building population mapping with aerial imagery and GIS data. *Int. J. Appl. Earth Obs. Geoinf.,* 13, 841-852. https://doi.org/10.1016/j.jag.2011.06.004

Vörösmarty, C.J., Green, P., Salisbury, J., and Lammers, R.B. (2000). Global water resources: vulnerability from climate change and population growth. *Science,* 289, 284-288. https://doi.org/10.1126/science.289.5477.284

Wang, J. and Feng, M. (2000). China Spatial Infrastructure Data, *The Institute of Geographic Sciences and Natural Resources Research (IGSNRR),* C. (Ed.), Beijing.

Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., and Buckee, C.O. (2012). Quantifying the Impact of Human Mobility on Malaria. *Science,* 338, 267-270. https://doi.org/10.1126/science.1223467

Wu, S.S., Qiu, X., and Wang, L. (2005). Population estimation methods in GIS and remote sensing: a review. *GISci. Remote Sens.,* 42, 80-96. https://doi.org/10.2747/1548-1603.42.1.80

Yin, Z.-Y., Stewart, D.J., Bullard, S., and MacLachlan, J.T. (2005). Changes in urban built-up surface and population distribution patterns during 1986-1999: A case study of Cairo, Egypt. *Comput., Environ. Urban Syst.,* 29, 595-616. https://doi.org/10.1016/j.compenvurbsys.2005.01.008

Yu, H., Deng, W., and Liu, S. (2013). The impact of relief amplitude on population and economic development in three Gorges Reservoir. *Res. Environ. Yangtze Basin,* 22, 686-690.

Yue, T.X., Wang, Y.A., Chen, S.P., Liu, J.Y., Qiu, D.S., Deng, X.Z., and Tian, Y.Z. (2003). Numerical simulation of population dis-tribution in China. *Popul. Environ.,* 25(2), 141-163. https://doi. org/10.1023/B:POEN.0000015562.88915.01

Yue, T.X., Wang, Y.A., Liu, J.Y., Chen, S.P., Qiu, D.S., Deng, X.Z., Liu, M.L., Tian, Y.Z., and Su, B.P. (2005). Surface modelling of human population distribution in China. *Ecol. Model.,* 181, 461-478. https://doi.org/10.1016/j.ecolmodel.2004.06.042