GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: https://www.tandfonline.com/loi/tgrs20

Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data

Hasi Bagan & Yoshiki Yamagata

To cite this article: Hasi Bagan & Yoshiki Yamagata (2015) Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data, GIScience & Remote Sensing, 52:6, 765-780, DOI: 10.1080/15481603.2015.1072400

To link to this article: https://doi.org/10.1080/15481603.2015.1072400

GIScience & Remote Sensing, 2015
Vol. 52, No. 6, 765–780, http://dx.doi.org/10.1080/15481603.2015.1072400

Taylor & Francis
Taylor & Francis Group

Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data

Hasi Bagan* and Yoshiki Yamagata

Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

(*Received 17 April 2015; accepted 10 July 2015*)

We investigated the spatiotemporal dynamics of urban expansion in Japan from 1990 to 2006 by using gridded land-use data, population census data, and satellite images of nighttime lights. First, we mapped Defense Meteorological Satellite Program (DMSP) nighttime lights and land-use data onto the 1 km^2 grid cell system of Japan to determine the proportional areas of DMSP and urban land use within each grid cell. Then, we investigated the relationships among population density, DMSP, and urban area. The urban/built-up area was strongly positively correlated with population density, and rapid expansion of the urban/built-up area around megacities was associated with population increases. In contrast, population density dropped steeply in rural areas and in small towns. Statistical analysis showed that correlation coefficients between population density and DMSP increased as the DMSP nighttime lights brightness value increased. We next estimated population density in the Hokkaido region using an ordinary least squares (OLS) regression model. Numerical evaluation of the results showed that the combination of land-use data and DMSP could be used to predict the population density. Finally, we compared OLS and geographically weighted regression (GWR) model for Sapporo city, Hokkaido. Compared with the OLS, the GWR can improve predictions of population density.

Keywords: urban growth; population density; correlation analysis; OLS; GWR

1. Introduction

Urbanization is commonly represented by population density and urban growth data (Seto, Guneralp, and Hutyra 2012). The global urbanization level increased steadily throughout the 20th century (Brockerhoff 2000). In this century, the world population is projected to grow substantially from the current 7.2 billion people to 9.6 billion in 2050 and 10.9 billion in 2100, and this increase is unlikely to be halted (Gerland et al. 2014). As urban populations grow, the area of urban land cover is expected to increase to about three times its current area by 2030 (Seto et al. 2011).

Information on how cities are changing becomes increasingly important as the population becomes more urbanized, and reliable information on the spatial distribution of the human population is key to a better understanding of urban growth (Lu, Im, and Quackenbush 2011; Lu and Guldmann 2012).

Images of nighttime lights acquired by the NOAA Air Force Defense Meteorological Satellite Program (DMSP) Operational Linescan System sensor have been linked to the spatial distribution of population (Elvidge et al. 2007), and DMSP images of nighttime

*Corresponding author. Email: hasi.bagan@nies.go.jp

lights on the earth's surface obtained over a period of decades provide a metric of urban extent that has been correlated with other metrics of urban activity such as dense population clusters (Sutton et al. 1997; Pozzi, Small, and Yetman 2003) and monitoring socioeconomic activities (Doll, Muller, and Morley 2006; Witmer and O'Loughlin 2011; Butt 2012; Propastin and Kappas 2012; Henderson, Storeygard, and Weil 2012; Forbes 2013; Shi et al. 2015; Su et al. 2015). Therefore, it should be possible to use the relationships among population density, urban land use, and DMSP images of nighttime lights for mapping urban growth and developing effective urban planning strategies.

Japan is one of the most urbanized countries in the world; the rapid growth of the Japanese economy has concentrated its population, industry, and other economic activities in metropolitan regions to an extreme degree (Bagan and Yamagata 2012). Thus, urban growth is one of the most important land-cover/land-use change events in recent Japanese history.

Spatiotemporal analyses of urban growth based on grids composed of 1 km^2 cells have demonstrated that gridded data are useful for obtaining spatial and temporal information about areas that are smaller than municipal scale and are uniform in size (Bagan and Yamagata 2012; Qian et al. 2014; Maimaitijiang et al. 2015). Moreover, detailed gridded population census (grid cell size, 1 km^2) and land-use (0.01 km^2) data are available for Japan. Analysis based on grid cells has many advantages: grid cells all have the same size allowing for easy comparison; grids integrate easily with other scientific data; grids are stable over time and thus facilitate the modeling and analysis of very large multivariate spatial data sets.

Multiple regression analyses are commonly employed in the field of spatial modeling. The primary multiple regression to analyze the spatial relationship between a quantitative variable (dependent variable) and explanatory variables (independent variables) is ordinary least squares (OLS) regression. OLS is easy to compute and lead to easier determination of statistical parameters. Geographically weighted regression (GWR) is an extension of OLS by including the geographic location for each data point. GWR is excellent at picking up broad-scale regional differences to visualize geographical interactions. As a consequence, OLS and GWR have become popular in spatial data analysis (e.g., Wang, Zhang, and Li 2012; Yan et al. 2014; Yuan, Wang, and Mitchell 2014) and have been used to examine the relationship between population patterns and land characteristics (Joseph, Wang, and Wang 2012; Yu, Morton, and Peterson 2014).

Although numerous studies have used a combination of DMSP nighttime lights data and population census data to investigate urban growth (e.g., Small, Pozzi, and Elvidge 2005; Zhang, He, and Liu 2014), to our knowledge, quantitative estimation of population density based on the combined use of population census, land-use, and DMSP nighttime lights data has not been published.

In this article, we first examine the relationships among gridded population census, land-use, and DMSP nighttime lights data. Then, we use OLS model to estimate the distribution of population density in the Hokkaido region. Finally, we use GWR and OLS regression models for Sapporo city to quantitatively compare the modeled population density with population census data.

2. Data collection and methods

2.1. Data

We analyzed three different types of data to quantify the urban spatial extent in Japan: population census data from the Statistical Information Institute for Consulting and

Analysis (Sinfonica; Basic Grid Square data, spatial resolution, 1 km; Sinfonica 2012) for 1990 and 2005, land-use map data from the Geospatial Information Authority of Japan (GSI; spatial resolution, 100 m; GSI 2012) for 1991 and 2006, and mean annual stable nighttime lights data from DMSP for 1992 and 2006 (30 arc second spatial resolution).

The Basic Grid Square population census data are available from the Statistics Bureau of Japan. Each square (area about 1 km^2) is assigned a unique 8-digit ID number based on its longitude and latitude that is used as location information in the National Land Numerical Information databank (Statistics Bureau, Japan 1973).

The GSI land-use maps use the Basic Grid Square system and show 11 land-use categories; thus, the grid cell ID numbers are the same as those used for the population census data. Here, the terms of land use and land cover are as follows: land cover is the observed biophysical cover on the earth's surface; land use is characterized by the actions of people in a certain land-cover type to produce, change, or maintain it (Maimaitijiang et al. 2015). To minimize issues of collinearity between land-use classes and to smooth the analysis, we aggregated the GSI land-use data from 11 to 5 classes. We aggregated cropland and other agricultural land-use categories into a common "cropland" class; we merged bare soil, golf courses, and beaches and riverbanks into a single "grassland" class; and we combined all classes dominated by man-made structures such as buildings and roads to create a broadly defined urban/built-up land-use class (hereafter, "urban" for brevity). The water-dominated categories such as rivers and lakes were merged into "water." The "forest" class was unchanged. Figure 1 shows the five-category land-use map for Japan in 2006.

The mean annual stable nighttime lights images from DMSP have a 30 arc second spatial resolution. Each image pixel is associated with a DMSP digital number (DN) between 0 (no light) and 63. To link the DMSP data to the Basic Grid Square cells, it has to be decided how to process a DMSP pixel that is located on a grid line (or spans across multiple grid lines). If we assigned a DMSP pixel located on a grid line to the cell that covered the largest part of the pixel, it would introduce an error due to the spatial resolution of the DMSP pixel, which may affect the accuracy of the spatiotemporal analysis and the calculation of the correlation coefficients. For example, in the worst case, almost 50% of the DMSP pixel area may be incorrectly assigned to neighboring cells.

To reduce the error caused by the pixel size, we resampled the DMSP maps to a 1-m spatial resolution pixel size. After resampling, the area inaccurately assigned to a neighboring cell's pixel area became less than 0.5 m^2. We then assigned such a pixel located on a grid line (or spanning across multiple grid lines) to the cell that covered the largest part of the pixel.

Then we used Esri® ArcGIS 10.2 software (Redlands, CA, USA) to overlay the resampled DMSP images onto a map with 1 km^2 grid cells and to compute the percentage of each DN value within each cell and store the results in a new attribute table. Thus, the resulting attribute table included 64 newly added DMSP attributes (i.e., DN values from 0 to 63).

2.2. Statistical methods

We used a classic correlation coefficient analysis to reveal the relationships of DMSP values with population density and with urban area at local, regional, and national levels. Then, we used OLS regression and GWR to model the spatial relationship of population density with urban area and DMSP values at local and regional levels.

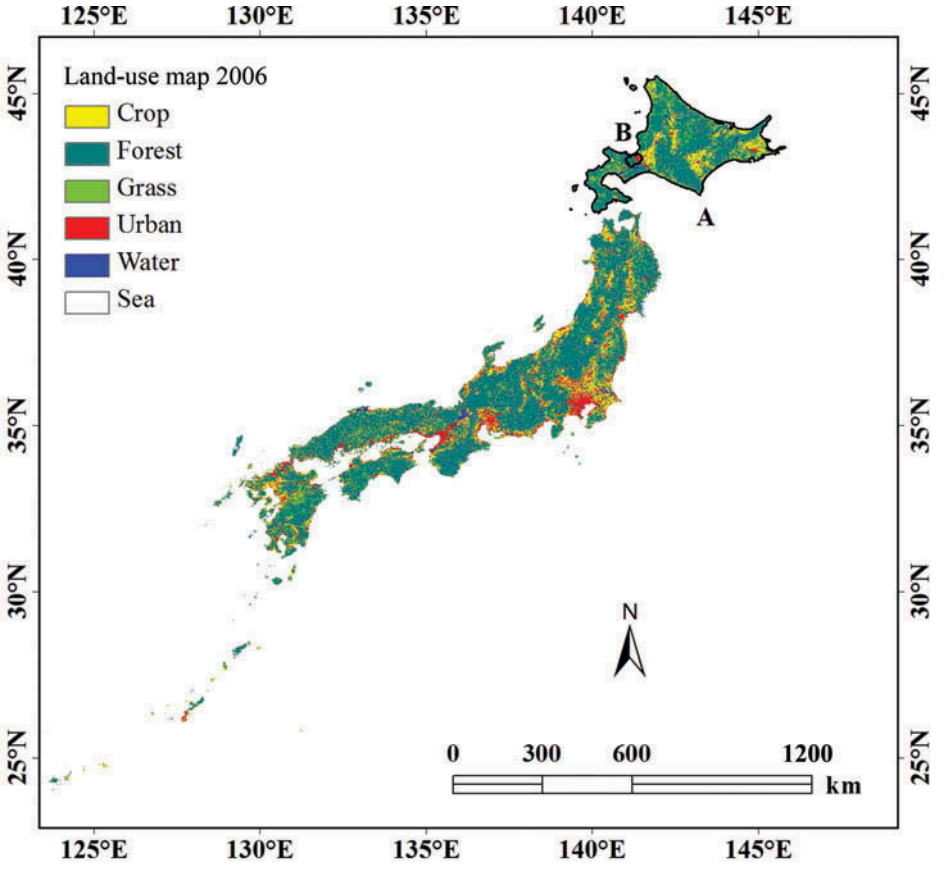

Figure 1. A five-category land-use map of Japan for 2006 based on GSI data. Hokkaido (labeled A, see Figure 5) and Sapporo city in Hokkaido (labeled B, see Fig. 10) are outlined in black. For full color versions of the figures in this paper, please see the online version.

OLS regression utilizes all observations within a data set to determine the coefficients of a linear model that result in the best fit of the model to the data. Therefore, the OLS linear model does not account for any spatial heterogeneity present in the relationship. The OLS model is calculated as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{i,j} + E_i \tag{1}$$

where Y_i (i = 1, 2,. . ., n) is the ith observation of the dependent variable, and $X_{i,j}$ (i = 1, 2,. . ., n; j = 1, 2,. . ., m) is the ith observation of the jth independent variable. β_0 and β_j are the regression coefficients, which are determined from a sample of n observations by the least squares method. E_i is the residual, which is the difference between the observed value of the dependent variable (Y_i) and the predicted value ($\hat{Y}_i$). Each data point has one residual, which is calculated as follows:

$$E_i = Y_i - \hat{Y}_i = Y_i - \left(\beta_0 + \sum_{j=1}^{m} \beta_j X_{i,j}\right) \tag{2}$$

The standardized residual, which is the residual divided by the standard deviation of the residuals, is used to compare the relative importance of different variables:

$$\varepsilon_i = \frac{E_i}{\sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-m-1}}} \tag{3}$$

In GWR, Equation (1) is modified so that individual parameters are estimated at each observation location. Thus, GWR attempts to capture spatial variation by calibrating a regression equation for each observation in which neighboring observations are weighted according to their distance from the focal observation, under the assumption that observations that are closer together have more impact on each other than on more distant observations. For each calibration location $i = 1, 2, \ldots, n$, the GWR model at location i is expressed as (Fotheringham, Brunsdon, and Charlton 2002)

$$Y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{m} \beta_j(u_i, v_i) X_{i,j} + E_i \tag{4}$$

where Y_i is the dependent variable, (u_i, v_i) indicates the spatial location of observation i, $\beta_0(u_i, v_i)$ is the intercept for location i, $\beta_j(u_i, v_i)$ is the local parameter estimate for the independent variable $X_{i,j}$ at location i, and E_i is the residual. The vector of the estimated regression coefficients at location i is

$$\hat{\beta}_j(u_i, v_i) = \left(X^T W(u_i, v_i) X\right)^{-1} X^T W(u_i, v_i) Y \ \ (j = 0, 1, \ldots, m) \tag{5}$$

where Y is the $n \times 1$ vector whose components are the dependent variables Y_i; X is a matrix, called the design matrix, whose elements are values of the explanatory variables and a column of 1's for the intercept; and $W(u_i, v_i)$ is a weighting matrix whose diagonal elements w_{ij} represent the geographical weightings of observations around point i. The Gaussian kernel function and Euclidian distance are most commonly used to calculate the weight of each point as a continuous function of distance:

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{6}$$

where w_{ij} is the weight for data at location j in the model estimated for location i; d_{ij} is the distance between regression point i and data point j; and b is the kernel bandwidth. For point coordinates u, v, distance is usually defined as the Euclidean distance:

$$d_{ij} = \sqrt{\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2} \tag{7}$$

3. Results and discussion

3.1. Urban area change and population density change

As explained in Section 2, we mapped the land-use data onto a grid of empty basic grid cells and computed the proportions of each cell occupied by the different land-use categories. Figure 2(a) shows the proportion of urban area in 1 km^2 grid cells in 2006, and Figure 2(b) shows the population density in 1 km^2 grid cells in 2005. The generation of gridded land-use maps facilitated our calculation of the percentage change in urban area from 1991 to 2006 at the 1-km scale. For this calculation, we subtracted the urban area in each grid cell in 1991 from that in 2006 (Figure 2(c)). We similarly computed the difference in population density between 1990 and 2005 by subtracting the 1990 grid cell population density values from the 2005 grid cell values (Figure 2(d)).

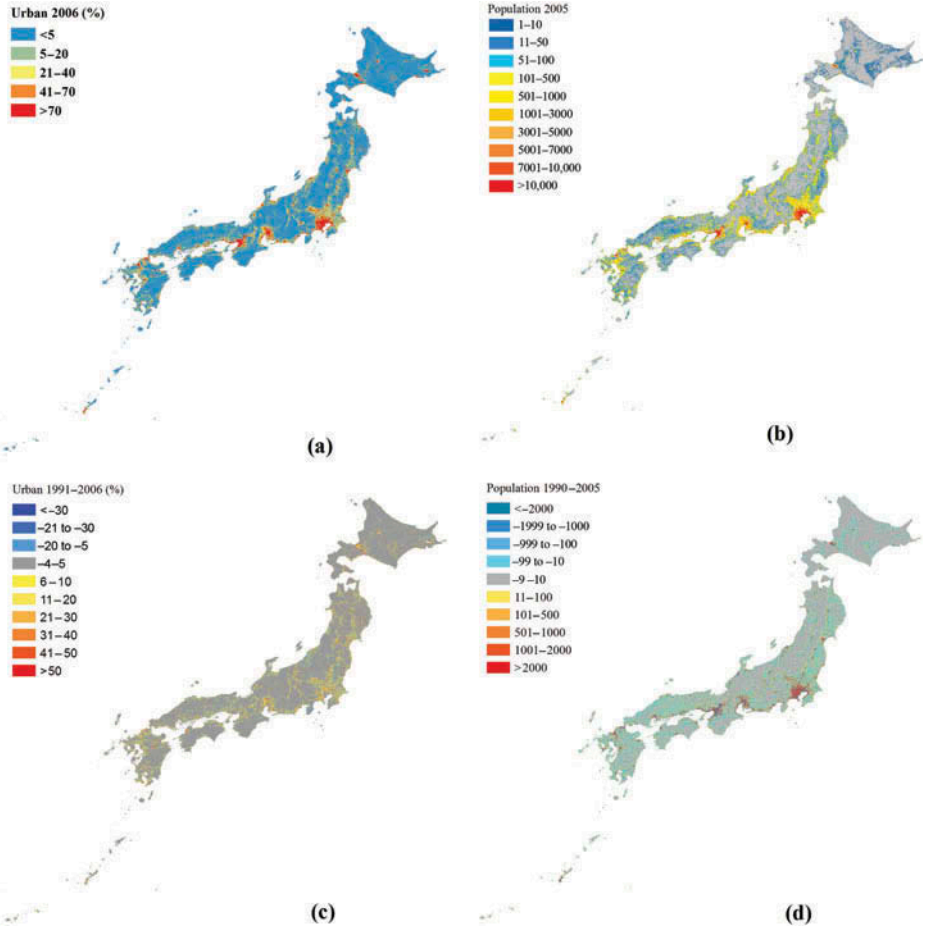

Figure 2. (a) Percentages of urban area in 1 km^2 grid cells in 2006. (b) Population densities in 1 km^2 grid cells in 2005; the gray color indicate no population. (c) Absolute percentage rates of change in urban areas from 1991 to 2006. (d) Absolute changes in population density between 1990 and 2005.

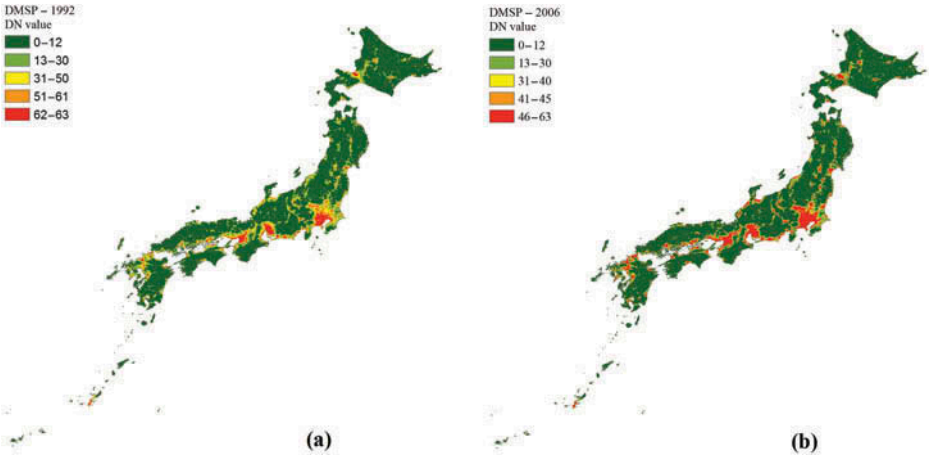

Figure 3. Urban land-use area based on DMSP data (a) in 1992 and (b) in 2006. The total area above the respective threshold DN value (red) in (a) and (b) is roughly equal to the urban land-use area according to the 1991 and 2006 land-use maps, respectively.

The total urban area increased from 24,250.7 km^2 in 1991 to 31,644.5 km^2 in 2006, whereas the overall population of Japan grew only slightly, from 123,611,000 in 1990 to 127,768,000 in 2005.

To represent the urban area in the DMSP data, we selected a DMSP DN threshold value of 62 for 1992 and a threshold value of 46 for 2006, so that the distribution of urban areas detected by DMSP would be roughly equivalent to their distributions on the corresponding 1991 and 2006 land-use maps (Figure 3). The total urban land-use area determined from the DMSP data was 23,892.1 km^2 in 1992 and 31,576.0 km^2 in 2006.

In Japan, where living standards are fairly uniform nationally, the higher concentrations of lights in megacities and around railway lines reflect higher population densities in those areas. As shown in Figures 1–3, the spatial distribution of population density was similar to both the percentage distribution of the urban land-use area and the distribution of DMSP brightness. In addition, the population density increased as DMSP DN values increased.

Although the population of Japan grew slowly between 1990 and 2005, regional population changes resulting from internal migration were substantial. As shown in Figure 2(c) and (d), in areas with large population increases, urban land-use area also tended to increase. In particular, the area of urban land use expanded rapidly around metropolitan regions (e.g., Tokyo, Osaka, Nagoya, Fukuoka, and Sapporo). In contrast, rural areas experienced population declines, and areas occupied by small cities became smaller. This pattern reflects the concentration of industry in megacities combined with deindustrialization and suburbanization in rural areas and small provincial cities (Okata and Murayama 2011; Frazier, Bagchi-Sen, and Knight 2013).

3.2. *Relationships among population census, DMSP, and land-use data*

We calculated the correlation coefficients between the land-use categories in 2006 and the census population density data in 2005 at local (Sapporo city, Hokkaido), regional (Hokkaido), and national (all of Japan) levels (Table 1). The results show that the urban area category was strongly and positively correlated with population density at all three levels.

Table 1. Correlations (r) between land-use categories in 2006 and census population density data in 2005 at three levels: (a) Sapporo city, (b) Hokkaido, and (c) all of Japan.

r	Cropland	Forest	Grassland	Urban	Water	Population
a. Sapporo						
Cropland	1					
Forest	−0.42939	1				
Grassland	0.085074	−0.17909	1			
Urban	0.004234	−0.77025	−0.2151	1		
Water	0.26466	−0.28621	−0.03701	0.003906	1	
Population	−0.0577	−0.672	−0.21297	**0.932712**	−0.01937	1
b. Hokkaido						
Cropland	1					
Forest	−0.73764	1				
Grassland	−0.14575	−0.27424	1			
Urban	0.051087	−0.33095	−0.04198	1		
Water	0.042273	−0.27699	−0.01997	0.035644	1	
Population	−0.02583	−0.20385	−0.03752	**0.847868**	0.012191	1
c. Japan						
Cropland	1					
Forest	−0.67467	1				
Grassland	−0.11086	−0.15794	1			
Urban	0.135436	−0.55014	−0.08755	1		
Water	0.051239	−0.28391	−0.03456	0.059643	1	
Population	0.015612	−0.35282	−0.06709	**0.770934**	0.029145	1

We next examined the correlations of DMSP DN values for 1992 and 2006 with urban land-use area in 1991 and 2006, respectively, and with population density in Japan in 1990 and 2005, respectively (Figure 4). In both 1990–1992 and in 2005/2006, the correlation coefficients between DMSP DN values and urban land-use area and between DMSP DN values and population density tended to increase as the DN value increased (Figure 4).

In 2006 (Figure 4(b)), the correlation coefficients suddenly decreased at the maximum DMSP DN value of 62, because DMSP nighttime lights are saturated in core city areas where the population density is very high (Raupach, Rayner, and Paget 2010; Frolking et al. 2013).

3.3. Predicting population density from DMSP and land-use data

These relationships among population census, land-use, and DMSP data suggested that it would be possible to predict population density by using a multivariate regression model. To investigate this possibility, we focused on Hokkaido, the largest region in Japan (Figure 1), and applied a global-level OLS regression model. Because the economic activity and living standards are fairly uniform across the country, Hokkaido can be viewed as a representative area for all Japan. Our focus here was the spatial distribution of the statistical relationships among urban land use, population density, and DMSP DN values, because we considered these to be the most important variables reflecting the socioeconomic and physical aspects of urban sprawl.

First, we visually compared the spatial distributions of the DMSP DN values (2006), population density (2005), and urban land use (2006) in Hokkaido (Figure 5) and found that their spatial distributions were broadly consistent. This result suggested that all three data sets are spatially correlated.

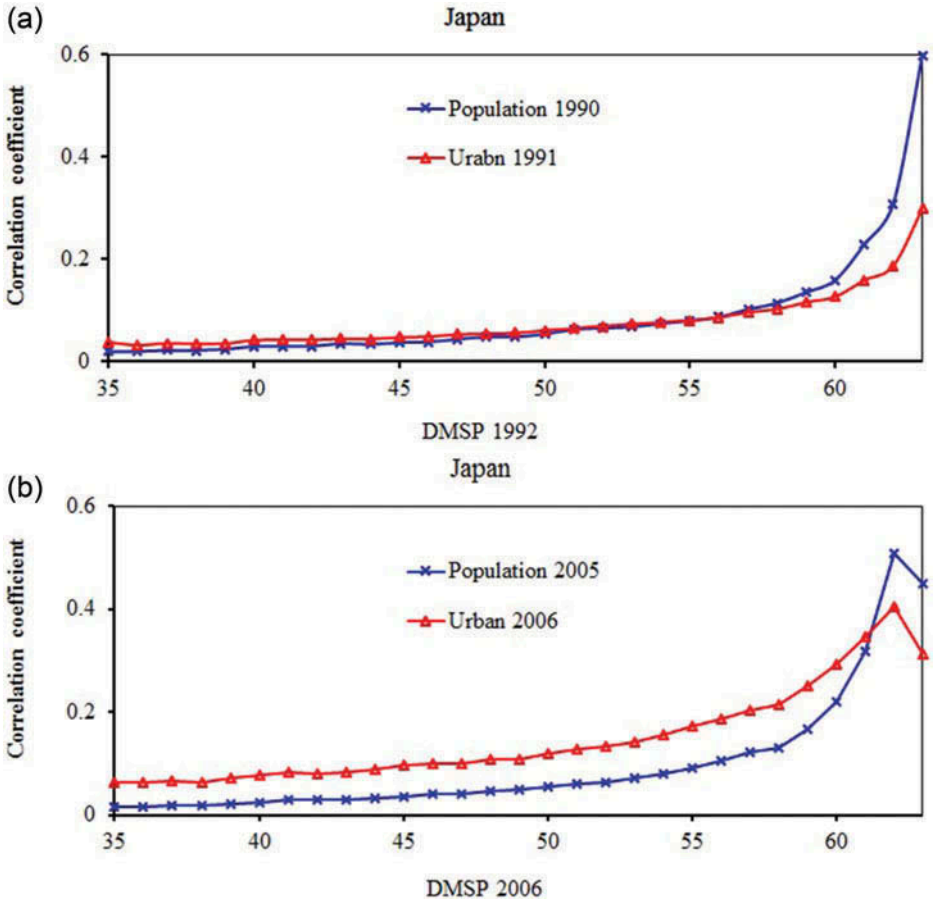

Figure 4. Correlation coefficients of DMSP DN values with urban land-use area and with population density in (a) 1992 and (b) 2006. The correlations were calculated using data for all 379,281 grid cells in Japan.

The total area of DMSP DN values between 35 and 63 in Hokkaido was 3215 km^2, which is nearly equal to the urban land-use area there (3196 m^2). The correlation coefficient between population density in 2005 and urban land-use area in 2006 was 0.6565. We also calculated the correlation coefficients between DMSP DN values and urban land-use area in 2006, and those between DMSP DN values in 2006 and population density in Hokkaido in 2005 (Figure 6).

The results showed that as the DN values of DMSP nighttime lights increased, the correlation coefficients between the DMSP DN values and population density, and between DMSP DN values and urban land-use area, also increased.

A total of 85,107 grid cells were used to develop and validate the OLS regression model for Hokkaido; one-third of the grid cells (28,369 cells) were randomly chosen to use for parameterization of the OLS model, and the remaining two-thirds (56,738 cells) were reserved for model validation. In this model, the dependent variable was census population data for 2005. The total number of explanatory variables was 35, including DMSP DN values ≥31 in 2006 and the urban land-use area in 2006. We omitted from our

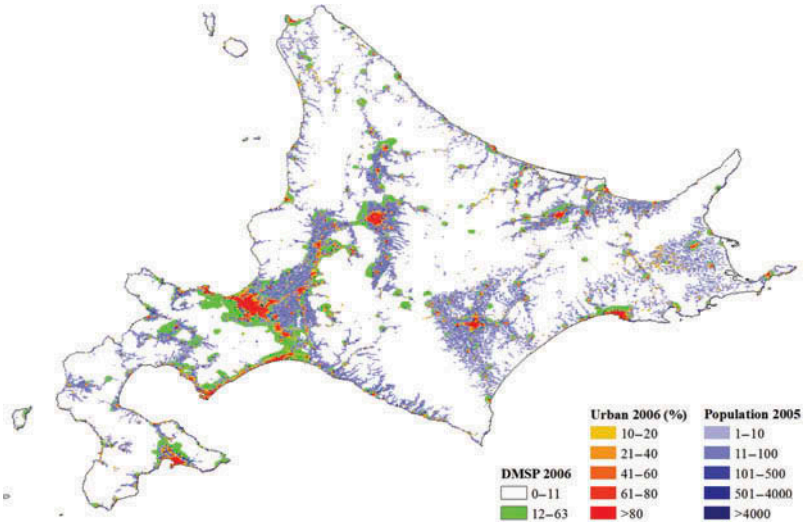

Figure 5. Spatial distributions of the urban land-use area in 2006 (red, top layer), DMSP in 2006 (green, middle layer), and population density in 2005 (blue, bottom layer).

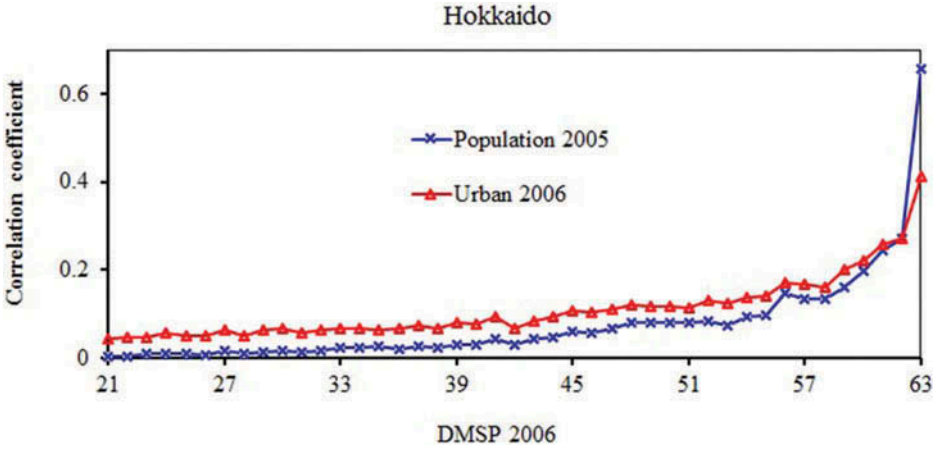

Figure 6. Correlation coefficients between DMSP DN values and urban land-use area and between DMSP DN values and population density in Hokkaido.

analysis DMSP DN values <31 because correlations between population density and small DMSP DN values were very low (Figure 6).

In the OLS model derived from the training data set, the adjusted overall R^2 was 0.8041, indicating that 80.41% of the variability in population density could be explained by the data set. Figure 7 shows the spatial distribution of the standardized residuals in Hokkaido. In this figure, the variances appear to be randomly distributed, although higher standardized residuals are apparently associated with city centers. The main reason for this result is that the values of the independent variable DMSP nighttime lights saturated in core city areas where the population density is very high (Raupach, Rayner, and Paget 2010). Thus, to improve the prediction accuracy of the model in the future, the analysis

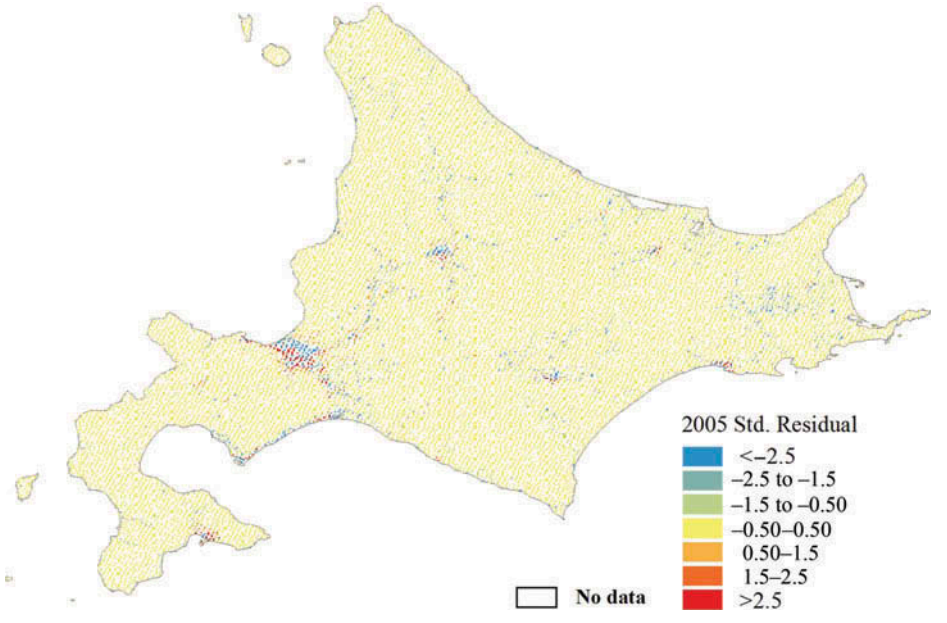

Figure 7. Spatial distribution of standardized residuals in Hokkaido, based on the training data set (28,369 grid cells).

should include data of urban 3D structures (Lu et al. 2010; Susaki, Kajimoto, and Kishimoto 2014). Indeed, when it comes to this analysis, urban 3D structures are not available. In addition, there are several methods been developed to minimize DMSP light saturation issues, including logarithm curve methods (Lo 2001) and cluster-specific threshold techniques (Zhou et al. 2014).

Figure 8 shows a scatter plot of the OLS-estimated population density in 2005, based on the validation data set (56,738 cells), versus the census population density in 2005. We found a strong, positive linear relationship between the OLS-estimated population density and the census population density ($R^2 = 0.8$) (Figure 8). In contrast, when only DMSP DN

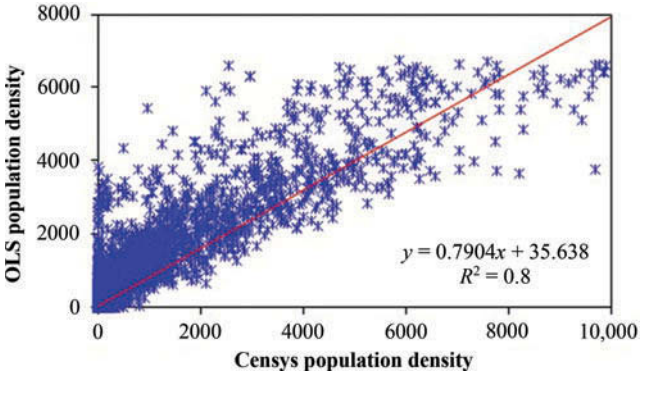

Figure 8. Scatter plot of the OLS-estimated population density in 2005, based on the validation data set (56,738 grid cells) versus census population density in 2005. The red line indicates the linear least-squares fit, and R^2 is the squared correlation coefficient.

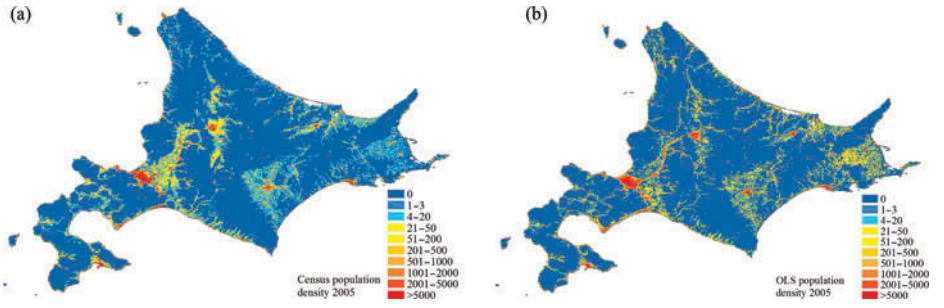

Figure 9. Distribution of population density in Hokkaido. (a) Actual population density in 2005 (census data). (b) Predicted population density in 2005 (OLS).

values ≥31 or only urban land-use area were used as explanatory variables, the adjusted overall R^2 was 0.5738 or 0.7236, respectively. Thus, by using DMSP data only or urban land-use data only, we would be unable to produce high-quality population density maps.

Moreover, the spatial distributions of census population density and OLS-estimated population density were very similar (Figure 9).

The regression result obtained with the training data set (adjusted overall $R^2 = 0.8041$) together with the spatial distribution of the standardized residuals shows that the OLS regression model fitted the data well. The results from the model indicate that satellite remote sensing technology can provide a simple but effective way of estimating regional population density, especially in areas without ground-based statistics data.

The findings also suggest that DMSP data and urban land-use data can be used for simulation of future population density using GIS and models at fine spatial resolution (Pijanowski et al. 2014).

3.4. *Predicting population density: comparison of OLS and GWR results*

Some previous research suggests that GWR may be better than OLS regression at explaining spatially varying relationships (Chen et al. 2012). To investigate this possibility with our data, we focused on Sapporo city, Hokkaido, and compared the performance of the GWR and OLS regression models using the same training data sets to create the OLS and GWR models, as before. We used data from all 1308 grid cells in Sapporo city (Figure 1) for the statistical analyses. To calculate the GWR, a Gaussian kernel function was employed as the weighting function, with an adaptive kernel bandwidth. The optimal bandwidth was determined automatically in ArcGIS 10.2 from the corrected Akaike information criterion (AICc). The GWR and OLS results were compared by comparing their adjusted R^2 and AICc values.

For Sapporo city, the adjusted R^2 (=0.8833) was higher and the AICc (=21,917) was smaller with GWR than with OLS regression ($R^2 = 0.8739$, AICc = 21,990). This adjusted R^2 of the GWR model indicates that 88.33% of the population density could be explained by the model, corresponding to a gain of +1%, compared with the explanatory power of the OLS model.

When we applied the GWR model to the entire Hokkaido region, the analysis failed, partly because of multicollinearity among the explanatory variables (Wheeler and Tiefelsdorf 2005) and more particularly because, in most of Hokkaido, both the urban area and DMSN DN values were zero (Figure 5). Future research should address this

limitation by using additional measures to examine the spatial dependence and spatial heterogeneity issues in population estimation and to avoid possible over-fitting issues. Furthermore, with regard to the future research, a number of individual-generated data sets (e.g., Geo-located mobile phone records, social media or social networking data) are available to estimate population distributions (Ratti et al. 2006; Reades, Calabrese, and Ratti 2009; Kang et al. 2012; Liu et al. 2015). Therefore, it is a promising approach that integrates such data sets with nightlight remote sensing data to estimate population density and reduce DMSP light saturation issues.

4. Conclusion

In this study, we demonstrated that the area of urban land use in Japan expanded rapidly around metropolitan regions from 1990 to 2006, and the spatial distribution of the expanded urban land-use area coincided with that of areas showing population increases. We also demonstrated that correlation coefficients between DMSP DN values and urban land-use area, and between DMSP DN values and population density, tended to increase as the DMSP DN value increased. These results suggest that DMSP data and land-use data have the potential to predict population density.

We employed OLS regression and GWR models to estimate population density at a spatial resolution of about 1 km. Case studies for Sapporo and all of Hokkaido showed that the OLS model showed good stability, robustness, and predictive capacity at both local and regional levels. Although the GWR model performed better than the OLS model at the local level, it failed at the regional level because of multicollinearity problems.

The models described in this article are applicable to other countries, because gridded population data with a 2.5 arc-minute resolution (Gridded Population of the World) are available from the Center for International Earth Science Information Network (2005) and can be used in conjunction with global land-cover maps and DMSP data to model population distributions. Future research should integrate DMSP data, land-cover maps and models (e.g., regression tree, random forest, logistic regression model, neural network, and cellular automata methods) to simulate population density and compare them to other models such as LandScan project (Dobson et al. 2000).

Highlights

- ► We examined relationships among population, nighttime lights, and urban land use.
- ► Both population density and urban area increased in metropolitan regions.
- ► Associations among nighttime lights, land use, and population density are strong.
- ► Land-use and nighttime lights data can be used to predict population density.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by The Environment Research and Technology Development Fund (S-10) of the Ministry of the Environment, Japan [S-10].

References

Bagan, H., and Y. Yamagata. 2012. "Landsat Analysis of Urban Growth: How Tokyo Became the World's Largest Megacity during the Last 40 Years." *Remote Sensing of Environment* 127: 210–222. doi:10.1016/j.rse.2012.09.011.

Brockerhoff, M. 2000. "An Urbanizing World." *Population Bulletin* 55 (3). Accessed 15 April 2015. http://www.prb.org/source/acfac3f.pdf.

Butt, M. J. 2012. "Estimation of Light Pollution Using Satellite Remote Sensing and Geographic Information System Techniques." *GIScience & Remote Sensing* 49 (4): 609–621. doi:10.2747/1548-1603.49.4.609.

Center for International Earth Science Information Network (CIESIN), Columbia University, & Centro Internacional de Agricultura Tropical (CIAT). (2005). *Gridded Population of the World*, Version 3 (Gpwv3): *Population Density Grid*. Accessed 15 April 2015. http://sedac.ciesin.columbia.edu/gpw.

Chen, G., K. G. Zhao, G. J. McDermid, and G. J. Hay. 2012. "The Influence of Sampling Density on Geographically Weighted Regression: A Case Study Using Forest Canopy Height and Optical Data." *International Journal of Remote Sensing* 33 (9): 2909–2924. doi:10.1080/01431161.2011.624130.

Dobson, J. E., E. A. Bright, P. R. Coleman, R. C. Durfee, and B. A. Worley. 2000. "A Global Population Database for Estimating Populations at Risk." *Photogrammetric Engineering & Remote Sensing* 66 (7): 849–857.

Doll, C. N. H., J. P. Muller, and J. G. Morley. 2006. "Mapping Regional Economic Activity from Night-Time Light Satellite Imagery." *Ecological Economics* 57: 75–92. doi:10.1016/j.ecolecon.2005.03.007.

Elvidge, C. D., B. T. Tuttle, P. C. Sutton, K. E. Baugh, A. T. Howard, C. Milesi, B. Bhaduri, and R. Nemani. 2007. "Global Distribution and Density of Constructed Impervious Surfaces." *Sensors* 7 (9): 1962–1979. doi:10.3390/s7091962.

Forbes, D. J. 2013. "Multi-Scale Analysis of the Relationship between Economic Statistics and DMSP-OLS Night Light Images." *GIScience & Remote Sensing* 50 (5): 483–499.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, Sussex: John Wiley.

Frazier, A. E., S. Bagchi-Sen, and J. Knight. 2013. "The Spatio-Temporal Impacts of Demolition Land Use Policy and Crime in a Shrinking City." *Applied Geography* 41: 55–64. doi:10.1016/j.apgeog.2013.02.014.

Frolking, S., T. Milliman, K. C. Seto, and M. A. Friedl. 2013. "A Global Fingerprint of Macro-Scale Changes in Urban Structure from 1999 to 2009." *Environmental Research Letters* 8 (2): 024004. doi:10.1088/1748-9326/8/2/024004.

Geospatial Information Authority of Japan (GSI). 2012. Maps, Aerial Photographs and Survey Results (in Japanese). Accessed 15 April 2015. http://www.gsi.go.jp/tizu-kutyu.html

Gerland, P., A. E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, et al. 2014. "World Population Stabilization Unlikely This Century." *Science* 346 (6206): 234–237. doi:10.1126/science.1257469.

Henderson, J. V., A. Storeygard, and D. N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028. doi:10.1257/aer.102.2.994.

Joseph, M., L. Wang, and F. Wang. 2012. "Using Landsat Imagery and Census Data for Urban Population Density Modeling in Port-Au-Prince, Haiti." *GIScience & Remote Sensing* 49 (2): 228–250. doi:10.2747/1548-1603.49.2.228.

Kang, C., Y. .Liu, X. Ma, and L. Wu. 2012. "Towards Estimating Urban Population Distributions from Mobile Call Data." *Journal of Urban Technology* 19 (4): 3–21. doi:10.1080/10630732.2012.715479.

Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. "Social Sensing: A New Approach to Understanding Our Socioeconomic Environments." *Annals of the Association of American Geographers* 105 (3): 512–530. doi:10.1080/00045608.2015.1018773.

Lo, C. P. 2001. "Modelling the Population of China Using DMSP Operational Linescan System Nighttime Data." *Photogrammetric Engineering & Remote Sensing* 67: 1037–1047.

Lu, J., and J.-M. Guldmann. 2012. "Landscape Ecology, Land-Use Structure, and Population Density: Case Study of the Columbus Metropolitan Area." *Landscape and Urban Planning* 105 (1–2): 74–85. doi:10.1016/j.landurbplan.2011.11.024.

Lu, Z., J. Im, and L. J. Quackenbush. 2011. "A Volumetric Approach to Population Estimation Using Lidar Remote Sensing." *Photogrammetric Engineering & Remote Sensing* 77 (11): 1145–1156. doi:10.14358/PERS.77.11.1145.

Lu, Z., J. Im, L. J. Quackenbush, and K. Halligan. 2010. "Population Estimation Based on Multi-Sensor Data Fusion." *International Journal of Remote Sensing* 31: 5587–5604. doi:10.1080/01431161.2010.496801.

Maimaitijiang, M., A. Ghulam, J. S. O. Sandoval, and M. Maimaitiyiming. 2015. "Drivers of Land Cover and Land Use Changes in St. Louis Metropolitan Area over the past 40 Years Characterized by Remote Sensing and Census Population Data." *International Journal of Applied Earth Observation and Geoinformation* 35: 161–174. doi:10.1016/j.jag.2014.08.020.

Okata, J., and A. Murayama. 2011. "Tokyo's Urban Growth, Urban Form and Sustainability." In *Megacities: Urban Form, Governance, and Sustainability*, edited by A. Sorensen and J. Okata, 15–41. Tokyo: Springer.

Pijanowski, B. C., A. Tayyebi, J. Doucette, B. K. Pekin, D. Braun, and J. Plourde. 2014. "A Big Data Urban Growth Simulation at A National Scale: Configuring the GIS and Neural Network Based Land Transformation Model to Run in A High Performance Computing (HPC) Environment." *Environmental Modelling & Software* 51: 250–268. doi:10.1016/j.envsoft.2013.09.015.

Pozzi, F., C. Small, and G. Yetman. 2003. "Modeling the Distribution of Human Population with Nighttime Satellite Imagery and Gridded Population of the World." *Earth Observation Magazine* 12 (4): 24–30.

Propastin, P., and M. Kappas. 2012. "Assessing Satellite-Observed Nighttime Lights for Monitoring Socioeconomic Parameters in the Republic of Kazakhstan." *GIScience & Remote Sensing* 49 (4): 538–557. doi:10.2747/1548-1603.49.4.538.

Qian, T., H. Bagan, T. Kinoshita, and Y. Yamagata. 2014. "Spatial–Temporal Analyses of Surface Coal Mining Dominated Land Degradation in Holingol, Inner Mongolia." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (5): 1675–1687. doi:10.1109/JSTARS.2014.2301152.

Ratti, C., R. M. Pulselli, S. Williams, and D. Frenchman. 2006. "Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis." *Environment and Planning B: Planning and Design* 33: 727–748. doi:10.1068/b32047.

Raupach, M. R., P. J. Rayner, and M. Paget. 2010. "Regional Variations in Spatial Structure of Nightlights, Population Density and Fossil-Fuel CO_2 Emissions." *Energy Policy* 38 (9): 4756–4764. doi:10.1016/j.enpol.2009.08.021.

Reades, J., F. Calabrese, and C. Ratti. 2009. "Eigenplaces: Analysing Cities Using the Space–Time Structure of the Mobile Phone Network." *Environment and Planning B: Planning and Design* 36: 824–836. doi:10.1068/b34133t.

Seto, K. C., M. Fragkias, B. Güneralp, and M. K. Reilly. 2011. "A Meta-Analysis of Global Urban Land Expansion." *PLoS ONE* 6 (8): e23777. doi:10.1371/journal.pone.0023777.

Seto, K. C., B. Guneralp, and L. R. Hutyra. 2012. "Global Forecasts of Urban Expansion to 2030 and Direct Impacts on Biodiversity and Carbon Pools." *Proceedings of the National Academy of Sciences of the United States of America* 109 (40): 16083–16088. doi:10.1073/pnas.1211658109.

Shi, K., B. Yu, Y. Hu, C. Huang, Y. Chen, Y. Huang, Z. Chen, and J. Wu. 2015. "Modeling and Mapping Total Freight Traffic in China Using NPP-VIIRS Nighttime Light Composite Data." *GIScience & Remote Sensing* 52 (3): 274–289. doi:10.1080/15481603.2015.1022420.

Small, C., F. Pozzi, and C. D. Elvidge. 2005. "Spatial Analysis of Global Urban Extents from the DMSP-OLS Night Lights." *Remote Sensing of the Environment* 96 (3–4): 277–291.

Statistical Information Institute for Consulting and Analysis (Sinfonica). 2012. Grid Square Statistics of Population Census (in Japanese). Accessed 15 April 2015. http://www.sinfonica.or.jp/datalist/index.html

Statistics Bureau, Japan. 1973. Standard Grid Square and Grid Square Code Used for the Statistics. Accessed 15 April 2015. http://www.stat.go.jp/english/data/mesh/02.htm

Su, Y., X. Chen, C. Wang, H. Zhang, J. Liao, Y. Ye, and C. A. Wang. 2015. "A New Method for Extracting Built-Up Urban Areas Using DMSP-OLS Nighttime Stable Lights: A Case Study in the Pearl River Delta, Southern China." *GIScience & Remote Sensing* 52 (2): 218–238. doi:10.1080/15481603.2015.1007778.

Susaki, J., M. Kajimoto, and M. Kishimoto. 2014. "Urban Density Mapping of Global Megacities from Polarimetric SAR Images." *Remote Sensing of Environment* 155: 334–348. doi:10.1016/j.rse.2014.09.006.

Sutton, P., D. Roberts, C. D. Elvidge, and H. Meij. 1997. "A Comparison of Nighttime Satellite Imagery and Population Density for the Continental United States." *Photogrammetric Engineering and Remote Sensing* 63 (11): 1303–1313.

Wang, K., C. Zhang, and W. Li. 2012. "Comparison of Geographically Weighted Regression and Regression Kriging for Estimating the Spatial Distribution of Soil Organic Matter." *GIScience & Remote Sensing* 49 (6): 915–932. doi:10.2747/1548-1603.49.6.915.

Wheeler, D., and M. Tiefelsdorf. 2005. "Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression." *Journal of Geographical Systems* 7 (2): 161–187. doi:10.1007/s10109-005-0155-6.

Witmer, F., and J. O'Loughlin. 2011. "Detecting the Effects of Wars in the Caucasus Regions of Russia and Georgia Using Radiometrically Normalized DMSP-OLS Nighttime Lights Imagery." *GIScience & Remote Sensing* 48 (4): 478–500.

Yan, Y., R. T. James, F. Miralles-Wilhelm, and W. Tang. 2014. "Geographically Weighted Spatial Modelling of Sediment Quality in Lake Okeechobee, Florida." *GIScience & Remote Sensing* 51 (4): 366–389. doi:10.1080/15481603.2014.929258.

Yu, D., C. M. Morton, and N. A. Peterson. 2014. "Community Pharmacies and Addictive Products: Sociodemographic Predictors of Accessibility from a Mixed GWR Perspective." *GIScience & Remote Sensing* 51 (1): 99–113. doi:10.1080/15481603.2014.886457.

Yuan, F., C. Wang, and M. Mitchell. 2014. "Spatial Patterns of Land Surface Phenology Relative to Monthly Climate Variations: US Great Plains." *GIScience & Remote Sensing* 51 (1): 30–50. doi:10.1080/15481603.2014.883210.

Zhang, Q., C. He, and Z. Liu. 2014. "Studying Urban Development and Change in the Contiguous United States Using Two Scaled Measures Derived from Nighttime Lights Data and Population Census." *GIScience & Remote Sensing* 51 (1): 63–82. doi:10.1080/15481603.2014.883212.

Zhou, Y., S. J. Smith, C. D. Elvidge, K. Zhao, A. Thomson, and M. Imhoff. 2014. "A Cluster-Based Method to Map Urban Area from DMSP/OLS Nightlights." *Remote Sensing of Environment* 147: 173–185. doi:10.1016/j.rse.2014.03.004.