Modeling the Population of China Using DMSP Operational Linescan System Nighttime Data

C.P. Lo

Abstract

Radiance-calibrated DMSP-OLS nighttime lights data of China acquired between March 1996 and January-February 1997 were evaluated for their potential as a source of population data at the provincial, county, and city levels. The light clusters were classified into six categories of light intensity, and their areal extents were extracted from the image. Mean pixel values of light clusters corresponding to the settlements were also extracted. A light volume measure was developed to gauge the three-dimensional capacity of a settlement. A density of light cluster measure known as percent light area was also calculated for each spatial unit. Allometric growth models and linear regression models were developed to estimate the Chinese population and population densities at the three spatial levels using light area, light volume, pixel mean, and percent light area as independent variables. It was found that the DMSP-OLS nighttime data produced reasonably accurate estimates of non-agricultural (urban) population at both the county and city levels using the allometric growth model and the light area or light volume as input. Non-agricultural population density was best estimated using percent light area in a linear regression model at the county level. The total sums of the estimates for non-agricultural population and even population overall closely approximated the true values given by the Chinese statistics at all three spatial levels. It is concluded that the 1-km resolution radiance-calibrated DMSP-OLS nighttime lights image has the potential to provide population estimates of a country and shed light on its urban population from space.

Introduction

The Defense Meteorological Satellite Program (DMSP), which was started in the 1970s (Brandli, 1978), consists of several satellites placed in circular, sun-synchronous, near-polar orbits at an altitude of about 830 km. Because each satellite crosses any point on the Earth up to two times a day with an orbital period of about 101 minutes, nearly complete global coverage can be obtained every six hours. These satellites carry different sensors, i.e., Operational Linescan System (OLS), Microwave Imager (SSM/I), Atmospheric Water Vapor Profiler (SSM/T-2), Precipitating Electron and Ion Spectrometer (SSJ/4), Atmospheric Temperature Profiler (SSM/T), Ion Scintillation Monitor (SSI/ES), and Magnetometer (SSM) (http://www.ngdc.noaa.gov/dmsp). They provide valuable data on the atmosphere, the ocean, and the solar-geophysical environment of the Earth.

This paper focuses only on the nighttime image data acquired from the Operational Lincescan System (OLS). The OLS is used to monitor the distribution of clouds and cloud top temperatures twice daily, once during the day and once at night. It consists of two telescopes and a photo multiplier tube (PMT), which scan the Earth in a whiskbroom scanning mode with a swath width of about 3,000 km. One telescope is sensitive to radiation in the visible and near infrared wavelengths (0.40 to 1.10 μm), while the other telescope images in the thermal infrared wavelengths (10.0 to 13.4 μm). The spatial resolution is 0.55 km in fine mode and 2.7 km in smooth mode. The *PMT*, which detects radiation in the visible and near-infrared wavelength region of 0.47 to 0.95 μm under very weak lighting condition (with radiance measurable to 10^{-9} watts cm^{-2} $sr^{-1}\mu^{-1}$), is used to detect clouds at night using moonlight, with a spatial resolution of 2.7 km at nadir (Elvidge *et al.*, 1997). It produces a unique data set of city lights, gas flares, fires, and lighted objects, which makes it possible to visualize the size and distribution of human population on a continental basis (Croft, 1978). By using the size and shape of the bright spots representing the cities, an estimation of power utilization can be assessed (Brandli, 1978).

For the past 20 years, these OLS nighttime data were archived in the form of film strips only. As a result, radiance values could only be obtained by measuring directly from a positive film transparencies using a microdensitomer, as demonstrated by the work of Welch and Zupko (1980), who calculated the volume of the "light dome" of each city based on microdensitomer scanned profiles. It was found that the volume correlated strongly with the city's energy consumption. Since 1992, a digital archive for the DMSP program was established at the NOAA National Geophysical Data Center, thus making quantitative analysis of DMSP imagery easier. This has already rekindled interest in the use of the DMSP-OLS nighttime images for population estimation and for the creation of a global population database (Sutton *et al.*, 1997; Dobson *et al.*, 2000). Sutton *et al.* (1997) compared the population density map of the United States derived from the 1990 U.S. census with the composite DMSP-OLS image of the continental United States. They concluded that the two exhibited strong correlation at aggregate scales, and that clusters of saturated areas in the DMSP-OLS images strongly correlated with populations covered by those clusters.

This paper is an attempt to extend the evaluation of the potential of the DMSP-OLS imagery for population modeling to another cultural realm, China, which has about the same area of the United States but a population of over 1.2 billion, or 4.5 times that of the United States. Therefore, China has a much greater population density than that of the United States. This evaluation makes use of both the areal extents of the saturated clusters on the images and the mean pixel values of these clusters to model population at three different spatial scales: province, county, and city.

Department of Geography, The University of Georgia, Athens, GA 30602 (chpanglo@uga.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 67, No. 9, September 2001, pp. 1037–1047.

0099-1112/01/6709–1037$3.00/0

Data

The DMSP-OLS data of China for this study are a composite of cloud-free radiance-calibrated low-light data obtained between March 1996 and January-February 1997 (Elvidge *et al.*, 1999). This type of data has an advantage over the "percent lighted" data, a version of the OLS data, which recorded a percentage of times light was seen relative to the cloud free orbits times as the value for each pixel. The "percent lighted" data suffer from "blooming" caused by saturation of a pixel even with very weak light sources, thus causing urban areas to be exaggerated or picking up non-existent urban areas (Imhoff *et al.*, 1997). On the other hand, the radiance-calibrated nighttime lights data have been found to be capable of showing population density more accurately and providing detailed zonation within the city (Elvidge *et al.*, 1999).

The radiance-calibrated data have been specially obtained by using an OLS gain setting alternating between 24 db and 50 db. The radiances of the data range from the lowest of $1.54*10^{-9}$ W cm^{-2} sr^{-1} μm^{-1} to a maximum of $3.17*10^{-7}$ W cm^{-2} sr^{-1} μm^{-1}. The pixel is a digital number (DN) in byte format with a value between 0 and 255 . The value between 0 and 254 can be converted to radiance using the following equation:

$$\text{Radiance} = DN^{3/2} * 10^{-10} \text{ watts } cm^{-2}sr^{-1}\mu m^{-1} \quad (1)$$

The data have been processed by an algorithm developed by Elvidge *et al.* (1997; 1999), which detects the visible/near-infrared emission sources and clouds from a large number of nighttime OLS images and maps them to a consistent reference grid. Using a time series of images, stable lights produced by cities, towns, and industrial establishments can be separated from ephemeral lights caused by gas flares, boats, fires, and lightning. The reference grid used is the NASA-USGS Global 1-km AVHRR project interrupted Goode Homolosine Projection, which is best to provide a uniform grid cell size at all latitudes and contiguous land masses (except Antarctica) (Steinwand, 1994).

In this research, the radiance-calibrated data allow more direct correlation with population density and energy consumption, along the same line of reasoning employed by Welch and Zupko (1980) and the conventional approach to population estimation using satellite images (Lo, 1995). Figure 1 shows the DMSP-OLS radiance data of China, and Figure 2 shows the frequency distribution curve of the data with DN values greater than zero. The DN value of the stable lights varied from 5 to 166 with a mean of 12.52 (standard deviation ±8.27), a median of 10, and a mode of 9. The frequency distribution curve is highly skewed to the left or the lower value side. Over 40 percent of the pixels have DN values of 8, 9, and 10. Therefore, the radiance-calibrated DMSP-OLS low-light data have rather limited range in pixel values.

Data on China administrative regions at the provincial and county levels at the scale of 1:1 million in Arc/Info format for 1990 were obtained from the China Dimension project of the Consortium for International Earth Science Information Network (CIESIN) (http://www.ciesin.org). A ChinaA Dataset, which contains China's 1990 population and economic data at the county level in dbf format, is also available. The dbf format allows these data to be associated with the county boundary data for display in ArcView GIS (or in Arc/Info). Additional demographic and socio-economic data of China at the provincial level for the year 1997 were also obtained from the *China Statistical Yearbook, 1998* (State Statistical Bureau, 1998). All these data are needed in order to investigate the potential of the radiance-calibrated DMSP-OLS nighttime images of China for population estimation and modeling in this study. It is difficult to get population and socio-economic data of China at the county level at exactly the years when the DMSP-OLS images were acquired.

Methodology

Previous works on population estimation using a global scale satellite image made use of the areal extent of a settlement as measured from the image as input to an allometric growth model, based on Tobler's observation of settlement sizes in the Nile River Delta (Tobler, 1969), in the following form:

$$P = aA^b \quad (2a)$$

or its logarithmic form,

$$\log P = \log a + b \log A, \quad (2b)$$

where P is population, A is the built-up area of the settlement, a

Figure 1. Radiance-calibrated DMSP-OLS nighttime composite image of China acquired between March 1996 and January-February 1997.

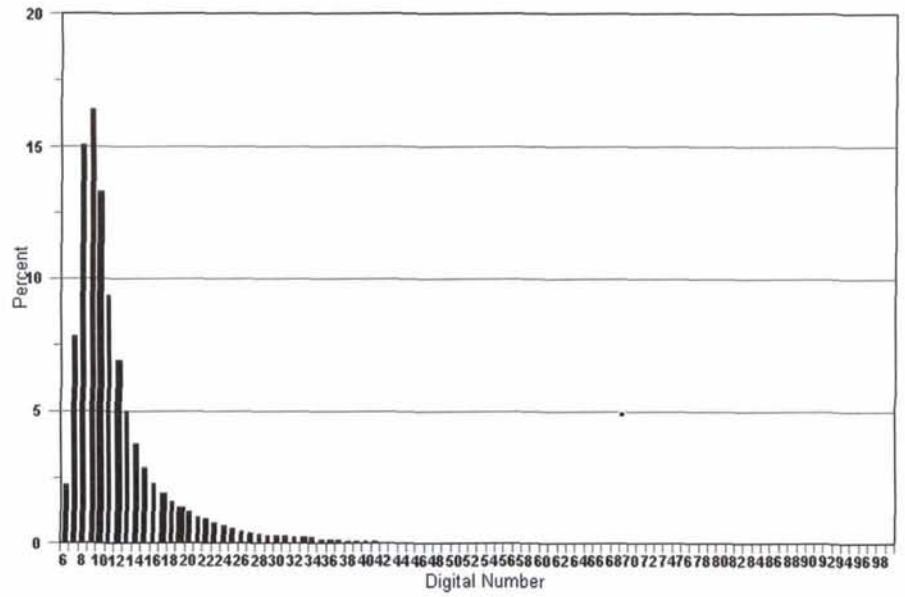

Figure 2. Frequency distribution curve of DN values of the radiance-calibrated DMSP-OLS nighttime image of China shown in Figure 1.

is a coefficient, and b is an exponent (Lo and Welch, 1977). Both a and b, which have to be empirically determined, tend to vary based on population density in a particular cultural realm. Another approach, which is more commonly used with large-scale multispectral satellite images such as Landsat MSS and TM, is to develop a linear regression model based on the pixel value of, say, four bands of the satellite data: i.e.,

$$PD = a + bX_1 + cX_2 + dX_3 + eX_4 \quad (3)$$

where PD is population density per pixel; X_1, X_2, X_3, and X_4 are pixel values for bands 1 through 4 of the satellite images; and a through e are the coefficients that have to be empirically determined (Lo, 1995).

In this analysis, both approaches will be used. The radiance-calibrated DMSP-OLS nighttime lights show different areal extents, which, when judged visually, seem to correlate positively with the sizes of the settlements. The areal extents of these light clusters can be easily extracted from the image using an image processing program such as ERDAS Imagine. The lights also show variations in light intensity (brightness) from one part of the city to another. It appeared that areas of higher building densities exhibited brighter light, and a positive correlation can therefore be established between population density and light intensity. Because the DMSP-OLS data have only one band, the linear regression model can be simplified to the following:

$$PD = a + bX \quad (4)$$

where X is the light intensity (DN value) per pixel of the settlement.

The spatial variation of light intensity in a settlement is useful information, and should be utilized in population modeling. The non-zero radiance-calibrated DMSP-OLS nighttime data (DN values) were classified using a parallelepiped classifier into the following six classes based on natural breaks: (CL_1) 5–7, (CL_2) 8–10, (CL_3) 11–14, (CL_4) 15–26, (CL_5) 27–99, and (CL_6) 100–166. The areal extent of each class of DN values was extracted. Following the approach of Welch and Zupko (1980), a volume of light of a settlement (V) was also calculated using the following formula:

$$V = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 \quad (5)$$

where $V_1 = 7*CL_1$, $V_2 = 9*CL_2$, $V3 = 11*CL_3$, $V_4 = 15*CL_4$, $V_5 = 27*CL_5$, and $V_6 = 100*CL_6$. Each of the multipliers for V_1 through V_6 is the highest occurrence DN value of that class. In this way, the correct intensity (height) was multiplied by the area of that class to give the volume. It is noteworthy that light volume is really the areal extent of the light cluster weighted by its light intensity. This simplified approach realistically approximated the volume of light in a complex multi-centered settlement. The mega-city of Shanghai, which exhibited complex spatial variation in light intensity, is shown in Figure 3. When compared with a simplified land-use map of the city (Figure 4), it is clear that the center of the built-up area (city center) has much stronger light than the surrounding area. The market gardening belt in turn is much brighter than the surrounding cropland area. The light intensity has a clear relationship with the type of land use/land cover cover, and depicts best the built-up area.

For comparison with volume in the analysis, the light area (without the light intensity) was also calculated simply as

$$LA = A_{CL1} + A_{CL2} + A_{CL3} + A_{CL4} + A_{CL5} + A_{CL6} \quad (6)$$

where LA is the light area and A_{CL1} through A_{CL6} are areas of DN classes 1 through 6, respectively. In this case, the light area is the surrogate of the built-up area of the settlement.

In the subsequent analysis using different spatial units (province and county), a percent of light was also calculated as follows:

$$PCLA = LA*100/LANDAREA \quad (7)$$

where $PCLA$ is the percent of light areas in a spatial unit, or light density, LA is the light area in the spatial unit, and $LANDAREA$

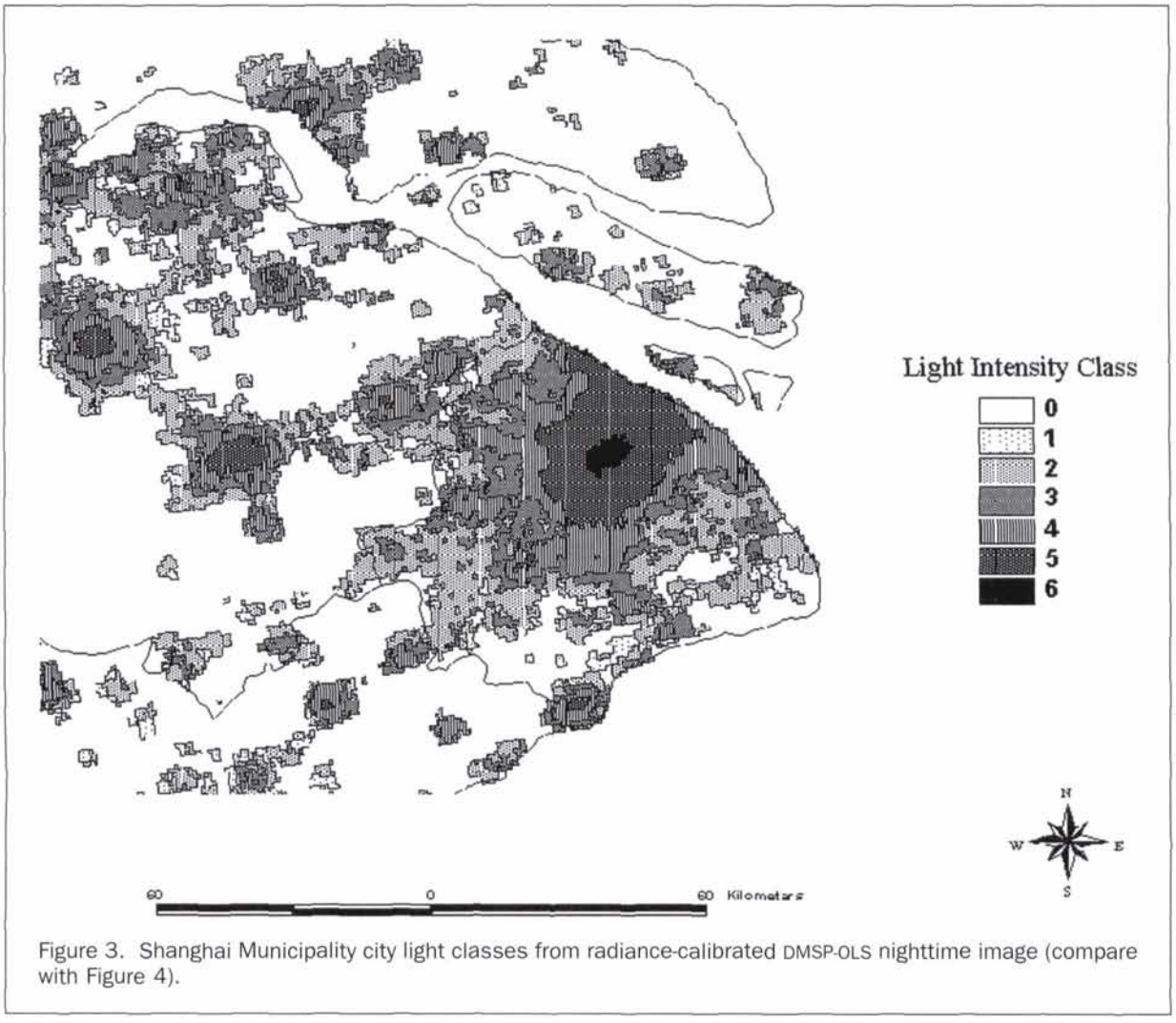

Figure 3. Shanghai Municipality city light classes from radiance-calibrated DMSP-OLS nighttime image (compare with Figure 4).

is the land area of the spatial unit. Intuitively, *PCLA* should have some associations with population density.

Population Modeling at the Provincial Level

Because more current demographic and socio-economic data of China are available at the provincial level, investigation of the relationship between the DMSP-OLS nighttime image data and the following variables is possible: (1) population in 1997, (2) number of households in 1997, (3) population density in 1997, (4) non-agricultural population in 1991, (5) urban per capita income in 1995, (6) rural per capita income in 1997, (7) energy consumption in 1988, (8) electricity consumption in 1988, and (9) Gross Value of Industrial Output in 1997. To register the DMSP-OLS image to the provincial units, the classified nighttime image in geographic coordinates was converted into polygons (vector format) using ERDAS Imagine. With the aid of the Arc/Info Project function, the geographic coordinates were transformed into the Lambert Conformal Projection of the China province boundary file. The vectorized DMSP-OLS coverage was then unioned with the province coverage using ArcView geoprocessing (Figure 5). As a result, the light intensity classes of the DMSP-OLS coverage were related to the above-mentioned nine variables at the provincial level. Only 28 provinces and municipalities in China were used. Because of the lack of data, the province of Hainan and the autonomous region of Xizhang (Tibet) were not included in the analysis. Also excluded from the analysis were Taiwan and the two Special Administrative Areas (SAR) of Hong Kong and Macao because their economic systems are quite different from that of Mainland China.

The tabulated data as extracted by ArcView GIS were exported for input to the SAS program, which computed the light areas (*LA*), percent of light areas in the province (PCLA), and the light volume (*V*) based on Equations 5, 6, and 7 given above. The relationships between these variables derived from the DMSP-OLS image and the nine population and socio-economic variables mentioned were investigated by computing the Pearson correlation coefficients (Table 1). It was clearly revealed that light area (*LA*) was strongly correlated with non-agricultural population (NAGPOP91), energy consumption (EGYCON88), electricity consumption (ELECON88), and Gross Value of Industrial Output (GVIO97). Light volume (*V*) has slightly improved on the strength of correlation with non-agricultural population only. However, the common logarithmic transformation of light volume (LOGV) has correlated even stronger with the common logarithmic transformation of non-agricultural population (LOGNAG91). This suggests the validity of the allometric growth model (Equation 2b) with the use of light volume *in lieu* of built-up area as input. The percent of light area by province (*PCLA*) was very strongly correlated with population density (POPDEN97), and strongly correlated with per capita rural income (RURINC97) and per capita urban income (URBINC95). It was also found that the percent of light area (*PCLA*) was very strongly correlated with pixel mean (PIX-MEAN) at the provin-

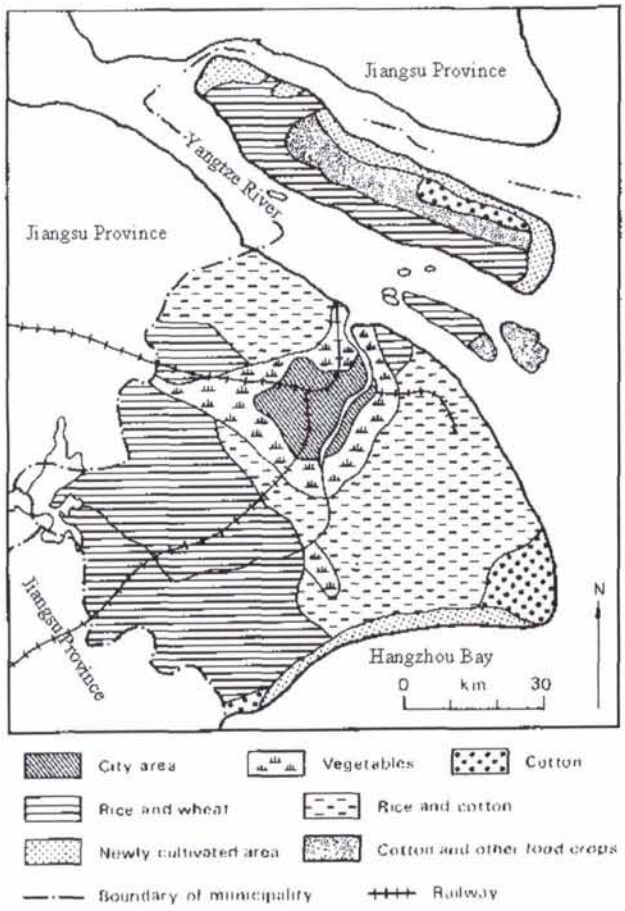

Figure 4. Land-use/land-cover map of Shanghai Municipality in the early 1970s (source: Lo, 1980).

cial level ($r = 0.95$ at the 0.0001 level of significance), which exhibited an even stronger correlation with population density, per capital urban income, and per capita rural income (Table 1). The correlation analysis showed that the light intensity of the DMSP-OLS nighttime image data could reveal some economic characteristics of the population (per capita income and the economic status of the province), while the areal extent of the city light (light area) was related to the size of the non-agricultural (i.e., urban) population as well as the consumption of energy and electricity in the province.

These findings suggested that it is possible to use light area, percent light area, and pixel mean to predict non-agricultural population and population density at the provincial level. To calibrate these models, a sample of ten provinces was randomly selected to provide the ground-truth data. These ten provinces/municipalities were Anhui, Fujian, Gansu, Guangdong, Hebei, Hubei, Liaoning, Qinghai, Shaanxi, and Shanghai, which exhibited a mix of population densities from the very high (Shanghai with 1821.25 persons/km^2) to the very low (Qinghai with 6.89 persons/km^2). The following three models were obtained:

$$\text{LOGNAGPOP} = -0.1290231314 + 0.6433115899*\text{LOGVOL} \quad (\text{with an } R^2 = 0.66) \quad (8)$$

NAGPOP = 10**LOGNAGPOP (the anti-log to get the actual number of non-agricultural population)

$$\text{POPDEN} = 94.15476782 + 29.59337367*PCLA \quad (\text{with an } R^2 = 0.94) \quad (9)$$

$$\text{POPDEN} = 166.064761 + 121.0099841*\text{PIX-MEAN} \quad (\text{with an } R^2 = 0.96) \quad (10)$$

These three models were then applied to the remaining 18 provinces or municipalities of China to determine the degree of accuracy achievable. The results are shown in Table 2. It is clear from the table that the allometric growth model, as represented by Equation 8, gave a mean relative error of 12.11 percent and a mean absolute relative error (which did not take into account of the signs) of 47.96 percent. Bad positive errors (i.e., overestimation) occurred in Ningxia, Shanxi, and Xinjiang, which have low population densities. The total estimated non-agricultural population of the 18 provinces (144,172,288) was only 5.2 percent underestimated (152,096,860). In applying this model (Equation 8) to all the 28 provinces, the total estimated non-agricultural population was 228,166,613, which is only about 4.6 percent underestimated from the actual 239,272,126 figure. In other words, it is possible to use this approach to estimate or check urban population size in China based on the light volumes of settlements in each province as extracted from DMSP-OLS nighttime images.

Although the population density models (Equations 9 and 10) appeared to have a much stronger R^2 in each case, the relative errors were large (103 percent for the *PCLA* model and 154 percent for the PIX-MEAN model) and the estimation results were disappointing, due to some huge relative errors caused by low-density provinces. Bearing this in mind, none of the three models should be used to estimate population in low-density provinces.

Population Modeling at the County Level

There are a total of 1880 counties in China (Figure 6). By moving to the county level, the spatial unit of population estimation using DMSP-OLS nighttime images has been greatly reduced as compared with provinces. The ChinaA Dataset from CIESIN provides county population data of the 01 July 1990 census. In this analysis, the following three population variables were used: (1) total population, (2) non-agricultural population, and (3) agricultural population.

Arc/View GIS was used to register the county boundary coverage with the vectorized light class coverage from the DMSP-OLS nighttime image. Because of the large number of counties involved, which exceeded the limit of the union function in ArcView GIS, the tabulated area in the Spatial Analyst extension was used instead to obtain all the area statistics of the variables at the county level. These data were then input to a SAS computer program for further statistical analysis. The correlation coefficients between population variables and DMSP-OLS derived variables by counties were computed (Table 3).

In interpreting the results of the correlation, one should note that, because of the very much larger sample size ($n = 1880$) at the county level, the correlation coefficients computed would be weaker compared with those obtained at the provincial level ($n = 28$) although they were all highly significant (at 0.0001). It is interesting to note that at the county level the areal extent of the light of the settlement or light area (*LA*) was not too strongly correlated with non-agricultural (urban) population. However, the pixel means (PIX-MEAN) exhibited very strong correlation with non-agricultural population (A117). The percent light area (*PCLA*) showed very strong correlation with all the three population density variables, in particular, the population density of the county (which included both agricultural and non-agricultural population). At the county level, the pixel mean did not correlate strongly with percent

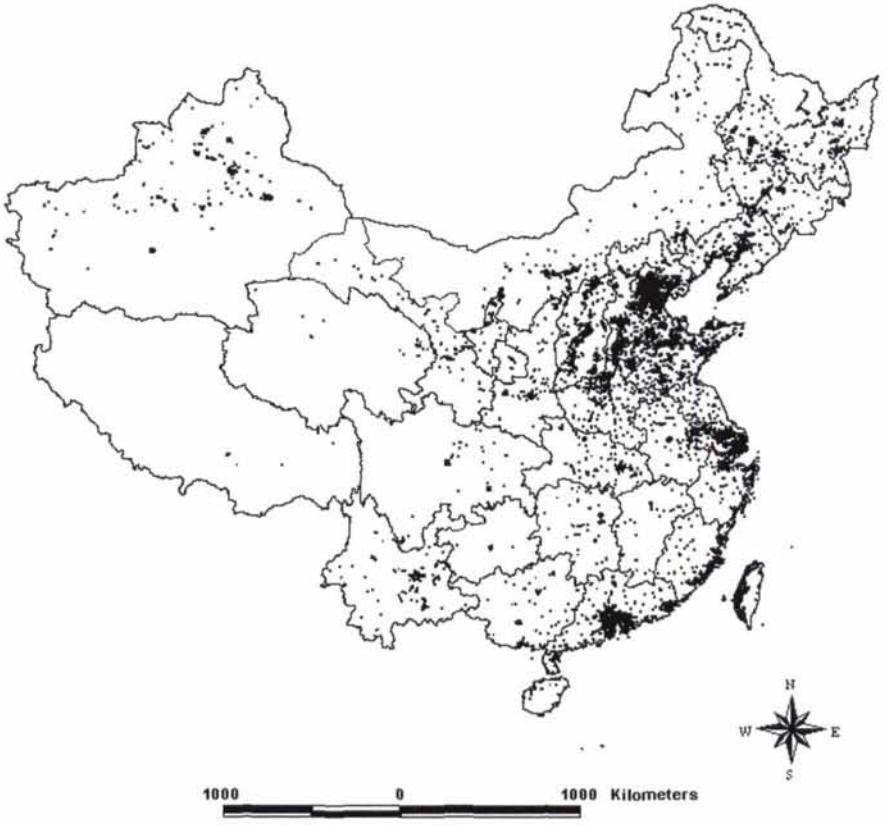

Figure 5. Radiance-calibrated DMSP-OLS nighttime image of China by province.

TABLE 1. CORRELATION COEFFICIENTS AMONG DMSP-OLS DERIVED VARIABLES AND POPULATION VARIABLES AT THE PROVINCIAL LEVEL (N = 28) (THE NUMBER IN PARENTHESES UNDER THE CORRELATION COEFFICIENT INDICATES THE LEVEL OF SIGNIFICANCE).

	DMSP-OLS Derived Variables				
Population Variables	*LA*	*PCLA*	*V*	LOG*V*	PIX-MEAN
POP97	0.44 (0.02)	-	0.42 (0.03)	0.34 (0.07)	−0.29 (0.14)
LOGPOP97	0.41 (0.03)	−0.31 (0.11)	0.40 (0.04)	0.39 (0.04)	−0.32 (0.10)
HH97	0.45 (0.02)	-	0.43 (0.02)	0.37 (0.05)	−0.24 (0.21)
POPDEN97	-	0.84* (0.0001)	-	-	−0.92* (0.0001)
NAGPOP91	0.57* (0.002)	-	0.59* (0.001)	0.58* (0.001)	0.02 (0.93)
LOGNAG91	0.53* (0.004)	0.11 (0.59)	0.55 (0.002)	0.64* (0.0003)0.09 (0.66)	
URBINC95	-	0.59* (0.0009)	-	-	0.64* (0.0002)
RURINC97	-	0.76* (0.0001)	-	-	0.79* (0.0001)
EGYCON88	0.65* (0.0002)	-	0.61* (0.0005)	0.61* (0.0006)	0.02 (0.92)
ELECON88	0.67* (0.0001)	-	0.66* (0.0001)	0.65* (0.0002)	0.13 (0.52)
GVIO97	0.57* (0.002)	-	0.62* (0.0004)	0.58* (0.001)	0.17 (0.39)

*Highly significant.

light area (and hence population density) as was the case at the provincial level. Light volume (V) was significantly correlated with total population and non-agricultural population. However, the logarithmic transformation of light volume (LOGV) exhibited a much stronger correlation with the logarithmic transformation of the non-agricultural population (LOGA117), suggesting again the validity of the allometric growth model as at the provincial level. Based on this simple correlation analysis, it is clear that both pixel mean (PIX-MEAN), light volume (V), and its logarithmic transformation (LOGV) as derived from the DMSP-OLS nighttime image can be used to estimate non-agricultural population (A117), while the percent of light area (*PCLA*) is best used to estimate the overall population density (A102/KM2) at the county level.

To calibrate the models, a random sample of 185 counties (or approximately 10 percent) was used. The following linear models were generated:

$$(1)\ \text{ESA102} = 307226.3575 + 0.0001 * V \quad (\text{with an } R^2 = 0.40) \tag{11}$$

$$(2)\ \text{ESA117} = -27505.39545 + 0.0001 * V \quad (\text{with an } R^2 = 0.56) \tag{12}$$

TABLE 2. ACCURACY OF ESTIMATING NON-AGRICULTURAL POPULATION AND POPULATION DENSITY BY PROVINCE FROM DMSP-OLS NIGHTTIME IMAGES OF CHINA BY APPLYING EQUATIONS 8, 9, AND 10.

PROV	NAGP91	ESTNAGP+	REL. ERR. (%)	POPDEN (p/km²)	EPOPDEN* (p/km²)	REL. ERR. (%)	ESPOPDEN** (p/km²)	REL. ERR. (%)
BJ	6485922	8006126	23.44	729.41	1309.27	79.50	961.49	31.82
GX	5814339	4199469	−27.77	196.31	128.46	−34.56	183.89	−6.33
GZ	4031296	1904486	−52.76	204.89	106.49	−48.03	173.45	−15.35
HLJ	15059574	12807727	−14.95	81.54	202.92	148.86	216.60	165.63
HN	11446843	10844929	−5.26	556.81	340.56	−38.84	276.70	−50.31
HU	9442439	4168887	−55.85	304.95	132.12	−56.68	186.37	−38.88
JS	14285106	11413046	−20.11	680.76	475.36	−30.17	367.29	−46.05
JX	7228163	3168373	−56.17	248.50	124.02	−50.09	182.98	−26.37
JL	9662230	8084616	−16.33	139.05	221.47	59.28	227.22	63.41
NMG	6710463	10204630	52.07	20.37	125.52	516.22	180.04	783.86
NX	1138938	2884367	153.25	80.30	171.14	113.13	211.76	163.71
SD	16500391	16995711	3.02	556.01	601.09	8.11	406.34	−26.92
SX	6384608	13139205	105.80	200.06	433.13	116.50	323.93	61.92
SC	6187070	4288185	−73.51	148.94	109.71	−26.34	173.77	16.67
TJ	4921945	7352310	49.38	794.17	1798.18	126.42	1068.02	34.48
XJ	5096430	9972694	95.68	10.33	111.12	975.73	177.43	1617.60
YN	4641061	6676327	43.85	106.61	137.67	29.13	189.35	77.61
ZJ	7060042	8061198	14.18	422.38	305.09	−27.77	286.93	−32.07
Mean	**8449826**	**8009572**	**12.11**	**304.52**	**379.63**	**103.36**	**321.87**	**154.14**
Sum	**152096860**	**144172288**	**−5.21**					
Mean Absolute Relative Error (%)			**47.96**		**138.08**	**181.05**		

+ ESTNAGP = non-agricultural population estimated using logarithm of light volume (Equation 8).
*EPOPDEN = population density estimated using percent of light areas (Equation 9).
** ESPOPDEN = population density estimated using pixel means (Equation 10).
Key to Province Names:
BJ = Beijing, GX = Guangxi, GZ = Guizhou, HLJ = Heilongjiang, HN = Henan, HU = Hunan, JS = Jiangsu, JL = Jilin, NMG = Neimenggu (Inner Mongolia), NX = Ningxia, SD = Shandong, SX = Shanxi, SC = Sichuan, TJ = Tianjin, XJ = Xinjiang, YN = Yunnan, Zj = Zhejiang.

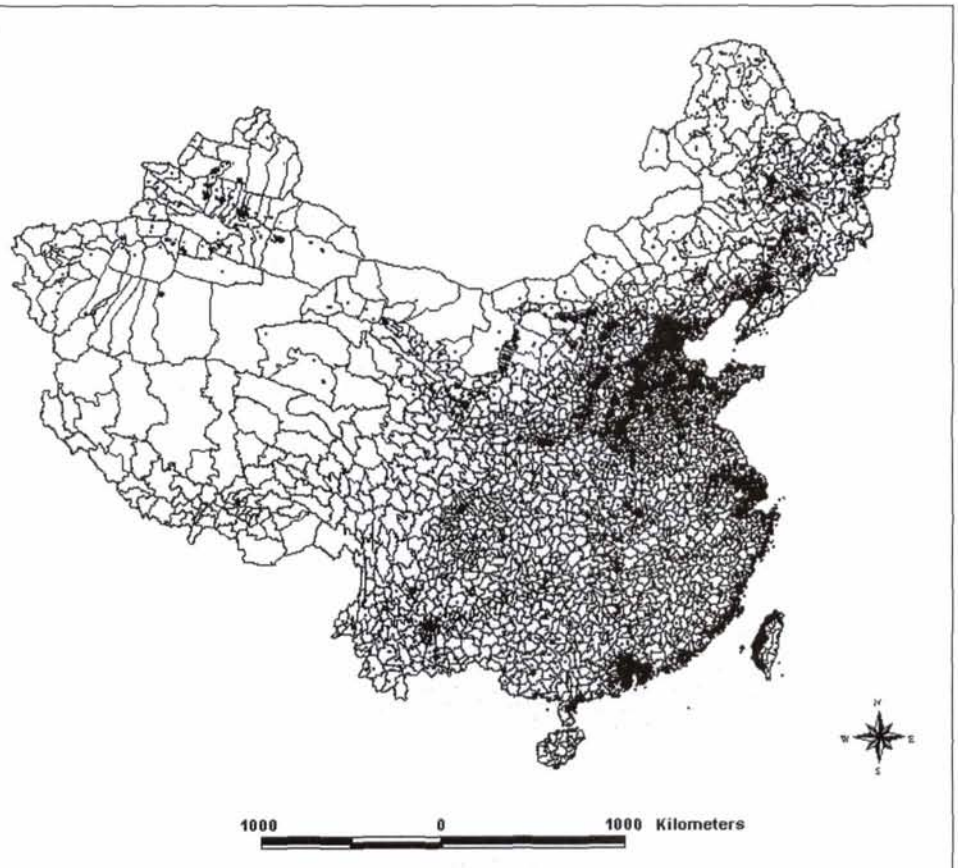

Figure 6. Radiance-calibrated DMSP-OLS nighttime image of China by county.

TABLE 3. CORRELATION COEFFICIENTS AMONG DMSP-OLS DERIVED VARIABLES AND POPULATION VARIABLES AT THE COUNTY LEVEL (N = 1880) (ALL SIGNIFICANT AT 0.0001 LEVEL WITH THE EXCEPTION OF THE ONE INDICATED).

Population Variables	DMSP-OLS				
	LA	*PCLA*	*V*	LOG*V*	PIX-MEAN
A102	0.35		0.42*	0.29	0.44*
A117	0.35		0.47*	0.31	0.70*
A120	0.14		0.12	0.10	−0.09
LOGA102	0.28			0.30	0.25
LOGA117	0.42			0.51*	0.53*
LOGA120	0.12			0.10	−0.08
A102/KM2		0.97*			0.14
A117/KM2		0.81*			0.17
A120KM2		0.76*			0.04 (0.10)

*Correlation coefficient > 0.40.
Key: A102 = total number of population, A117 = non-agricultural population, A120 = agricultural population.
A102/KM2, A117/KM2, and A120/KM2 are densities of the above defined population variables in persons per square kilometer.

$$\text{(3) PRA102} = 398109.6940 + 54765.2039*\text{PIX-MEAN} \quad (\text{with an } R^2 = 0.17) \tag{13}$$

$$\text{(4) PRA117} = 17506.73862 + 51065.43016*\text{PIX-MEAN} \quad (\text{with an } R^2 = 0.41) \tag{14}$$

$$\text{(5) ESLOGA117} = 0.8629622934 + 0.4338528040*\text{LOG}V \quad (\text{with an } R^2 = 0.29) \tag{15}$$

$$\text{ESTA117} = 10**\text{ESLOGA117}$$

$$\text{(6) EA102KM2} = 150.1919726 + 25.7192801*PCLA \quad (\text{with an } R^2 = 0.99) \tag{16}$$

$$\text{(7) EA117KM2} = 101.4798288 + 5.0239884*PCLA \quad (\text{with an } R^2 = 0.99) \tag{17}$$

$$\text{(8) EA120KM2} = 49.08312235 + 20.34672823*PCLA \quad (\text{with an } R^2 = 0.99) \tag{18}$$

It is noteworthy that the population density estimation models (Equations 16 through 18) produced almost perfect correlation coefficients.

These models were applied to 1608 counties, which were withheld for the purpose of accuracy assessment. The results are shown in Table 4, which only summarizes the overall estimation accuracy. The individual county breakdown is too large to reproduce here. However, the standard deviation (S.D.) of the mean relative error is shown for each model, which helps to indicate the variability of accuracy of the estimate by county. Clearly, models which produced a small mean relative error, a small standard deviation, and a small mean absolute relative error are best used to estimate population at the county level. These include model 1, which estimates population using light volume (V) as the independent variable; model 3, which estimates population using pixel mean (PIX-MEAN); and model 5, which estimates non-agricultural (urban) population using logarithmic transformation of the light volume (LOGV). Model 3 is also best in estimating the total population of China by adding the individual county population together (with an error of only −6.94 percent). Model 5, which is the allometric growth model, estimates non-agricultural (urban) population only, and is not too accurate in estimating the total non-agricultural population by adding all the county estimates together. It seems that pixel mean is better than light volume in estimating the total population of China, while light volume is best used to estimate the population of individual counties. The allometric growth model is most suited to estimate the non-agricultural population only.

Despite the very strong correlation between percent light area (*PCLA*) with population densities (overall, non-agricultural, and agricultural), models 6, 7, and 8 did not produce very accurate estimates. The best model is model 6 which has the lowest absolute mean errors and standard deviation. In other word, one can use percent light area (*PCLA*) to estimate population density by county or even the overall population density of the whole country. The model failed to give good results in low population density counties. It is possible to improve this model by dividing counties into high, medium, and low density groups, and derive one model for each group. This approach should be applicable for the improvement of all other models shown in Table 4.

Population Modeling at the City Level

Because the light areas of the DMSP-OLS nighttime images show very strong visual correlation with the built-up areas of the settlements, it is logical to assume that they are best used to estimate population at the city level. To test this hypothesis, 34 cities which have population data listed in the *China Statistical Yearbook 1998* were selected. China used the *hukou* (household registration) system to define its urban population (Chan and Xu, 1985). A city has two types of population statistics: (1) the overall population of the city, and (2) the non-agricultural population of the city. The first category includes population living in villages under the administrative control of the city, while the second category is population classified as non-agricultural by the government based on the fact that they live in the urban area. In the strictest sense, the second category (non-agricultural population) is the true urban population.

ArcView GIS was used to extract light area (*LA*) and pixel mean (PIX-MEAN) of each city from the classified DMSP-OLS nighttime image (vectorized) of China. Light volume (V) was

TABLE 4. RELATIVE ERRORS OF THE EIGHT MODELS APPLIED TO ESTIMATE POPULATION AND POPULATION DENSITIES AT THE COUNTY LEVEL (N = 1608).

Model No.	Dependent Variable	Independent Variable	Relative Mean (absolute)	Error % S.D.	Estimated Total	Correct Total from Census	% Error in Total
1	ESA102	V	20.94 (59.60)	103.33	742733583	891418404	−16.68
2	ESA117	V	77.02 (195.27)	388.17	204484924	194357262	5.21
3	PRA102	PIX-MEAN	39.40 (72.26)	127.92	829588299	891418404	−6.94
4	PRA117	PIX-MEAN	85.28 (136.02)	339.17	204781563	194357262	5.36
5	ESLOGA117	LOGV	26.51 (79.06)	134.33	94673112	194357262	−51.29
6	EA102KM2	PCLA	76.71 (109.72)	196.24	10160846	10866021	−6.49
7	EA117KM2	PCLA	607.10 (650.45)	904.81	2100817	6200673	−66.12
8	EA120KM2	PCLA	61.44 (121.97)	472.56	7926193	4603407	72.18

also computed as explained above. Logarithmic transformation of light area (LOG*LA*), light volume (LOG*V*), population (LOGPOP97), and non-agricultural population (LOGNAGP97) was carried out. Densities for both population (POP97D) and non-agricultural population (NAGPOP97D) were also calculated. Correlation coefficients among these variables were computed, and the results are shown in Table 5.

It is clear that light area, light volume, and pixel mean were all strongly correlated with non-agricultural population only (not the overall population of the city). The implication is that light clusters in the DMSP-OLS nighttime image display the built-up areas of a cities, and, as a result, they best reveal the true urban population size of Chinese cities. Light volume (*V*) was somewhat better correlated to non-agricultural population (NAGPOP97) than light area (*LA*), but in their logarithmic forms, both were equal in correlating with the logarithmic transformation of non-agricultural population (LOGNAGP97). As expected, percent light area (*PCLA*) was very strongly correlated with non-agricultural population density (NAGPOP97D) (0.87), but not so with the overall population density (POP97D)(0.43), thus reconfirming that the radiance-calibrated DMSP-OLS nighttime image reveals correctly the urban area.

Using a random sample of 17 cities out of the 34, the following models were generated, based on the observations given above:

$$\text{(1) ESNAGPOP1} = 746078 + 0.002537 * LA \quad (\text{with } R^2 = 0.51) \tag{19}$$

$$\text{(2) ESLOGNAGP1} = 0.338618 + 0.681704 * \text{LOG}LA \quad (\text{with } R^2 = 0.43) \tag{20}$$

$$\text{ESNAGP1} = 10^{**}\text{ESLOGNAGP1}$$

$$\text{(3) ESLOGPOP1} = 2.880938 + 0.435099 * \text{LOGLA} \quad (\text{with } R^2 = 0.13) \tag{21}$$

$$\text{ESPOP1} = 10^{**}\text{ESLOGPOP1}$$

$$\text{(4) ESLOGPOP2} = 2.959203 + 0.378141 * \text{LOGV} \quad (\text{with } R^2 = 0.12) \tag{22}$$

$$\text{ESPOP2} = 10^{**}\text{ESLOGPOP2}$$

$$\text{(5) ESLOGNAGP2} = 0.357406 + 0.602976 * \text{LOGV} \quad (\text{with } R^2 = 0.43) \tag{23}$$

$$\text{ESNAGP2} = 10^{**}\text{ESLOGNAGP2}$$

$$\text{(6) ESNAGPOP2} = 1038460 + 0.000154 * \text{V} \quad (\text{with } R^2 = 0.53) \tag{24}$$

$$\text{(7) ESNAGPOP3} = 1073326 + 143417 * \text{PIX-MEAN} \quad (\text{with } R^2 = 0.40) \tag{25}$$

$$\text{(8) ESNAGPDEN} = 397.256679 + 31.491779 * \text{PCLA} \quad (\text{with } R^2 = 0.30) \tag{26}$$

These models were applied to the DMSP-OLS nighttime image variables of the 17 cities withheld for accuracy evaluation. It was found that the best estimate of non-agricultural population (with a mean relative error of 28.66 percent and a mean absolute relative error of 51.62 percent) was achieved by model 2, the allometric growth model using logarithmic light area as input (Table 6). Model 5, which is also an allometric growth model but uses logarithmic light volume as input, gave the second best result (with a mean relative error of 40.28 percent and a mean absolute relative error of 61.44 percent) (Table 6). The third best estimate of non-agricultural population was achieved by model 1 (with a mean relative error of 48.27 percent and a mean absolute relative error of 66.86 percent), which simply used light area as input. It appeared that model 7 which used pixel mean as input gave the poorest result (with a relative error of 156.76 percent and a mean absolute relative error of 162.84 percent). The density of non-agricultural population estimated using model 8 and percent light area as input was moderately accurate (with a relative error of 40.84 percent and a mean absolute relative error of 61.58 percent). For each of these models, the sum of the estimates of non-agricultural population for individual cities was very close to the actual total, with the following results: +6.20 percent for model 1, −8.03 percent for model 2; and −4.32 percent for model 5. For model 8, the sum of the non-agricultural population densities was overestimated only by 3.58 percent. In other words, models 1, 2, and 5 can be used to estimate the total non-agricultural (urban) population in China.

On the whole, light area and light volume as employed in models 3 and 4, respectively, could not produce accurate estimates of the total population of the city (which by definition includes both agricultural and non-agricultural population as explained above). However, both models produced an extremely accurate estimate of the sum total of population of all the cities combined: +0.37 percent for model 3 and +3.65 percent for model 4. In other words, the radiance-calibrated DMSP-OLS low-light nighttime image can be used to extract up-to-date total population as well as total non-agricultural population of China. These are very valuable data to scholars studying urbanization in China. The DMSP-OLS data will allow an accurate assessment of the level of urbanization in China to be made.

To supplement Table 6, Table 7 shows city by city the estimation accuracy of models 2 and 8, the two best models which estimate the non-agricultural population and non-agricultural population density, respectively.

TABLE 5. CORRELATION COEFFICIENTS AMONG DMSP-OLS DERIVED VARIABLES AND POPULATION VARIABLES AT THE CITY LEVEL (N = 34) (THE NUMBER IN PARENTHESES UNDER THE CORRELATION COEFFICIENT INDICATES THE LEVEL OF SIGNIFICANCE).

	DMSP-OLS Derived Variables					
Population Variables	*LA*	LOGLA	*PCLA*	*V*	LOG*V*	PIX-MEAN
POP97	0.24 (0.17)	0.26 (0.14)	−0.15 (0.39)	0.28 (0.11)	0.24 (0.18)	0.23 (0.20)
LOGPOP97	0.42 (0.01)	0.54* (0.001)	−0.39 (0.02)	0.44 (0.009)	0.50* (0.003)	0.26 (0.14)
NAGPOP97	0.60* (0.0002)	0.62* (0.0001)	−0.10 (0.59)	0.72* (0.0001)	0.68* (0.0001)	0.67* (0.0001)
LOGNAGP97	0.62* (0.0001)	0.72* (0.0001)	−0.27 (0.12)	0.67* (0.0001)	0.72* (0.0001)	0.48* (0.004)
POP97D	−0.23 (0.20)	−0.29 (0.10)	0.43 (0.01)	−0.17 (0.34)	−0.24 (0.17)	0.08 (0.65)
NAGPOP97D	−0.16 (0.36)	−0.16 (0.36)	0.87* (0.0001)	−0.06 (0.76)	−0.17 (0.35)	0.31 (0.08)

*Highly significant (0.001 or smaller).

TABLE 6. ACCURACY OF THE EIGHT MODELS IN ESTIMATING POPULATION AND POPULATION DENSITY FROM DMSP-OLS NIGHTTIME IMAGES AT THE CITY LEVEL (N = 17)

Model No.	Variable Estimated	Relative Mean	Error (%) Absolute	Sum Total Estimated	Sum Total Actual	Sum Total Error%
1	non-agricultural population	48.27	66.86	45855870	43180000	+6.20
2	non-agricultural population	28.66	51.62	39713685	43180000	−8.03
3	population in city	72.81	99.52	88458247	88130000	+0.37
4	population in city	86.72	109.54	91344140	88130000	+3.65
5	non-agricultural population	40.28	61.44	41313395	43180000	−4.32
6	non-agricultural population	59.40	76.10	46568706	43180000	+7.85
7	non-agricultural population	156.76	162.84	71440840	43180000	+65.45
8	non-agricultural population density	40.84	61.58	54618.94	52732.67	+3.58

TABLE 7. ACCURACY OF ESTIMATING NON-AGRICULTURAL POPULATION AND NON-AGRICULTURAL POPULATION DENSITIES OF CITIES FROM DMSP-OLS NIGHTTIME IMAGES BY APPLYING EQUATIONS 20 AND 26

City	NAGP97**	ESNAGP1+	R.E. (%)	NAGPDEN# (p/km²)	ESNAGPDEN* (p/km²)	R.E. (%)
Shanghai	9430000	3005551	−68.13	8346.51	3222.23	−61.39
Chengdu	3180000	1615690	−49.19	2482.45	1399.66	−43.62
Tianjin	5150000	6602392	28.20	1306.57	2965.90	127.00
Wuhan	4230000	2620252	−38.06	3154.14	2343.34	−25.71
Changchun	2720000	2567738	−5.60	1814.31	2087.18	15.04
Shenyang	4230000	3834457	−9.35	1255.93	1751.87	39.49
Hangzhou	2040000	1532980	−24.85	4335.99	2924.07	−32.56
Zhengzhou	1980000	2377276	20.06	1921.08	2592.61	34.96
Fuzhou	1500000	1581076	5.41	1482.86	1626.97	9.72
Hefei	1300000	1307727	0.59	2719.15	2366.77	−12.96
Kunming	1640000	2372763	44.68	828.90	1537.69	85.51
Taiyuan	1920000	3126725	62.85	1372.84	2815.62	105.09
Lanzhou	1510000	2261479	49.77	908.24	1662.10	83.00
Xiamen	570000	1564664	174.50	1019.79	2588.96	153.87
Shenzhen	850000	1184203	39.32	3265.47	3524.66	7.94
Yinchuan	510000	1356500	165.98	388.81	1154.73	196.99
Haikou	420000	802214	91.00	16129.65	18054.61	11.93
Sum:	**43180000**	**39713685**	**−8.03**	**52732.67**	**54618.94**	**+3.58**
Mean Relative Error (%)			**28.66**			**40.84**
Mean Absolute Relative Error (%)			**51.62**			**61.58**

**NAGP97 = non-agricultural population for 1997 (data source: *China Statistical Yearbook 1998*, p.363).
+ ESNAGP1 = non-agricultural population estimated using logarithm of light area (Equation 20).
#NAGPDEN = non-agricultural population density calculated by dividing NAGP97 by the area of the city.
*ESNAGPDEN = non-agricultural population density estimated using per cent of light area (Equation 26).

Conclusions

Despite the limitations imposed by the unavailability of population and socio-economic data of China at the county level as "ground truth" during the same years as DMSP-OLS image acquisition, radiance-calibrated DMSP-OLS low-light nighttime images have been found to be capable of producing population data of China at three different spatial scales: province, county, and city. At the provincial level, it was found that the allometric growth model using light volume as the independent variable estimated non-agricultural population, while mean pixel values of the light areas predicted population densities only. It is noteworthy that light volume is a hybrid of light intensity and light area. At the county level, the allometric growth model using light volume continued to give accurate estimates of non-agricultural (urban) population. However, light volume and mean pixel value in the form of a simple linear regression model estimated the population (including both agricultural and non-agricultural population) quite well. This is an interesting finding because it implies that the light clusters on the radiance-calibrated DMSP-OLS image also reflect large rural settlements. Population densities were best estimated with the use of percent light area (or the density of saturated clusters in the DMSP-OLS image). At the city level, the allometric growth model has been successfully applied to estimate non-agricultural (urban) population. The independent variable of the model can be either light area or light volume, both giving good results. Pixel mean is not useful at all for population estimation. Percent light area is best for estimating non-agricultural population densities. The DMSP-OLS light clusters correspond best with the built-up areas of the cities, and hence are best used to estimate non-agricultural population at the city level. On the other hand, the DMSP-OLS images also have the potential to be used to estimate overall population (both rural and urban population) at the county level.

It is noteworthy that, at all spatial levels, the sums of the estimates have closely matched the true sums in nearly all the models. This observation applies to both total population and non-agricultural (urban) population. As expected, the best results were obtained at the city level (for the two types of city population used by the Chinese government). However, even at the provincial level, a reasonably accurate estimate of the total non-agricultural population can be obtained. Even population overall (including both rural and urban population) can be extracted from the radiance-calibrated DMSP-OLS image at the county level. The implication is that it is now possible to use remote sensing and statistical modeling to produce population estimates of a country quickly from space. The fact that urban population data can be extracted from the radiance-calibrated DMSP-OLS data is good news to many Chinese urban geographers, who will now have a means to check on China's

urban population size and, hence, its level of urbanization in a more objective manner.

The models developed in this paper can be further improved. It has already been suggested that separating the provinces, counties, or cities into different population density groups and developing allometric growth models in each group using light area and light volume extracted from the DMSP-OLS image will produce more accurate population estimates. There is also the potential to estimate energy consumption, personal income, and gross value of industrial output from the DMSP-OLS data, which have found to be strongly correlated with light intensity and light volume. All these are useful measures of economic development in a country. Therefore, it is possible to infer socio-economic implications from the pixels of the radiance-calibrated DMSP-OLS nighttime images, or "socializing the pixel," thus making the image data useful for social science research (Geoghegan *et al.*, 1998).

Acknowledgments

The author wishes to thank Dr. Kevin Gallo of NOAA for supplying the DMSP-OLS data of China for this research. Mr. Kai Wang, my research assistant, has helped with image processing and GIS analysis.

References

Brandl, H.W., 1978. The night eye in the sky, *Photogrammetric Engineering & Remote Sensing*, 44:503–505.

Chan, K.W., and X. Xu, 1985. Urban population growth and urbanization in China since 1949: Reconstructing a baseline, *The China Quarterly*, (104):583–613.

Croft, T.A., 1978. Nighttime images of the Earth from space, *Scientific American*, 239(July):86–98.

Dobson, J.E., E.A. Bright, P.R. Coleman, R.C. Durfee, and B.A. Worley, 2000. LandScan: A global population database for estimating populations at risk, *Photogrammetric Engineering & Remote Sensing*, 66:849–857.

Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis, 1997. Mapping city lights with nighttime data from the DMSP Operational Linescan System, *Photogrammetric Engineering & Remote Sensing*, 63:727–734.

Elvidge, C.D., K.E. Baugh, J.B. Dietz, T. Bland, P.C. Sutton, and H.W. Kroehl, 1999. Radiance calibration of DMSP-OLS low-light imaging data of human settlements, *Remote Sensing of Environment*, 68:77–88.

Geoghegan, J., L. Pritchard, Jr., Y. Ogneva-Himmelberger, and R.R. Chowdhury, 1998. "Socializing the pixel" and "pixeling the social" in land-use and land-cover change, *People and Pixels: Linking Remote Sensing and Social Science* (D. Liverman, E.F. Moran, R.R. Rindfuss, and R.C. Stern, editors), National Academy Press, Washington, D.C., pp. 51–69.

Lo, C.P., 1980. Shaping socialist Chinese cities: a model of form and land use, *China: Urbanization and National Development* (C.K. Leung and N. Ginsburg, editors), The University of Chicago Department of Geography, Research Paper No. 196, The University of Chicago, Chicago, Illinois, pp. 130–155.

———, 1995. Automated population and dwelling unit estimation from high-resolution satellite images: A GIS approach, *International Journal of Remote Sensing*, 16:17–34.

Lo, C.P., and R. Welch, 1977. Chinese urban population estimates, *Annals of the Association of American Geographers*, 67:246–253.

State Statistical Bureau, People's Republic of China, 1998. *China Statistical Yearbook 1998*, Chinese Statistical Publishing House, Beijing, China, 910 p.

Steinwand, D.R., 1994. Mapping raster imagery into the Interrupted Goode Homolosine Projection, *International Journal of Remote Sensing*, 15:3463–3472.

Sutton, P., D. Roberts, C. Elvidge, and H. Meij, 1997. Comparison of nighttime satellite imagery and population density for the continental United States, *Photogrammetric Engineering & Remote Sensing*, 63:1303–1313.

Tobler, W.R., 1969. Satellite confirmation of settlement size coefficients, *Area*, 1:30–34.

Welch, R., and S. Zupko, 1980. Urbanized area energy utilization patterns from DMSP data, *Photogrammetric Engineering & Remote Sensing*, 46:201–207.

(Received 28 September 2000; accepted 26 February 2001; revised 04 April 2001)